

Guilherme Menegon Arantes

Bioquímica e Biofísica Computacionais:
Simulações da Reatividade e da Estrutura de Biomoléculas

2016

Guilherme Menegon Arantes

**Bioquímica e Biofísica Computacionais:
Simulações da Reatividade e da Estrutura de Biomoléculas**

Texto apresentado como requisito para obtenção de título de Livre-Docente junto ao Departamento de Bioquímica da Universidade de São Paulo

Departamento de Bioquímica
Instituto de Química
Universidade de São Paulo

2016

*Dedico este trabalho a toda minha família, em especial
a minha filha, Julia, e a minha irmã, Fernanda.
Porque a família é como a raiz de uma árvore.
Sem ela, seríamos facilmente derrubados.*

Agradecimentos

Agradeço primeiramente a toda comunidade do Instituto de Química pelo convívio. Mesmo após 20 anos frequentando este Instituto, ainda me empolgo como se fosse a primeira vez que ando pelos seus longos corredores.

Aos colegas do Departamento de Bioquímica, personificados nas chefias da Maria Júlia Manso e do Chuck Farah, pela ajuda em estabelecer meu grupo de pesquisa física e tecnicamente. Ao Sandro Marana pelas conversas, constante apoio, colaborações e por ter sempre aberto as portas do seu laboratório. Ao Roberto Salinas e novamente ao Chuck pelas recentes colaborações que espero que floresçam.

Ao meu grupo de pesquisa, alunos e estagiários, em particular Ariane Alves, Murilo Teixeira, Raphael Sayegh e Vanesa Galassi. Com eles descobri que também é divertido acompanhar o crescimento de um aluno.

Ao Martin J. Field, meu principal colaborador e também amigo. Não seria exagero ele ter dividido o Nobel de Química em 2013.

Ao Luís Gustavo Dias, grande amigo e antigo mentor, pelas discussões científicas (ou nem tanto!).

Às agências de fomento: Pró-Reitoria de Pesquisa-USP pelo auxílio, CAPES e CNPq pelas bolsas, e FAPESP pelas bolsas e auxílios, principalmente o Jovem Pesquisador e os Regulares que permitiram a instalação do meu grupo de pesquisa e a montagem da nossa infraestrutura de pesquisa, usada em todos trabalhos apresentados aqui. Espero que esta Fundação volte a ser generosa com projetos individuais e independentes.

Sumário

1	Introdução	9
1.1	Organização do texto	10
2	Modelos e metodologias computacionais	11
2.1	Química quântica	11
2.2	Mecânica molecular	13
2.3	Potenciais híbridos QC/MM	15
2.4	Dinâmica molecular	16
2.5	Estimativa de energia livre	17
3	Enzimologia Computacional	19
4	Metaloproteínas de Ferro-enzofre	23
5	Flexibilidade conformacional e complexação	28
6	Conclusões e Perspectivas	32
	Referências	34
ANEXO A	A computational perspective on enzymatic catalysis	40
ANEXO B	The catalytic acid in the dephosphorylation of the Cdk2-pTpY/CycA protein complex by Cdc25B phosphatase	48
ANEXO C	Approximate Multiconfigurational Treatment of Spin-Coupled Metal Complexes	53
ANEXO D	Ferric-Thiolate Bond Dissociation Studied with Electronic Structure Calculations	63
ANEXO E	Homolytic cleavage of Fe-S bonds in Rubredoxin under mechanical stress	71
ANEXO F	Force induced chemical reactions on the metal center in a single metalloprotein molecule	75
ANEXO G	Flexibility and inhibitor binding in Cdc25 phosphatases	85
ANEXO H	Conformational Flexibility of the Complete Catalytic Domain of Cdc25B Phosphatases	102
ANEXO I	Ligand-receptor affinities computed by an adapted linear interaction model for continuum electrostatics and by protein conformational averaging	111
ANEXO J	Partition, orientation and mobility of ubiquinones in a lipid bilayer	123

Lista de abreviaturas e siglas

$\Delta\Delta G_{cat}$	Diminuição da barreira para uma reação catalisada, página 19
ΔG_{enz}^\ddagger	Energia de ativação no ambiente enzimático, página 19
h	Constante de Planck, página 17
k_B	Constante de Boltzmann, página 17
T	Temperatura, página 17
LJ	Lennard-Jones, página 15
MM	Mecânica molecular, página 13
PMF	Potencial de força média, página 17
PTP	Proteínas-fosfatase de tirosina, página 20
QC	Química quântica, página 12
QC/MM	Híbrido de química quântica e mecânica molecular, página 15
TS	Estado de transição, página 17

1 Introdução

Enzimas são catalisadores de extraordinária eficiência, capazes de acelerar reações químicas até vinte ordens de magnitude^[1]! Também possuem grande seletividade pelos seus substratos e sua atividade é microscopicamente controlável. São essenciais para catalisar reações bioquímicas e regular processos celulares. Do ponto de vista molecular, a vida como nós a conhecemos é orquestrada por enzimas.

Mas, como proteínas aceleram reações químicas? Um dos maiores desafios da ciência moderna é desvendar os mecanismos que propiciam tamanho poder e controle catalítico. Proponho neste texto contribuir com respostas para esta pergunta. Significa, a grosso modo, estudar o que os “elétrons estão fazendo” no sítio ativo das biomoléculas. Ou seja, a investigação das interações moleculares que podem influenciar uma reação química requer uma descrição da estrutura eletrônica da biomolécula, pelo menos da sua região reativa.

A melhor teoria que conhecemos para determinar a estrutura eletrônica de uma molécula é a mecânica quântica^[2]. Para sistemas simples, digamos uma molécula de água isolada, as equações da mecânica quântica permitem o cálculo de propriedades como geometria, energia da ligação covalente, espectro de absorção de luz, entre outras, com maior exatidão e precisão do que qualquer método experimental existente.

No entanto, as equações da mecânica quântica são muito complexas. Sua aplicação para estudar reações químicas de sistemas com mais do que uma dúzia de átomos apenas tornou-se viável nos últimos 40 anos, graças ao espetacular desenvolvimento das metodologias de cálculo e simulação, e da capacidade de processamento dos computadores.

Atualmente, a pesquisa científica em diversas áreas moleculares é permeada por cálculo e simulação computacional, com crescente impacto e importância^[3, 4, 5, 6]. Diversos prêmios Nobel em Química já foram concedidos para pesquisadores na área (Fukui & Hoffmann em 1981, Marcus em 1991, Kohn & Pople em 1998). O prêmio Nobel concedido em 2013 para Karplus, Levitt e Warshel pelo desenvolvimento de potenciais híbridos para simulação molecular é um claro reconhecimento de que métodos computacionais podem ser usados com grande sucesso para estudar biomoléculas e para responder perguntas, por exemplo, sobre os mecanismos microscópicos de catálise enzimática.

O texto apresentado aqui está apoiado nestas tradições teóricas. Descreveremos desenvolvimentos metodológicos e aplicações de cálculo e simulação computacionais de enzimas como proteínas-fosfatases e metaloproteínas de ferro-enxofre a fim de compreender seus mecanismos de catálise.

Seja na reatividade química ou em processos biofísicos que não envolvem a quebra ou formação de ligações covalentes, como na complexação e no reconhecimento intermolecular, é fundamental que aconteça um encaixe ou ajuste estrutural mediados por forças de interação entre as moléculas participantes. De fato, pode-se afirmar que todas as

funções e mecanismos biomoleculares são determinados pela estrutura tridimensional dos participantes e suas interações.

Uma propriedade fundamental de soluções aquosas em condições normais, como no interior celular, é que proteínas e outras biomoléculas solvatadas assumem um imenso conjunto de configurações, graças às flutuações térmicas e à fricção induzida por moléculas de solvente. Tanta flexibilidade estrutural sugere que proteínas podem ser consideradas nanomáquinas que funcionam graças a sua movimentação^[7].

Assim, outro objetivo deste texto é responder as perguntas: Qual é o papel da flexibilidade estrutural na formação de complexos e no reconhecimento entre biomoléculas? Como determinar o conjunto de configurações que melhor descreve a estrutura de uma biomolécula em solução? Nestas investigações também apresentaremos desenvolvimentos metodológicos e aplicações de simulação computacional baseadas em aproximações da mecânica quântica como a teoria de forças intermoleculares^[8, 9], e em metodologias da mecânica estatística^[10] que permitem conectar propriedades microscópicas com observáveis macroscópicas, principalmente a energia livre.

1.1 Organização do texto

Este documento compila nove artigos de minha autoria já publicados em revistas indexadas e de prestígio em suas áreas do conhecimento e um manuscrito em fase final de preparação. O primeiro critério que usei para escolha dos artigos foi a coerência dos temas pesquisados. Foi uma escolha natural visto que meu interesse e curiosidade em compreender a estrutura e a reatividade de biomoléculas utilizando modelagem molecular mantém-se constantes desde meu doutoramento.

O segundo critério foi usado para demonstrar minha independência científica. Assim, estão incluídos apenas artigos que sou autor para correspondência. Sou o único autor de parte destes trabalhos, o que reforça tal independência, mas também apresento artigos cujos resultados foram obtidos ou em colaboração com colegas do Departamento e do exterior, ou por alunos e pesquisadores membros de meu grupo de pesquisa e sob minha orientação direta.

Nos próximos capítulos estes artigos serão analisados criticamente para sistematizar minhas contribuições científicas. Reconhecendo a profundidade e a interdisciplinaridade dos temas, incluí um capítulo inicial que brevemente introduz as metodologias de cálculo e simulação computacionais empregadas. Em seguida, três capítulos são usados para apresentar as linhas de pesquisa. Todas são intimamente relacionadas pelo uso de métodos de simulação molecular como ferramenta de investigação. Além disso, a segunda linha (Capítulo 4) pode ser considerada uma extensão ou continuação da primeira (Capítulo 3), mas aplicada a metaloproteínas.

2 Modelos e metodologias computacionais

A atividade científica constituiu-se desde seus primórdios num diálogo entre experimento e teoria. Observações empíricas apenas não trazem qualquer explicação sobre a natureza. A interpretação destas observações, sim. Uma boa teoria traz explicações sobre um conjunto de fenômenos, deve ser experimentalmente testável e capaz de fazer previsões sobre fenômenos similares. Teorias são compostas por conjuntos de modelos e abstrações internamente consistentes que permitem a interpretação dos fenômenos^[11].

Modelos são idealizações ou simplificações de um objeto ou processo. Por construção, nenhum modelo é real. As aproximações e idealizações usadas para construir um modelo muitas vezes permitem sua descrição matemática, com grande abstração e, logo, generalidade. Por exemplo, uma série de fenômenos (bio)físicos é modelada por equações diferenciais^[12, 13].

A descrição matemática de um modelo é útil se as propriedades calculadas forem compatíveis às observáveis experimentais, proporcionando, assim, credibilidade ou uma validação para o modelo. Como muitas vezes um modelo depende de parâmetros, que podem ou devem ser ajustados, é necessária uma calibração inicial dos parâmetros antes da aplicação do modelo. Finalmente, modelos cada vez mais complexos são usados atualmente para idealizar ou simular a natureza. Daí vem o imenso impacto que métodos computacionais tem na ciência contemporânea já que computadores permitem que as descrições matemáticas subjacentes a tais modelos complexos sejam numericamente testadas e analisadas^[6].

O enfoque deste texto e da pesquisa científica apresentada aqui é em modelos de fenômenos bioquímicos e biofísicos moleculares. Não procuro estudar apenas um único sistema, como uma proteína, ou uma via metabólica ou de sinalização. Mas, através do estudo detalhado de vários sistemas, procuro por mecanismos gerais de funcionamento, que podem explicar o comportamento molecular de uma gama de processos ou sistemas biológicos.

Energia talvez seja o conceito abstrato mais importante neste texto. Parto do princípio que energias de interação e as forças interatômicas derivadas controlam todo comportamento microscópico. No restante deste capítulo serão apresentados os modelos teórico-computacionais usados na pesquisa descrita. Estes são usualmente classificados quanto ao nível teórico ou forma funcional que a energia é descrita.

2.1 Química quântica

Permita-me iniciar com uma famosa citação de Paul Dirac, um dos pais da mecânica quântica, em um artigo em 1929^[14]:

The general theory of quantum mechanics is now almost complete... The underlying physical laws necessary for a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed.

De fato, as leis físicas necessárias para descrição de toda química e, portanto, toda bioquímica já são conhecidas. O problema é que sua aplicação exata é impossível na maioria dos casos de interesse. Assim, nas últimas décadas, parte significativa da pesquisa realizada buscou aproximações e tratamentos simplificados das equações da mecânica quântica. Sua aplicação para determinação da estrutura eletrônica de átomos e moléculas é usualmente chamada de química quântica (QC).

Se considerarmos uma única geometria estática de uma molécula, a energia cinética dos elétrons e as interações entre os elétrons e os núcleos que compõem a molécula pode ser dada pela função hamiltoniana^[2, 15]:

$$\hat{H}_{QC} = - \sum_i^N \frac{1}{2} \nabla_i^2 - \sum_A^M \sum_i^N \frac{Z_A}{r_{Ai}} + \sum_{i>j}^N \frac{1}{r_{ij}} + \sum_{A>B}^M \frac{Z_A Z_B}{R_{AB}} \quad (2.1)$$

onde a equação é dada em unidades atômicas, ∇^2 é o operador laplaciano, Z_A é o número atômico do átomo no centro A , r_{Ai} é a distância entre o centro A e o elétron i , r_{ij} é a distância entre os elétrons i e j , e R_{AB} é a distâncias entre os centros atômicos A e B . As somatórias são efetuadas sobre os M núcleos atômicos e os N elétrons da molécula.

A estrutura eletrônica é, então, descrita por uma *função de onda* Ψ , que é determinada pela solução da equação de Schrödinger não-relativística e independente do tempo:

$$\hat{H}_{QC}\Psi = E\Psi \quad (2.2)$$

onde E é a energia do sistema molecular referente ao estado eletrônico descrito por Ψ que é função de $3N$ coordenadas eletrônicas.

A função de onda Ψ depende parametricamente das coordenadas nucleares ($\{\mathbf{R}_A\}$), ou seja, para cada geometria dos núcleos, teremos uma Ψ e E diferentes. A solução da equação 2.2 sob diferentes geometrias e acoplada a uma equação de propagação de movimento (seção 2.4) permite a descrição da evolução temporal do sistema molecular.

Pelo princípio da superposição^[2], a estrutura eletrônica ou a função de onda de uma molécula é uma simples combinação (soma) de diferentes configurações eletrônicas. Por exemplo, a molécula de hidrogênio tem sua estrutura eletrônica descrita por uma combinação de configurações iônicas, onde os dois elétrons estão conjuntamente mais próximos de um dos núcleos (como em um par H^+H^-), e de configurações covalentes, onde cada elétron está mais próximo de um dos núcleos, tipicamente ao longo do eixo que conecta os dois núcleos (*i.e.*, a ligação química, como indicado pela fórmula plana $H-H$).

Cada uma destas configurações eletrônicas é usualmente construída como um produto (antissimetrizado) de orbitais moleculares que descrevem a localização de pares de elétrons

com spins opostos. Teorias envolvendo orbitais moleculares estão no cerne dos estudos modernos sobre reatividade já que uma reação química nada mais é que um rearranjo dos orbitais moleculares e das configurações eletrônicas entre reagentes e produtos^[16, 17].

A medida que ampliamos ambas expansões, em configurações e em orbitais moleculares, nos aproximamos de uma solução exata da equação 2.2. Métodos QC que ampliam as expansões sistematicamente e não utilizam parâmetros empíricos são chamados de métodos *ab initio* ou de alto nível^[15]. Os métodos *ab initio* de maior exatidão são computacionalmente custosos, o que dificulta sua aplicação para sistemas complexos.

Uma maneira encontrada para melhorar a eficiência computacional foi truncar as expansões de configurações eletrônicas e dos orbitais para o menor número possível de termos. Os termos ausentes ou de difícil computação podem ser substituídos por parâmetros derivados de medidas experimentais ou de cálculos QC de alto nível. Neste caso, temos os chamados métodos QC semiempíricos, que são os mais eficientes empregados para modelagem de reatividade molecular^[18]. No entanto, paga-se um preço pela eficiência com a diminuição da acuracidade e da generalidade de aplicação.

A metodologia mais comum em QC atualmente é baseada na teoria do funcional da densidade (DFT)^[19]. Aqui, a descrição da estrutura eletrônica é feita pela densidade eletrônica que depende de apenas três coordenadas espaciais, ao contrário da função de onda que depende de $3N$ coordenadas espaciais. Esta redução na dimensionalidade implica que um potencial de troca-correlação descrevendo as interações entre elétrons seja usado na equação 2.1. No entanto, a forma exata deste potencial não é conhecida e diversas aproximações são usadas. A eficiência computacional e a exatidão dos cálculos de DFT é intermediária entre aquelas dos métodos *ab initio* de alto nível e dos métodos semiempíricos. Embora esta seja uma metodologia com boa acuracidade para moléculas orgânicas, a descrição de reações envolvendo centros metálicos por DFT pode ser problemática, porque a natureza multiconfiguracional da estrutura eletrônica destes centros não é descrita corretamente^[19, 20].

2.2 Mecânica molecular

A descrição da energia de interação entre biomoléculas solvatadas por métodos QC é custosa e, muitas vezes, desnecessária. Uma maneira muito mais eficiente é a descrição das interações por campos de força aditivos chamados de mecânica molecular (MM). Estas funções de energia são derivadas da teoria de forças intermoleculares^[8, 9] e da interpretação de espectroscopia vibracional^[21]. Mecânica molecular é largamente empregada para amostrar geometrias próximas ao equilíbrio. Em conjunto com a técnica de dinâmica molecular (seção 2.4), MM pode descrever a estrutura e as mudanças conformacionais de biomoléculas em processos e condições que simulam o comportamento no ambiente celular^[21, 22].

A energia de interação em uma descrição de MM, \mathcal{V}_{MM} , é tipicamente dividida em partes ligantes (ou covalentes), \mathcal{V}_{cov} , responsáveis pela manutenção da conectividade entre os átomos e da estrutura interna de cada molécula, e partes não ligantes, \mathcal{V}_{non} , responsáveis pela descrição das interações entre os átomos de diferentes moléculas ou entre regiões covalentemente distantes da mesma molécula:

$$\mathcal{V}_{MM} = \mathcal{V}_{cov} + \mathcal{V}_{non} \quad (2.3)$$

$$\mathcal{V}_{cov} = \mathcal{V}_{lig} + \mathcal{V}_{ang} + \mathcal{V}_{died} \quad (2.4)$$

$$\mathcal{V}_{non} = \mathcal{V}_{elet} + \mathcal{V}_{LJ} \quad (2.5)$$

onde cada um dos termos são descritos por funções específicas, explicadas abaixo.

As energias de estiramento de ligação e de ângulos entre ligações assumem uma forma harmônica:

$$\mathcal{V}_{lig/ang} = \sum_{lig/ang} \frac{1}{2} k_b (b - b_0)^2 \quad (2.6)$$

onde k_b é uma constante de força, b é o comprimento da ligação entre dois átomos ou o ângulo de ligação entre três átomos conectados sequencialmente e b_0 é a distância de ligação de equilíbrio ou o ângulo de equilíbrio. A somatória corre sobre todas as ligações ou os ângulos entre ligações definidos no sistema. Graças a sua forma harmônica, a energia aumentará indefinidamente a medida que a distância ou o ângulo de ligação se distorcerem dos valores de equilíbrio.

A energia de torção em torno de uma ligação, ou seja, a energia de rotação de um ângulo diedral ϱ , é descrita tipicamente por uma função periódica contínua como:

$$\mathcal{V}_{died} = \sum_{diedrais} \sum_n \frac{1}{2} K_n [1 + \cos(n\varrho - \delta)] \quad (2.7)$$

onde n é a periodicidade do ângulo, que varia normalmente de 0 a 4, δ é a fase, usualmente tomada como 0° , e K_n é uma constante de força. A primeira somatória corre sobre todos diedrais definidos no sistema e a segunda corre sobre as n funções cosseno usadas para cada diedral.

Cargas parciais (q_i), puntiformes e constantes são atribuídas aos centros atômicos de cada molécula para mimetizar sua distribuição de carga. Por exemplo, uma molécula de água teria uma carga $q_O = -0.8$ (unidades de carga eletrônica) sobre o oxigênio, mais eletronegativo, e $q_H = 0.4$ colocada sobre cada hidrogênio, de forma a reproduzir seu dipolo molecular e sua carga total neutra ($q_O + 2q_H = 0$). Assim, a interação eletrostática entre as moléculas do sistema é dada pela forma coulômbica clássica:

$$\mathcal{V}_{elet} = \sum_{i < j} \frac{q_i q_j}{r_{ij}} \quad (2.8)$$

onde a soma corre sobre todos os pares de centros atômicos i e j , e r_{ij} é a distância entre estes centros atômicos.

A repulsão entre as nuvens eletrônicas de cada molécula, que impede a sobreposição da matéria, e o efeito atrativo dispersivo de van der Waals, produzido pela correlação das flutuações da densidade eletrônica de dois grupos, são usualmente dados por um potencial de Lennard-Jones (LJ): :

$$\mathcal{V}_{LJ} = \sum_{i < j} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.9)$$

onde ϵ_{ij} é a profundidade (mínimo de energia) do potencial de LJ e σ_{ij} é a posição do fundo do poço de potencial.

Nota-se que um campo de força MM depende de diversos parâmetros como distâncias de equilíbrio b_0 , constantes de força k_b e K_n , cargas parciais q_i , etc., que são definidos para cada tipo atômico que compõe o sistema. Por exemplo, a carga parcial do hidrogênio na água é diferente da carga do hidrogênio ligado a um carbono alifático, assim como as distâncias de equilíbrio ou as constantes de força das ligações H–O e H–C são diferentes, etc. Assim, quando especificamos um campo de força como AMBER^[23] ou CHARMM^[24], tipicamente empregados para biomoléculas, definimos as formas funcionais (como as mostradas acima) e o conjunto de parâmetros associados que descrevem as energias de interação. Também nota-se que estas descrições aproximadas da energia de interação descartam diversas contribuições, como polarização ou transferência de carga, que são naturalmente incluídas em tratamentos mais rigorosos baseados em QC.

2.3 Potenciais híbridos QC/MM

Chegamos, então, em um breve dilema: Embora métodos QC sejam nossa melhor teoria para descrever o comportamento molecular, incluindo reatividade, são métodos computacionalmente custosos e impraticáveis para sistemas biológicos contendo muitos milhares de átomos. Por outro lado, métodos mais eficientes computacionalmente como MM não descrevem explicitamente a estrutura eletrônica das moléculas, impedindo a simulação da quebra e formação de ligações químicas, espectroscopia eletrônica, entre outros efeitos quânticos. Assim, como podemos estudar fenômenos eletrônicos em biomoléculas, por exemplo, o mecanismo de uma reação enzimática?

A solução óbvia é combinar ambas metodologias no que ficou conhecido como potenciais híbridos de química quântica e mecânica molecular (QC/MM, também chamado de QM/MM, onde QM vem de mecânica quântica). Embora inicialmente sugerida por Warshel e Levitt^[25], a implementação mais difundida e utilizada de potenciais híbridos foi proposta por Field e Karplus^[26]. O prêmio Nobel em Química foi concedido em 2013 justamente por estes desenvolvimentos metodológicos. Cabe notar aqui que o Prof. Martin Field (IBS, França) colabora regularmente com nosso grupo de pesquisa e parte dos estudos descritos aqui foi realizado em conjunto com seu grupo de pesquisa^[27, 28, 29, 30].

Potenciais híbridos QC/MM^[21, 26, 31, 32, 33] são representações particionadas da estrutura e da energia de um sistema molecular:

$$\hat{\mathcal{H}}_{eff} = \hat{\mathcal{H}}_{QC} + \mathcal{V}_{MM} + \hat{\mathcal{H}}_{QC/MM} \quad (2.10)$$

onde uma função de energia efetiva ($\hat{\mathcal{H}}_{eff}$) é definida pela adição de contribuições de uma região quântica do sistema ($\hat{\mathcal{H}}_{QC}$, seção 2.1), uma região de mecânica molecular clássica (\mathcal{V}_{MM} , seção 2.2) e uma contribuição da interação entre estas duas regiões usualmente dada por:

$$\hat{\mathcal{H}}_{QC/MM} = - \sum_i^{N_e} \sum_A^{M_m} \frac{q_A}{r_{iA}} + \sum_q^{M_q} \sum_A^{M_m} \frac{Z_q q_A}{r_{qA}} + \sum_q^{M_q} \sum_A^{M_m} \mathcal{H}_{QC/MM}^{vdW} \quad (2.11)$$

que inclui termos eletrostáticos de interação entre N_e elétrons (i) da região quântica e M_m cargas parciais (q_A) da região MM, entre as mesmas cargas e M_q núcleos quânticos (Z_q), e um termo para interações de van der Waals (vdW) entre centros MM e centros QC, usualmente descrito por um potencial de Lennard-Jones^[21].

Esta metodologia permite que fenômenos eletrônicos localizados, como a quebra de uma ligação química, ou a excitação eletrônica de um cromóforo, sejam simulados em fase condensada, na presença explícita de milhares de partículas descrevendo o restante do ambiente como a parte não-reativa de uma proteína, moléculas de água, ou lipídeos de uma membrana, que eventualmente solvatem ou complexem a região quântica ou reativa.

2.4 Dinâmica molecular

Qualquer que seja a descrição da energia de interação, por química quântica, mecânica molecular ou híbrida QC/MM, uma maneira para visitar as possíveis conformações e estruturas importantes do sistema é a simulação por dinâmica molecular^[21], equivalente à evolução temporal das partículas. Nesta técnica, a propagação ao longo do tempo das coordenadas do sistema pode ser feita seguindo-se a equação de movimento de Newton:

$$\mathbf{m} \frac{\partial^2 \mathbf{R}}{\partial t^2} = \mathbf{F} \quad (2.12)$$

em que \mathbf{m} é o vetor de massa das partículas, \mathbf{R} é o vetor de posições e \mathbf{F} é a força sobre o sistema. Esta equação da mecânica clássica é capaz de descrever com razoável rigor a evolução temporal das coordenadas de biomoléculas.

Dada uma estrutura e uma velocidade inicial das partículas que compõe o sistema biomolecular estudado, a integração da equação de Newton em um pequeno período de tempo e a conservação da energia podem ser usados para gerar outra configuração com novas posições atômicas e velocidades. Para proteínas, por exemplo, a estrutura inicial é usualmente retirada do banco de dados de proteínas, PDB^[34]. O gradiente da energia de

interação, ou seja, a derivada em relação às posições das funções de energia mostradas acima, determina as forças usadas na equação de movimento.

No limite de um longo tempo de simulação, podemos amostrar os estados conformacionais importantes visitados por um sistema molecular. São estes estados importantes que contribuirão significativamente para a energia livre e, logo, para as observáveis médias que pretendemos obter da simulação e comparar com as medidas experimentais.

2.5 Estimativa de energia livre

Energia livre é uma quantidade fundamental em sistemas termalizados pois determina a estabilidade e a cinética de processos e reações. Se considerarmos um equilíbrio hipotético entre dois estados, $A \rightleftharpoons B$, a constante de equilíbrio é:

$$K_{eq} = \exp \left[- \left(\frac{G_B - G_A}{RT} \right) \right] \quad (2.13)$$

onde R é a constante universal dos gases, T é a temperatura, e G_B é a energia livre do estado B. Pela teoria do estado de transição^[35], a constante cinética é:

$$k_{vel} = \left(\frac{k_B T}{h} \right) \exp \left[- \left(\frac{G^\ddagger - G_A}{RT} \right) \right] \quad (2.14)$$

onde k_B é a constante de Boltzmann, h é a constante de Planck e G^\ddagger é a energia livre do estado de transição (TS) que conecta o reagente A ao produto B.

Processos ativados como reações químicas ou mudanças conformacionais de biomoléculas podem ser convenientemente expressados em termos de uma coordenada de reação ξ que descreve o processo (por exemplo, uma distância de ligação que será quebrada) e um potencial de força média (PMF), que nada mais é do que a energia livre do processo ao longo da coordenada de reação. O PMF é definido a partir da função de distribuição $\langle \rho(\xi) \rangle$ ^[36]:

$$PMF(\xi) = k_B T \ln[\langle \rho(\xi) \rangle] \quad (2.15)$$

onde $\langle \dots \rangle$ significa uma média com peso de Boltzmann:

$$\langle \rho(\xi) \rangle = \frac{\int d\mathbf{R} \delta[\xi'(\mathbf{R}) - \xi] \exp[-\mathcal{V}(\mathbf{R})/k_B T]}{\int d\mathbf{R} \exp[-\mathcal{V}(\mathbf{R})/k_B T]} \quad (2.16)$$

onde $\mathcal{V}(\mathbf{R})$ é a energia potencial ou de interação do sistema em função das coordenadas \mathbf{R} e $\delta[x]$ é a função δ de Dirac que “filtra” as coordenadas \mathbf{R} diferentes de ξ . Na prática, a função de distribuição $\rho(\xi)$ é calculada a partir de um simples histograma normalizado da ocorrência da coordenada de reação ξ ao longo da simulação.

Uma vez que a presença de barreiras energéticas maiores que $k_B T$ ao longo da coordenada de reação dificulta a visitação de conformações importantes durante uma simulação, métodos de aumento de amostragem, como a amostragem por guarda-chuva (umbrella)^[36],

devem ser usados. Esta e outras metodologias similares foram empregadas em parte dos artigos compilados aqui^[37, 38, 39, 40].

Assim, estimativas computacionais de energia livre para processos de interesse podem ser diretamente comparadas com observações experimentais como constantes de equilíbrio (K_{eq}) e constantes cinéticas (k_{vel}) para validar os modelos de simulação empregados.

3 Enzimologia Computacional

Neste capítulo serão discutidos mecanismos catalíticos propostos para explicar o comportamento enzimático microscopicamente, conforme detalhado no Anexo A^[41]. Parte destes mecanismos, incluindo a importante contribuição eletrostática, é bem ilustrada pelas reações catalisadas por proteínas-fosfatase. A simulação dos mecanismos de reação da fosfatase Cdc25B é apresentada no Anexo B^[38].

Neste ponto acho pertinente introduzir duas definições. O termo “mecanismo catalítico” é usado para descrever as forças ou interações microscópicas usadas por enzimas para amplificar a velocidade de reações. Já “mecanismo de reação” é a sequência de transformações químicas (quebra e formação de ligações) e mudanças de estrutura observadas ao longo da reação.

Embora existam diversas sugestões de mecanismos catalíticos, foi novamente Arieh Warshel^[16, 42, 43] quem escrutinizou as propostas em comparações com resultados experimentais quantitativos e simulações computacionais. Sua primeira observação é que precisamos definir uma reação de referência em solução aquosa *na ausência de catalisador* e um ciclo termodinâmico que permitam a comparação com a reação enzimática. Catálise química somente ocorrerá quando a barreira para reação no sítio ativo da enzima (ΔG_{enz}^\ddagger) for menor do que a barreira em solução aquosa *para a mesma reação de referência* (ΔG_{sol}^\ddagger), ou seja, se a diminuição da barreira para uma reação catalisada $\Delta\Delta G_{cat} \equiv \Delta G_{enz}^\ddagger - \Delta G_{sol}^\ddagger < 0$.

Para uma série de enzimas analisadas, a principal espécie estabilizada é o estado de transição no sítio ativo da enzima, como proposto por Pauling^[44]. Outras sugestões como a desestabilização do estado reagente^[45], o modelo do sítio dividido proposto por Menger^[46] e a hipótese das conformações próximas ao ataque proposta por Bruice^[47] tem contribuições pequenas ou inexistentes para $\Delta\Delta G_{cat}$ (Anexo A)^[41].

Mas quais interações e forças moleculares promovem a estabilização do TS no sítio ativo? Novamente para uma série de enzimas, as contribuições para diminuição da energia de ativação foram quantificadas para diferentes mecanismos catalíticos^[48]. Catálise covalente, ou seja, a formação de um intermediário covalente ligado à enzima, contribui cerca de 5 kcal/mol em média para $\Delta\Delta G_{cat}$. Catálise ácido-base geral, quando ocorre a doação ou acepção de um H^+ diretamente pela enzima ou intermediada por moléculas de água ativada, também contribui cerca de 5 kcal/mol em média (Anexo A)^[41].

A principal força intermolecular otimizada pela evolução de enzimas tão eficientes é a interação eletrostática^[49]. Classifico como eletrostáticas todas interações do substrato ativado com cargas e dipolos da proteína, ligações de hidrogênio com as cadeias proteicas principais e laterais, assim como a coordenação não-covalente com metais ou grupos prostéticos carregados, já que estas interações podem ser razoavelmente bem descritas por

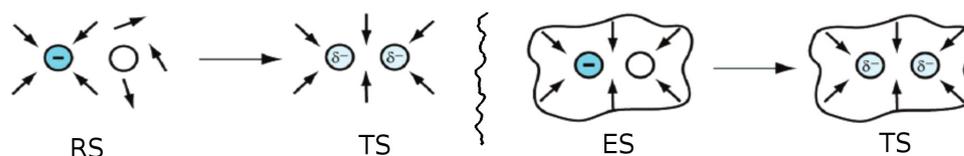


Figura 1 – Representação pictórica da distribuição dos dipolos (setas) que solvata reagentes (RS e ES, círculos) com carga localizada e estado de transição com carga deslocalizada numa reação em solução aquosa (esquerda) ou no sítio ativo de uma enzima (direita).

forças eletrostáticas dadas pelas equações 2.8 ou 2.11.

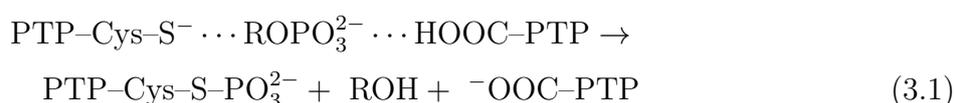
Em média a interação eletrostática contribui 17 kcal/mol para $\Delta\Delta G_{cat}$ ^[48, 41], graças principalmente a pre-orientação dos dipolos (e cargas parciais em geral) da proteína para estabilizar o TS. A Figura 1 indica que embora os dipolos do solvente sejam capazes de estabilizar o TS da reação em fase aquosa, estes dipolos devem se reorganizar, com um alto custo de energia livre. Por outro lado, a enzima uma vez enovelada tem os dipolos no sítio ativo pré-organizados para complementar a distribuição de carga do TS.

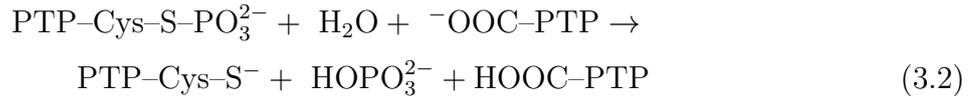
As contribuições para $\Delta\Delta G_{cat}$ de efeitos entrópicos, em que a flexibilidade do substrato estaria restringida no sítio ativo, e de impedimento estérico, em que a enzima poderia “forçar” estericamente o substrato numa determinada conformação reativa, são cerca de 2-3 kcal/mol, significativamente menores. Efeitos dinâmicos e quânticos como tunelamento também tem contribuição limitada para diminuição da energia de ativação enzimática^[48, 41].

Portanto, enzimas funcionam principalmente como um “super-solvente” eletrostático com baixa energia de reorganização. Mas cada enzima tem sua própria receita, dependente da reação química catalisada, em que uma combinação dos mecanismos catalíticos descritos acima pode ser utilizada (Anexo A)^[41].

Os mecanismos catalíticos apresentados podem ser ilustrados pelas reações catalisadas por proteínas-fosfatase de tirosina (PTP)^[50]. Estas enzimas atuam na sinalização celular ao contrabalancear a atividade de quinases e defosforilar fosfotirosina, além de fosfoerina e fosfotreonina no caso das fosfatases duais. Perturbações neste balanço estão envolvidas em diferentes tipos de câncer^[51], diabetes^[52], entre outras doenças. Assim, a compreensão do mecanismo de reação de diferentes fosfatases pode ser útil para o desenho de inibidores mais potentes e seletivos destas enzimas.

Investiguei as reações catalisadas por duas fosfatases: VHR^[37] e Cdc25B^[38]. Em ambas enzimas, a principal dúvida sobre o mecanismo da primeira etapa de reação era o estado de protonação do grupo fosfato no substrato: ou um diânion, conforme mostrado na equação química 3.1 abaixo, ou um monoânion monoprotonado:





As equações acima mostram as duas etapas de reação catalisadas por proteínas-fosfatase: tiólise de éster de fosfato que leva à desfosforilação do substrato; e hidrólise do intermediário tio-fosforilado que regenera a enzima livre. Em ambas reações temos um ácido-base geral HOOC-PTP e R representa o substrato.

Através de simulações híbridas QC/MM (seção 2.3) com uma descrição quântica semiempírica (seção 2.1) especialmente parametrizada para reações de transferência de fosfato^[53], levantei perfis de energia livre (seção 2.5) para diferentes propostas de mecanismo da primeira etapa de reação (Equação 3.1). Na VHR^[37], determinei que o substrato reage como um diânion já que a barreira calculada com este estado de protonação difere em apenas 1 kcal/mol da barreira obtida de experimentos cinéticos. A barreira calculada para as reações com o substrato monoprotonado é cerca de 15 a 25 kcal/mol mais alta, indicando que o substrato não reage quando protonado. Na VHR, também investiguei a reatividade do mutante sem o ácido-geral (D92A). A elevação da barreira calculada neste mutante foi de 5 kcal/mol, de acordo com a contribuição média indicada acima para o mecanismo de catálise ácido-base geral.

Para a fosfatase Cdc25B, utilizei um modelo contendo o substrato natural, a quinase dependente de ciclina 2 bisfosforilada, Cdk2-pTpY/CycA. Ou seja, o complexo de Michaelis simulado é um complexo ternário entre Cdc25B/Cdk2-pTpY/CycA^[54]. O sistema era composto por cerca de 130.000 átomos e, provavelmente, tratou-se da maior simulação híbrida QC/MM realizada até a época que o estudo foi conduzido (Anexo B)^[38].

A estrutura do sítio ativo no complexo pode ser visualizada no painel A da Figura 2. Estão mostrados apenas os átomos pesados do motivo catalítico conservado nas PTPs, o P-loop CX₅R. A estrutura deste sítio é similar àquela observada para os complexos de Michaelis da VHR com substratos artificiais^[37]. Nota-se aqui um preciso encaixe das interações eletrostáticas, principal mecanismo catalítico empregado por enzimas^[41]: os grupos NH da cadeia principal tem seus dipolos apontando e, portanto, estabilizando o grupo fosfato dianiônico do substrato. A cadeia lateral da arginina conservada (Arg479 na Cdc25B) tem carga positiva e também coordena o grupo fosfato. Contudo, a interação eletrostática ótima é atingida na região do TS ao longo do progresso da reação de transferência de fosfato quando as distâncias dos contatos NH...OPO₃²⁻ são minimizadas.

As superfícies de energia livre obtidas estão mostradas na Figura 2. Duas coordenadas de reação (seção 2.5) foram avaliadas explicitamente: $\xi_P = d(\text{PO}) - d(\text{PS})$, a diferença entre a distância da ligação quebrada e da ligação formada ao longo da transferência de fosfato; e $\xi_H = d(\text{OH})$, a distância entre o H⁺ doado e o oxigênio da cadeia lateral da treonina desfosforilada. A primeira observação relevante é que as reações do substrato mono- ou dianiônico se processam por mecanismos simultâneos de transferência de fosfato e de H⁺. Barreiras de mais de 45 kcal/mol teriam de ser transpostas para transferência por etapas,

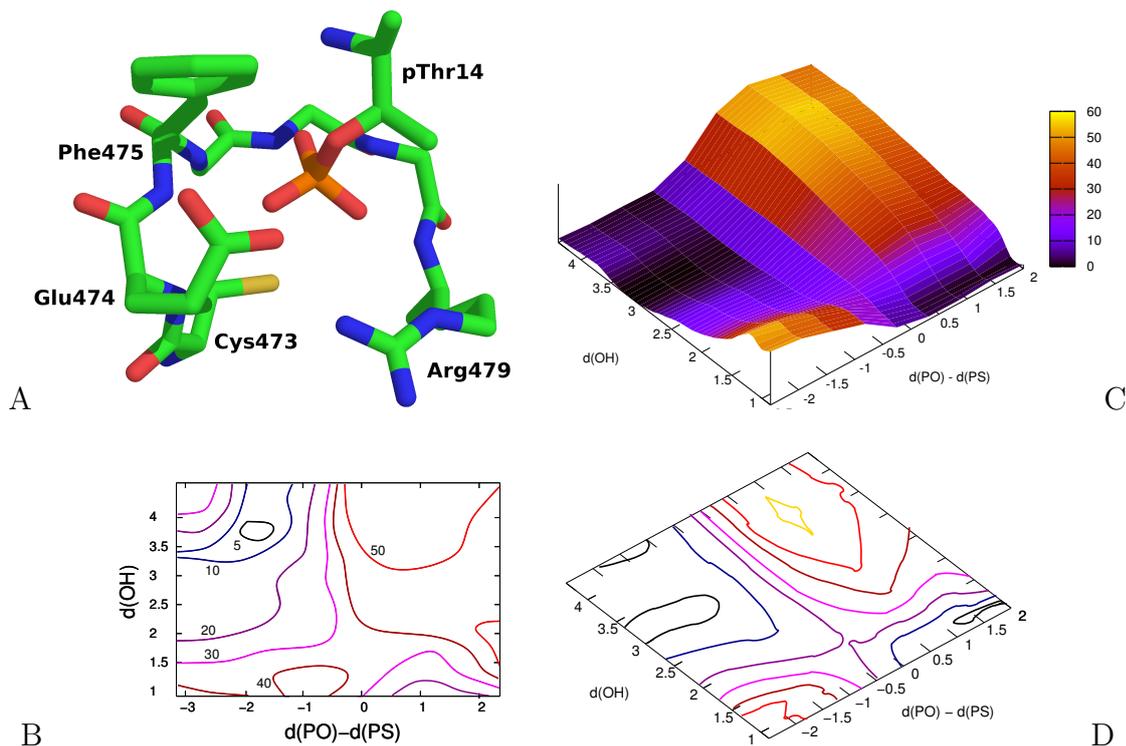


Figura 2 – Estrutura do sítio ativo da Cdc25B complexada ao seu substrato natural, fosfotreonina de Cdk2/CycA (painel A). Superfícies de energia livre em função de duas coordenadas de reação para desfosforilação catalisada pela Cdc25B do seu substrato natural na forma monoprotonada monoaniônica (painel B) e dianiônica (painel C e D). O painel D representa os mesmos dados do painel C na forma de curvas de nível com valores de energia livre indicados pela cor.

digamos, primeiro H^+ e depois fosfato, ou vice-versa.

A barreira calculada para a reação do substrato dianiônico com a transferência de H^+ a partir do ácido-geral E474 é de 17 kcal/mol, idêntica ao valor experimental^[38]. Por outro lado, a barreira calculada para a reação do substrato monoprotonado é de 38 kcal/mol, o que torna esta proposta mecanística energeticamente inviável. Portanto, o ótimo acordo entre as barreiras calculadas e experimentais valida as estimativas de energia livre com potenciais híbridos QC/MM e, logo, a determinação do mecanismo de reação catalisado pela Cdc25B frente ao seu substrato natural. Os mecanismos catalíticos subjacentes também estão de acordo com as teorias propostas para explicar a catálise enzimática^[43, 48, 41].

4 Metaloproteínas de Ferro-enxofre

Como mostrado no capítulo anterior, a simulação de reações de moléculas orgânicas na presença de milhares de átomos e interações em fase condensada pode ser muito bem sucedida. No entanto, o panorama é diferente para simular reações envolvendo átomos ou centros metálicos, principalmente metais de transição de camada aberta. A limitação aqui apresenta-se na descrição da estrutura eletrônica visto que as aproximações e os métodos existentes não são tão robustos ou exatos para o tratamento de metais quanto são para moléculas orgânicas^[15, 55, 28]. O problema é o imenso número de diferentes configurações eletrônicas com energia similar que estão acessíveis nos sistemas metálicos (seção 2.1).

De fato, o desenvolvimento de métodos para o tratamento de sistemas com estrutura eletrônica multiconfiguracional é a principal fronteira em aberto no cálculo de estrutura eletrônica^[15, 20, 56, 57, 30]. Investiguei uma possível solução deste problema para o caso de agregados polinucleares de metais de transição acoplados por interações de spin (Anexo C)^[20]. Desenvolvi uma série de aproximações fisicamente motivadas que resulta na seleção de configurações eletrônicas e na drástica diminuição do tamanho da expansão usada na função de onda destes agregados metálicos. Consequentemente, mostrei que o cálculo da estrutura eletrônica do complexo [2Fe-2S] torna-se factível e mais robusto, sem perda significativa na acuracidade (Anexo C)^[20].

Mas qual a relação entre a estrutura eletrônica de centros metálicos e a química de biomoléculas? Cerca de 30% das enzimas conhecidas contém metais essenciais para sua estrutura e atividade. Átomos metálicos possuem uma plástica e complexa estrutura eletrônica que permite às metaloenzimas estabelecer estruturas únicas assim como catalisar reações eficientemente, ou mesmo reações inacessíveis sem o acoplamento com o metal. Em particular, metaloproteínas de ferro-enxofre equipadas com agregados polinucleares como [2Fe-2S] e [4Fe-4S] são responsáveis pela reatividade e pelas transferências eletrônicas em diversos processos biológicos essenciais como a fotossíntese e a fosforilação oxidativa^[58]. Um dos meus objetivos de pesquisa a longo prazo é investigar a estrutura eletrônica e simular os mecanismos de reação destas metaloproteínas (Capítulo 6).

Rubredoxina é uma proteína carregadora de elétrons envolvida em processos de oxidorredução em bactérias^[59]. Seu centro redox possui apenas um átomo de ferro ligado a quatro cadeias laterais de cisteína ionizadas numa geometria tetraédrica (Figura 3A). É a mais simples proteína de ferro-enxofre conhecida e, portanto, foi escolhida como um alvo apropriado para estudos iniciais da estrutura eletrônica destas metaloproteínas.

Recentemente, fascinantes experimentos de microscopia de força atômica (AFM) “single-molecule” descreveram o desenovelamento da rubredoxina sob tensão mecânica. Como indicado na Figura 3B, polímeros lineares de rubredoxina são esticados e perfis de força por distância de tensionamento são medidos. O completo desenovelamento da rubredoxina

só ocorrerá se pelo menos duas ligações Fe–S forem rompidas. Dada a precisão e o grau de controle destes experimentos, diversas propriedades das ligações químicas Fe–S foram avaliadas^[60, 61, 62, 63].

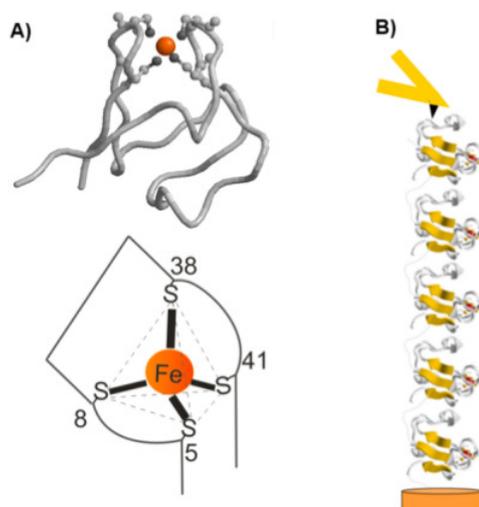


Figura 3 – A) No topo, a estrutura cristalográfica da rubredoxina cuja cadeia principal esta representada por um tubo cinza. O centro de ferro (laranja) e as cisteínas associadas estão representados por bolas e bastões. Embaixo, um esquema do centro de ferro e sua coordenação tetraédrica, com a numeração dos resíduos de cisteína indicados. B) Esquema do experimento de AFM “single-molecule” com poli-rubredoxina representada por um *cartoon*, ligada numa extremidade à ponta do microscópio e a uma base de sustentação na outra extremidade^[63].

Com o objetivo de simular os experimentos de AFM, iniciei o estudo da estabilidade da ligação Fe–S a partir de cálculos QC para a reação modelo (Anexo D)^[30]



que mimetiza a quebra de ligação observada na rubredoxina. Comparamos uma referência de estrutura eletrônica *ab initio* multiconfiguracional com diferentes funcionais de DFT e métodos semiempíricos (seção 2.1). O propósito foi encontrar um método de cálculo com acuracidade e eficiência computacional balanceadas. Determinamos que as eficientes funcionais OLYP e OPBE apresentam os melhores resultados em comparação com a referência *ab initio*, e que a funcional B3LYP é satisfatória, principalmente para descrições da geometria dos complexos (Anexo D)^[30]. Assim, obtivemos uma calibração de uma metodologia QC para ser aplicada nas simulações da proteína completa.

De volta à rubredoxina, uma das principais dúvidas que emergiram dos estudos por AFM foi o mecanismo de rompimento da ligação Fe–S^[60]. Poderia se tratar de uma quebra homolítica, em que o par de elétrons da ligação quebrada é dividido igualmente, resultando nos produtos radical tiolato (S^{-I}) e complexo ferroso (formalmente Fe^{2+} , com d^6 elétrons na camada de valência); ou um mecanismo heterolítico, em que o par de elétrons

ficaria somente com um dos produtos, resultando em ânion tiolato (S^{-II}) e complexo férrico (formalmente Fe^{3+} , com d^5 elétrons na camada de valência).

Investigamos ambas possibilidades por simulações com potenciais híbridos QC/MM (seção 2.3)^[27]. Uma estrutura da rubredoxina desenovelada, mas com as quatro ligações Fe–S ainda formadas, foi usada nas simulações, conforme mostrado na Figura 4. Esta foi obtida a partir da estrutura cristalográfica da rubredoxina e por simulação MM do desenovelamento forçado, empregando um procedimento também desenvolvido em nosso laboratório (veja Capítulo 6).

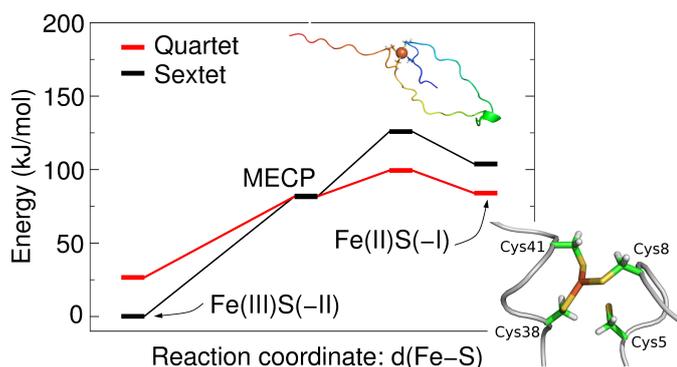


Figura 4 – Perfil de energia para a quebra da primeira ligação Fe–S da rubredoxina desenovelada nos estados de spin sexteto e quarteto. O inserto superior mostra a estrutura da rubredoxina parcialmente desenovelada e o inserto inferior mostra a geometria do centro ferro-enzofre após a quebra da ligação.

Nossos resultados mostram que a quebra da ligação é homolítica (Figura 4), já que os produtos da reação heterolítica são cerca de 50 kJ/mol mais instáveis (Anexo E)^[27]. Ainda mais interessante é a presença de um cruzamento entre estados de spin ao longo da reação. O estado fundamental do reagente tem estado de spin sexteto, com todos d^5 elétrons de valência do ferro desemparelhados. No entanto, o produto mais estável é um quarteto. Logo, identificamos um ponto de cruzamento entre as superfícies de energia dos dois estados de spin (MECP, Figura 4) ao longo da reação. Estes cruzamentos de spin e a consequente reatividade de dois estados eletrônicos é comum em complexos metálicos^[64] e também mostra-se relevante para estabilidade de agregados de ferro-enzofre^[27].

A investigação experimental da reatividade de centros metálicos em proteínas também é um desafio. Dada sua alta reatividade e a facilidade de oxidação, usualmente os metais encontram-se escondidos no interior proteico. Tipicamente sua reatividade é estudada em complexos modelo sintéticos ou miméticos daqueles encontrados em metaloenzimas, mas na ausência do ambiente proteico. Centros metálicos em proteínas também podem ser expostos com a desnaturação da estrutura proteica por agentes químicos ou físicos, mas de maneira descontrolada.

A vantagem de AFM é que a cadeia proteica pode ser manipulada e parcialmente desenovelada de forma controlada para expor o centro metálico e permitir o acesso do

solvente e de reagentes externos, mantendo parte dos contatos proteicos nativos. Assim, em colaboração com o grupo do Prof. Hongbin Li da University of British Columbia, Canadá, que conduziu os experimentos de AFM citados acima, investigamos as alterações causadas na estabilidade da ligação Fe–S na rubredoxina tensionada pela presença de reagentes como SCN^- , que compete pelo centro de ferro, e H^+ , que compete pelo grupo tiol das cisteínas.

Os resultados experimentais indicam que a presença destes reagentes diminui a força necessária para romper as ligações Fe–S^[29]. A Figura 5 mostra os resultados computacionais para as possíveis reações de substituição e rompimento de duas ligações Fe–S com o grupo tiol sequencialmente protonado, necessárias para desenovelar a rubredoxina completamente (veja acima). Além do rompimento puramente dissociativo^[27], também simulamos as reações de substituição por água. O caminho de reação mais provável é aquele com as menores barreiras, ou seja, a sequência de protonação e substituição por água (barreiras de 27 e 34 kJ/mol) das duas ligações Fe–S (Anexo F)^[29].

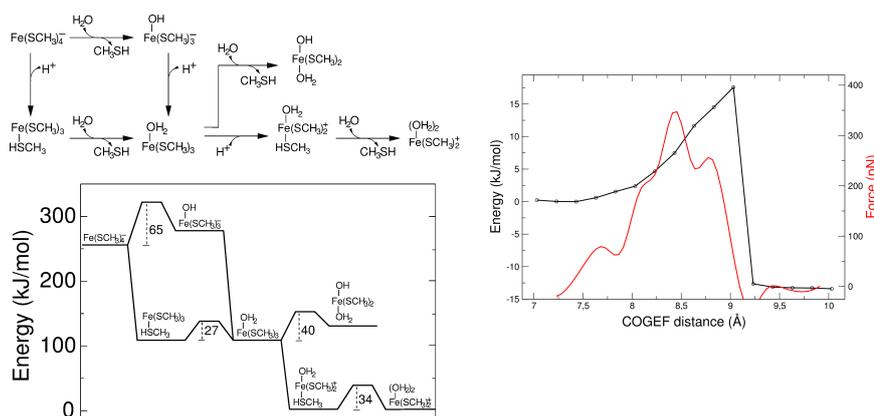


Figura 5 – Substituição e quebra das ligações Fe–S na rubredoxina. Painel da esquerda superior mostra as possíveis reações de duas ligações Fe–S sequencialmente protonadas. O painel da esquerda inferior mostra o perfil de energia calculado para as mesmas reações, com as barreiras indicadas e o painel da direita mostra os perfis de energia e de força calculados para etapa limitante.

Para discernir as diferentes propostas mecanísticas, calculamos a força necessária para rompimento da ligação Fe–S em cada etapa dos mecanismos propostos. Na etapa determinante da velocidade da reação com protonação (barreira de 34 kJ/mol, Figura 5), calculamos a força em cerca de 350 pN, em razoável acordo com a força medida por AFM, cerca de 160 ± 60 pN. As forças calculadas para outras propostas são ainda maiores e, portanto, em desacordo com os dados experimentais. Observamos também que todas reações de protonação e de substituição do centro de ferro-enxofre ocorrem por mecanismo heterolítico e no spin sexteto, sem o cruzamento entre estados eletrônicos^[29].

Assim, por uma combinação de experimentos de AFM com simulações computacionais, quantificamos e determinamos os possíveis mecanismos para desestabilização das ligações

Fe-S na rubredoxina quando o centro de ferro-enxofre é protonado ou tem seus ligantes substituídos (Anexo F)^[29].

5 Flexibilidade conformacional e complexação

Conforme levantado no Capítulo 1, a estrutura de biomoléculas é bastante flexível em condições normais. Em parte considerável do proteoma, segmentos contínuos com dezenas de aminoácidos possuem alta flexibilidade ou são intrinsecamente desordenados^[65, 66]. As conformações destes segmentos são alteradas quando interagem com outras proteínas ou até com pequenas moléculas^[67]. Então, como a distribuição estrutural destas proteínas e de outras biomoléculas flexíveis pode ser descrita? Qual o papel desta flexibilidade na formação de complexos?

De volta à fosfatase Cdc25B^[38], observei que a porção C-terminal do seu domínio catalítico apresentava consideráveis flutuações de posição durante uma simulação de dinâmica molecular^[68]. Estas flutuações e a possível desordem estrutural associada são compatíveis com observações experimentais, como a indefinição da densidade eletrônica correspondente à região C-terminal das fosfatases Cdc25B^[69] e sua isoforma Cdc25A^[70] a partir dos padrões de difração de raios-X obtidos de seus cristais.

Esta não é uma mera curiosidade estrutural. A porção C-terminal das Cdc25 tem um papel importante no reconhecimento do substrato natural^[54] e na atividade enzimática^[71]. Além disso, a porção C-terminal esta em contato direto com o sítio ativo e, portanto, também deve influenciar a complexação de pequenas moléculas que possivelmente atuem como inibidores competitivos da enzima. Assim, elegi a Cdc25B como um possível alvo para estudar as duas perguntas colocadas acima.

Para investigar o processo de complexação, utilizei três tipos de representações da estrutura da Cdc25B, com progressivo aumento de flexibilidade da região C-terminal: a estrutura cristalográfica^[69]; um conjunto de estruturas obtidos por simulação de dinâmica molecular^[68]; e um conjunto de estruturas obtidas de biblioteca de rotâmeros^[72]. Este último conjunto apresentava grandes flutuações e desordem na região C-terminal. Cada descrição foi usada para gerar complexos proteína-inibidor com duas moléculas orgânicas que são potentes inibidores da atividade enzimática (Anexo G)^[68].

Os resultados indicaram que o aumento da flexibilidade na cadeia principal da Cdc25B leva a uma maior diversidade de sítios e modos de ligação dos inibidores, principalmente pelo aparecimento de sítios crípticos ou cavidades transientes na superfície da região enovelada da proteína, que estão ocluídos nas representações estruturais menos flexíveis. Por exemplo, a Figura 6 mostra três modos de ligação obtidos para a representação estrutural de dinâmica molecular. Para o mesmo ligante, apenas um modo foi obtido para a estrutura cristalográfica, e oito modos foram observados para as estruturas obtidas de biblioteca de rotâmeros. A região C-terminal com maior flexibilidade, possivelmente

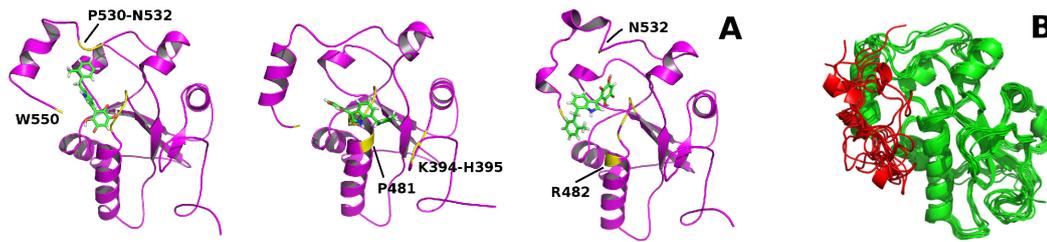


Figura 6 – Estrutura da Cdc25B obtida por simulação de dinâmica molecular. Painel A mostra três diferentes modos de ligação e uma variedade de contatos entre a Cdc25B e o ligante derivado de benzoquinona. Painel B mostra vinte estruturas sobrepostas do domínio catalítico completo da Cdc25B validadas por experimentos de RMN. A porção C-terminal desordenada está mostrada em vermelho.

desordenada^[72], não forma sítios ou contatos estáveis com os inibidores. Assim, essa região não deve contribuir significativamente para sua estabilidade, embora possa influenciar a cinética de entrada ou saída de inibidores competitivos (Anexo G)^[68].

No entanto, este trabalho computacional teve um caráter especulativo porque não possuía maiores evidências para discriminar qual seria a representação estrutural da Cdc25B mais apropriada^[68]. A etapa seguinte foi, então, buscar por validações experimentais que permitissem determinar a correta distribuição conformacional da proteína em solução.

Expressão, purificação e medidas de espectroscopia de ressonância magnética nuclear (RMN) do domínio catalítico completo da Cdc25B selvagem foram conduzidas pelo Raphael Sayegh sob minha orientação, e coorientado pelos colegas Profs. Sandro Marana e Roberto Salinas (Anexo H). RMN provém resolução atômica da estrutura e da dinâmica conformacional de biomoléculas em solução e, portanto, é ideal para comparações e validações das simulações moleculares.

Depois de superar problemas com a baixa estabilidade da proteína em solução concentrada, obtivemos diversos espectros de RMN que permitiram o assinalamento dos deslocamentos químicos de cerca de 85% dos átomos da cadeia principal, além de medidas de acoplamento residual dipolar e de parâmetros de relaxação dos grupos NH da cadeia principal. Estes dados indicaram claramente que a α -hélice presente na porção C-terminal da Cdc25B é estável e enovelada, contrariando os resultados de simulação anteriores^[68, 72]. Por outro lado, confirmaram que os últimos 16 aminoácidos da Cdc25B encontram-se desordenados^[68], conforme mostrado na Figura 6, e formam contatos metaestáveis com o restante da proteína (Anexo H)^[71, 54].

Assim como nos estudos de reatividade em que podemos calcular observáveis experimentais para validar quantitativamente as simulações (Capítulo 3), nos estudos estruturais podemos calcular todos parâmetros de RMN citados no parágrafo anterior a partir das trajetórias de simulação de dinâmica molecular. A comparação entre os valores pode ser usada para escolher um conjunto de configurações que esta de acordo com as observações

experimentais e descreve estatisticamente a estrutura em solução da Cdc25B (Capítulo 6). Procedimentos similares podem ser usados para qualquer outra biomolécula cujos dados estejam disponíveis^[73, 74].

Como podemos utilizar um conjunto de estruturas de uma proteína flexível para estimar sua afinidade por pequenas moléculas? Sugerir uma possível solução eficiente que foi executada pela Ariane Alves sob minha orientação (Anexo I)^[39]. A metodologia pode ser dividida em três etapas: dado conjunto de estruturas da proteína receptora, geramos complexos entre cada estrutura e a pequena molécula ligante utilizando ancoragem molecular^[75]; em seguida ranqueamos os complexos usando a aproximação de energia de interação linear^[76, 77]; finalmente agrupamos os complexos de acordo com seu ranqueamento numa única estimativa da energia livre de complexação (*i.e.*, a afinidade) do par receptor-ligante utilizando a teoria do ligante implícito^[78].

Para testar a metodologia, escolhemos três proteínas modelo: lisozima de fago T4, transcriptase reversa de HIV-1 e proteína ligante FK506 humana. Uma grande quantidade de complexos entre estas proteínas e diferentes ligantes tem estrutura cristalográfica e afinidade experimental conhecidas^[39]. Os resultados obtidos mostram que o método tem boa capacidade para prever afinidades para estes três receptores. É importante que o ranqueamento energético seja capaz de discriminar modos de ligação de falsos positivos se uma grande quantidade ($\sim 10^3$) de geometrias de complexos receptor-ligante for usada para estimar a afinidade. Conforme mostrado na Figura 7A, os complexos com maior afinidade correspondem ao sítio de ligação cristalográfico (Anexo I). Também notamos que o número e a diversidade de modos de ligação encontrados aumenta para as proteínas mais flexíveis, de acordo com meu estudo anterior^[68].

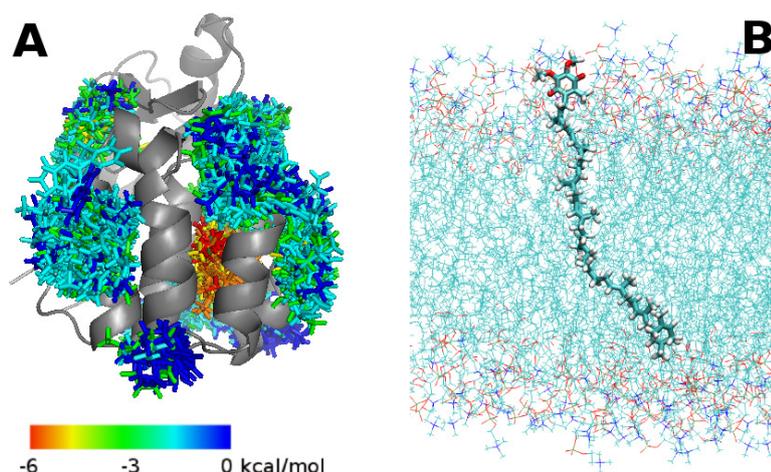


Figura 7 – A) Superposição de poses da lisozima T4 (cinza) complexada a 2-propilfenol, colorido de acordo com sua energia de ligação. O sítio de ligação cristalográfico esta populado com ligantes em vermelho e laranja. B) Conformação da ubiquinona (UQ10, em licorice) complexada à membrana de POPC (em linhas, cadeias acila em ciano, oxigênios em vermelho e nitrogênio da colina em azul).

Flexibilidade e dinâmica conformacional também são fundamentais para função de outras biomoléculas. Ubiquinona ou coenzima-Q é a molécula universal carregadora de carga em processos celulares de transferência eletrônica. Sua localização e partição membranar tem sido debatidas há décadas^[79, 80, 81, 82] e sua difusão entre os complexos respiratórios até já foi proposta como a etapa limitante da velocidade de toda fosforilação oxidativa^[83].

Com objetivo de resolver este longo debate e permitir futuros estudos da reatividade da ubiquinona nos complexos respiratórios (Capítulo 6), a Dra. Vanesa Galassi realizou sob minha supervisão a parametrização de um campo de força para ubiquinona e diversas simulações de sua complexação com membranas modelo (Anexo J)^[40].

A calibração do modelo molecular foi realizada meticulosamente pela parametrização da função de energia MM (seção 2.2) da ubiquinona em comparação com dados obtidos por cálculos QC de alto nível (seção 2.1) para complexos isolados, e pela estimativa de perfis de energia livre (seção 2.5) de complexação de ubiquinonas com diferentes comprimentos da cadeia isoprênica a membranas modelo de 1-palmitoil-2-oleoil-sn-glicero-3-fosfolina (POPC). As energias livres de ligação ubiquinona-membrana calculadas estão em ótimo acordo com as constantes de partição medidas experimentalmente para uma série homóloga de ubiquinonas. O mesmo acordo foi observado entre as constantes de difusão calculadas e experimentais^[40].

As simulações também indicam claramente a localização da ubiquinona inserida na membrana. A cabeça quinona esta na região interfacial da membrana, próxima aos grupos glicerol do lipídeo, altamente hidratada e com orientação normal ao plano da membrana. As primeiras unidades isoprênicas da cadeia estão estendidas e a cauda atravessa e interdiga a membrana nas ubiquinonas com seis ou mais unidades isopreno (Figura 7). Esta localização é notável porque coincide com a localização dos sítios ativos redox na citocromo bc_1 (Q_i e Q_o) e com a entrada do sítio de ligação do substrato na NADH:ubiquinona redutase. Logo, a evolução parece ter otimizado a localização destes sítios proteicos de acordo com as propriedades fisicoquímicas da ubiquinona para facilitar a (des)complexação deste substrato (Anexo J)^[40].

6 Conclusões e Perspectivas

Apresentei acima uma sistematização das minhas contribuições científicas recentes. Estes trabalhos foram realizados sob minha coordenação, como único autor ou em colaboração com colegas e estudantes sob minha orientação. Acredito que o texto demonstra a independência e a coesão da pesquisa realizada.

Sem dúvida, a importância de cálculo e simulação computacional em bioquímica e biofísica é crescente. Nestas áreas de pesquisa, simulações são tipicamente empregadas para ajudar ou aprofundar a interpretação de resultados experimentais. Antevejo que num futuro próximo, simulações terão ainda maior importância e capacidade preditiva, e serão usadas crescentemente para dirigir a pesquisa e propor experimentos validatórios. Isto já ocorre em diversas áreas da física, por exemplo. Nosso estudo da estrutura da Cdc25B e diversas outras linhas de pesquisa, notadamente o desenho de enzimas e estruturas proteicas^[84, 85], caminham por esta via.

Métodos computacionais, em particular híbridos QC/MM, contribuíram decisivamente para determinação das forças intermoleculares e dos mecanismos catalíticos que operam em diferentes enzimas. A estabilização eletrostática do estado de transição proporcionada pelo ambiente enzimático pré-organizado foi identificada como a principal contribuição para catálise. Em particular, determinei o mecanismo de reação para fosfatase Cdc25B, em ótimo acordo com observáveis experimentais (Capítulo 3).

A simulação molecular de metaloenzimas apresenta diversos desafios. Propomos aproximações para o estudo de agregados metálicos polinucleares, como os centros [2Fe-2S] encontrados na cadeia de transporte de elétrons da fosforilação oxidativa. Avaliamos quais metodologias aproximadas de cálculo de estrutura eletrônica seriam indicadas para estudar a reatividade das ligações Fe-S encontradas nesses agregados e investigamos a estabilidade desta ligação numa metaloproteína simples, a rubredoxina, em comparação com experimentos de AFM (Capítulo 4). Estes estudos formam uma sólida base para futuras investigações de metaloproteínas de ferro-enxofre mais complexas.

Investigamos o papel da flexibilidade estrutural no reconhecimento entre biomoléculas. Para a fosfatase Cdc25B, simulei a formação de complexos com pequenas moléculas para diferentes conjuntos de estruturas tridimensionais e conduzimos experimentos de RMN para validar qual o modelo estrutural mais apropriado. Propomos uma metodologia para prever a afinidade de complexos entre pequenas moléculas e proteínas flexíveis, representadas por conjuntos de estruturas. Determinamos a partição, localização e a difusão de ubiquinona complexada à membrana fosfolipídica, novamente em ótimo acordo com experimentos (Capítulo 5).

A construção de um *ensemble* conformacional que represente a estrutura da Cdc25B será realizada no futuro próximo pela combinação das medidas de RMN com observáveis

calculadas e pela seleção de geometrias estatisticamente bem distribuídas^[13, 73] obtidas de simulação de dinâmica molecular. Este *ensemble* conformacional será usado para previsão de modos de ligação da Cdc25B por inibidores e fosfato inorgânico, e de contatos transientes entre o núcleo proteico enovelado e a região C-terminal desordenada. Experimentos de RMN de perturbação do deslocamento químico serão usados na validação destas previsões.

Nossos futuros estudos de reatividade enzimática estarão focados nos mecanismos de reação de proteínas de ferro-enzofre. Em particular, estudaremos os três primeiros complexos respiratórios mitocondriais equipados com agregados de ferro-enzofre: NADH:ubiquinona redutase, succinato desidrogenase e citocromo bc_1 . Modelos tridimensionais destas proteínas embebidas em membranas lipídicas já foram construídos. Dentro desta linha, chamada de Bioenergética Molecular Computacional, investigaremos o mecanismo de acoplamento entre a redução de quinona e o bombeamento de H^+ no complexo I^[86] e as transferências eletrônicas observadas na etapa determinante de velocidade no ciclo-Q do complexo III^[87].

Finalmente, continuaremos a simulação e a interpretação de experimentos de AFM. No momento, terminamos o desenvolvimento e a calibração de um método de mecânica molecular capaz de descrever o rompimento de ligações covalentes mecanicamente induzidas.

Referências

- 1 LAD, C.; WILLIAMS, N. H.; WOLFENDEN, R. The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases. *Proc. Natl. Acad. Sci. USA*, v. 100, p. 5607–5610, 2003.
- 2 SLATER, J. C. *Quantum Theory of Atomic Structure*. 1st. ed. New York: McGraw-Hill, 1960.
- 3 FIELD, M. J. Simulating enzyme reactions: Challenges and perspectives. *J. Comp. Chem.*, v. 23, p. 48–58, 2002.
- 4 JORGENSEN, W. L. The many roles of computation in drug discovery. *Science*, v. 303, p. 1813–1818, 2004.
- 5 SHAW, D. E. et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, v. 330, p. 341–346, 2010.
- 6 WINSBERG, E. *Science in the Age of Computer Simulation*. 1st. ed. Chicago, EUA: The University of Chicago Press, 2010.
- 7 WEBER, J. K.; SHUKLA, D.; PANDE, V. S. Heat dissipation guides activation in signaling proteins. *Proc. Natl. Acad. Sci. USA*, v. 112, p. 10377–10382, 2015.
- 8 HIRSCHFELDER, J. O.; MEATH, W. J. The nature of intermolecular forces. *Adv. Chem. Phys.*, v. 12, p. 3–106, 1967.
- 9 JEZIORSKI, B.; MOSZYNSKI, R.; SZALEWICZ, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.*, v. 94, p. 1887–1930, 1994.
- 10 MCQUARRIE, D. A. *Statistical Mechanics*. 1st. ed. New York: Harper and Row, 1976.
- 11 GODFREY-SMITH, P. *Theory and Reality: An introduction to the philosophy of science*. 1st. ed. Chicago, EUA: The University of Chicago Press, 2003.
- 12 HIRSCH, M. W.; SMALE, S.; DEVANEY, R. L. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. 2nd. ed. Oxford, UK: Academic Press, 2004.
- 13 ZUCKERMAN, D. M. *Statistical Physics of Biomolecules: An Introduction*. 1st. ed. [S.l.]: CRC Press, 2010.
- 14 DIRAC, P. Quantum mechanics of many-electron systems. *Proc. R. Soc. London Ser. A*, v. 123, p. 714–733, 1929.
- 15 HELGAKER, T.; JØRGENSEN, P.; OLSEN, J. *Molecular Electronic-Structure Theory*. 1st. ed. New York: Wiley, 2000.
- 16 WARSHEL, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. 1st. ed. New York: Wiley, 1991.

- 17 SHAIK, S. S.; SCHLEGEL, H. B.; WOLFE, S. *Theoretical Aspects of Physical Organic Chemistry: The S_N2 Mechanism*. 1st. ed. New York: Wiley, 1992.
- 18 THIEL, W. Semiempirical methods. In: GROTEENDORST, J. (Ed.). *Modern Methods and Algorithms of Quantum Chemistry*. Jülich: John von Neumann Institute for Computing, 2001. v. 3, p. 261–283.
- 19 PARR, R. G.; YANG, W. *Density-Functional Theory of Atoms and Molecules*. 1st. ed. Oxford: Oxford University Press, 1996.
- 20 ARANTES, G. M.; TAYLOR, P. R. Approximate multiconfigurational treatment of spin-coupled metal complexes. *J. Chem. Theory Comput.*, v. 6, p. 1981–1989, 2010.
- 21 FIELD, M. J. *A Practical Introduction to the Simulation of Molecular Systems*. 1st. ed. Cambridge: Cambridge University Press, 1999.
- 22 JORGENSEN, W. L.; TIRADO-RIVES, J. Monte Carlo vs molecular dynamics for conformational sampling. *J. Phys. Chem.*, v. 100, p. 14508–14513, 1996.
- 23 CORNELL, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, v. 117, p. 5179–5197, 1995.
- 24 MACKERELL JR., A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, v. 102, p. 3586–3616, 1998.
- 25 WARSHEL, A.; LEVITT, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, v. 103, p. 227–249, 1976.
- 26 FIELD, M. J.; BASH, P. A.; KARPLUS, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, v. 11, p. 700–733, 1990.
- 27 ARANTES, G. M.; BHATTACHARJEE, A.; FIELD, M. J. Homolytic cleavage of Fe–S bonds in rubredoxin under mechanical stress. *Angew. Chem. Int. Ed.*, v. 52, p. 8144–8146, 2013.
- 28 BHATTACHARJEE, A. et al. Theoretical modeling of low energy electronic transitions in reduced cobaloximes. *ChemPhysChem*, v. 15, p. 2951–2958, 2014.
- 29 ZHENG, P. et al. Force induced chemical reactions on the metal center in a single metalloprotein molecule. *Nature Comm.*, v. 6, p. 7569, 2015.
- 30 ARANTES, G. M.; FIELD, M. J. Ferric-thiolate bond dissociation studied with electronic structure calculations. *J. Phys. Chem. A*, v. 119, p. 10084–10090, 2015.
- 31 FIELD, M. J. The pDynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials. *J. Chem. Theory Comput.*, v. 4, p. 1151–1161, 2008.
- 32 FIELD, M. J. et al. The DYNAMO library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.*, v. 21, p. 1088–1100, 2000.

- 33 ARANTES, G. M.; RIBEIRO, M. C. C. A microscopic view of substitution reactions solvated by ionic liquids. *J. Chem. Phys.*, v. 128, p. 114503, 2008.
- 34 The Protein Data Bank. 2002. Royal Society Protein Data Bank. <http://www.pdb.org>.
- 35 TRUHLAR, D. G.; GARRETT, B. C.; KLIPPENSTEIN, S. J. Current status of transition-state theory. *J. Phys. Chem.*, v. 100, p. 12771–12800, 1996.
- 36 ROUX, B. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Comm.*, v. 91, p. 275–282, 1995.
- 37 ARANTES, G. M. Free energy profiles for catalysis by dual-specificity phosphatases. *Biochem. J.*, v. 399, n. 2, p. 343–350, 2006.
- 38 ARANTES, G. M. The catalytic acid in the dephosphorylation of the Cdk2-pTpY/CycA protein complex by Cdc25B phosphatase. *J. Phys. Chem. B*, v. 112, p. 15244–15247, 2008.
- 39 NUNES-ALVES, A.; ARANTES, G. M. Ligand-receptor affinities computed by an adapted linear interaction model for continuum electrostatics and by protein conformational averaging. *J. Chem. Inf. Model.*, v. 54, p. 2309–2319, 2014.
- 40 GALASSI, V. V.; ARANTES, G. M. Partition, orientation and mobility of ubiquinones in a lipid bilayer. *Biochim. Biophys. Acta*, v. 1847, p. 1345, 2015.
- 41 ARANTES, G. M. A computational perspective on enzymatic catalysis. *Quim. Nova*, v. 31, p. 377–383, 2008.
- 42 ÅQVIST, J.; WARSHEL, A. Simulation of enzyme-reactions using valence-bond force-fields and other hybrid quantum-classical approaches. *Chem. Rev.*, v. 93, p. 2523–2544, 1993.
- 43 WARSHEL, A. et al. Electrostatic basis for enzyme catalysis. *Chem. Rev.*, v. 106, p. 3210–3235, 2006.
- 44 PAULING, L. Nature of forces between large molecules of biological interest. *Nature*, v. 161, p. 707–709, 1948.
- 45 JENCKS, W. P. *Catalysis in Chemistry and Enzymology*. 1st. ed. New York: Dover, 1987.
- 46 MENGER, F. M. Analysis of ground-state and transition-state effects in enzyme catalysis. *Biochemistry*, v. 31, p. 5368–5373, 1992.
- 47 BRUICE, T. C.; LIGHTSTONE, F. C. Ground state and transition state contributions to the rates of intramolecular and enzymatic reactions. *Acc. Chem. Res.*, v. 32, p. 127–136, 1999.
- 48 GARCIA-VILOCA, M. et al. How enzymes work: Analysis by modern rate theory and computer simulations. *Science*, v. 303, p. 186–195, 2004.
- 49 WARSHEL, A.; FLORIÁN, J. Computer simulation of enzyme catalysis: Finding out what has been optimized by evolution. *Proc. Natl. Acad. Sci. USA*, v. 95, p. 5950–5955, 1998.

- 50 GOLDSTEIN, B. J. *Tyrosine Phosphoprotein Phosphatases*. 2nd. ed. Oxford: Oxford University Press, 1998.
- 51 OSTMAN, A.; HELLBERG, C.; BOHMER, F. Protein-tyrosine phosphatases and cancer. *Nat. Rev. Cancer*, v. 6, p. 307–320, 2006.
- 52 JOHNSON, T.; ERMOLIEFF, J.; MR, M. J. Protein tyrosine phosphatase 1b inhibitors for diabetes. *Nat. Rev. Drug Discovery*, v. 1, p. 696–709, 2002.
- 53 ARANTES, G. M.; LOOS, M. Specific parametrisation of a hybrid potential to simulate reactions in phosphatases. *Phys. Chem. Chem. Phys.*, v. 8, p. 347–353, 2006.
- 54 SOHN, J. et al. Experimental validation of the docking orientation of Cdc25 with its Cdk2-CycA protein substrate. *Biochemistry*, v. 44, p. 16563–16573, 2005.
- 55 CRAMER, C. J.; TRUHLAR, D. G. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.*, v. 11, p. 10757–10816, 2009.
- 56 CHAN, G. K.-L. Low entanglement wavefunctions. *WIREs Comput. Mol. Sci.*, v. 2, p. 907–920, 2012.
- 57 OLIVARES-AMAYA, R. et al. The ab-initio density matrix renormalization group in practice. *J. Chem. Phys.*, v. 142, p. 034102, 2015.
- 58 BEINERT, H.; HOLM, R. H.; MUNCK, E. Iron-sulfur clusters: Nature’s modular, multipurpose structures. *Science*, v. 277, p. 653–659, 1997.
- 59 VOET, D.; VOET, J. *Biochemistry*. 2nd. ed. New York: Wiley, 1995.
- 60 ZHENG, P.; LI, H. Highly covalent ferric-thiolate bonds exhibit surprisingly low mechanical stability. *J. Am. Chem. Soc.*, v. 133, p. 6791–6798, 2011.
- 61 ZHENG, P. et al. Hydrogen bond strength modulates the mechanical strength of ferric-thiolate bonds in rubredoxin. *J. Am. Chem. Soc.*, v. 134, p. 4124–4131, 2012.
- 62 ZHENG, P. et al. Single molecule force spectroscopy reveals that iron is released from the active site of rubredoxin by a stochastic mechanism. *J. Am. Chem. Soc.*, v. 135, p. 7992–8000, 2013.
- 63 ZHENG, P. et al. Single molecule force spectroscopy reveals the molecular mechanical anisotropy of the fes4 metal center in rubredoxin. *J. Am. Chem. Soc.*, v. 135, p. 17783–17792, 2013.
- 64 SCHRODER, D.; SHAIK, S.; SCHWARZ, H. Two-state reactivity as a new concept in organometallic chemistry. *Acc. Chem. Res.*, v. 33, p. 139–145, 2000.
- 65 DUNKER, A. K. et al. The roles of intrinsic disorder in protein interaction networks. *FEBS Lett.*, v. 272, p. 5129–5148, 2005.
- 66 RADIVOJAC, P. et al. Intrinsic disorder and functional proteomics. *Biophys. J.*, v. 92, p. 1439–1456, 2007.
- 67 WRIGHT, P. E.; DYSON, H. J. Linking folding and binding. *Curr. Op. Struct. Biol.*, v. 19, p. 31–38, 2009.

- 68 ARANTES, G. M. Flexibility and inhibitor binding in Cdc25 phosphatases. *Proteins: Struct. Funct. Genet.*, v. 78, p. 3017–3032, 2010.
- 69 REYNOLDS, R. A. et al. Crystal structure of the catalytic subunit of Cdc25b required for G(2)/M phase transition of the cell cycle. *J. Mol. Biol.*, v. 293, p. 559–568, 1999.
- 70 FAUMAN, E. B. et al. Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell*, v. 93, p. 617–625, 1998.
- 71 WILBORN, M. et al. The c-terminal tail of the dual-specificity Cdc25B phosphatase mediates modular substrate recognition. *Biochemistry*, v. 40, p. 14200–14206, 2001.
- 72 MAMONOV, A. B. et al. General library-based monte carlo technique enables equilibrium sampling of semi-atomistic protein models. *J. Phys. Chem. B*, v. 113, p. 10891–10904, 2009.
- 73 LEUNG, H. T. A. et al. A rigorous and efficient method to reweight very large conformational ensembles using average experimental data and to determine their relative information content. *J. Chem. Theory Comput.*, v. 12, p. 383–394, 2016.
- 74 NODET, G. et al. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.*, v. 131, p. 17908–17918, 2009.
- 75 SOUSA, S. F.; FERNANDES, P. A.; RAMOS, M. J. Protein–ligand docking: current status and future challenges. *Proteins*, v. 65, p. 15–26, 2006.
- 76 GALLICCHIO, E.; KUBO, M. M.; LEVY, R. M. Enthalpy–entropy and cavity decomposition of alkane hydration free energies: numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B*, v. 104, p. 6271–6285, 2000.
- 77 ÅQVIST, J.; LUZHKOV, V. B.; BRANDSDAL, B. O. Ligand binding affinities from md simulations. *Acc. Chem. Res.*, v. 35, p. 358–365, 2002.
- 78 MINH, D. D. L. Implicit ligand theory: rigorous binding free energies and thermodynamic expectations from molecular docking. *J. Chem. Phys.*, v. 137, p. 104106, 2012.
- 79 KATSIKAS, H.; QUINN, P. The polyisoprenoid chain length influences the interaction of ubiquinones with phospholipid bilayers. *Biochim. Biophys. Acta*, v. 689, p. 363–369, 1982.
- 80 LENAZ, G. et al. Localization and preferred orientations of ubiquinone homologs in model bilayers. *Biochem. Cell. Biol.*, v. 70, p. 504–514, 1992.
- 81 METZ, G. et al. Nmr studies of ubiquinone location in oriented model membranes: Evidence for a single motionally-averaged population. *J. Am. Chem. Soc.*, v. 117, p. 564–565, 1995.
- 82 HAUSS, T. et al. Localization of coenzyme q10 in the center of a deuterated lipid membrane by neutron diffraction. *Biochim. Biophys. Acta*, v. 1710, p. 57–62, 2005.

-
- 83 LENAZ, G.; FATO, R. Is ubiquinone diffusion rate-limiting for electron transfer? *J. Bioener. and Biomem.*, v. 18, p. 369–401, 1986.
- 84 DAS, R.; BAKER, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.*, v. 77, p. 363–382, 2008.
- 85 FLEISHMAN, S.; BAKER, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell*, v. 149, p. 262–273, 2012.
- 86 SAZANOV, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex i. *Nature Rev Mol Cell Biol*, v. 16, p. 375–388, 2015.
- 87 CROFTS, A. R. et al. The Q-cycle reviewed: How well does a monomeric mechanism of the bc1 complex account for the function of a dimeric complex? *Biochim. Biophys. Acta*, v. 1777, p. 1001–1019, 2008.

ANEXO A – A computational perspective on enzymatic catalysis

UMA PERSPECTIVA COMPUTACIONAL SOBRE CATÁLISE ENZIMÁTICA

Guilherme M. Arantes*

Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900 São Paulo - SP, Brasil

Recebido em 24/11/06; aceito em 18/5/07; publicado na web em 19/12/07

A COMPUTATIONAL PERSPECTIVE ON ENZYMATIc CATALYSIS. Enzymes are extremely efficient catalysts. Here, part of the mechanisms proposed to explain this catalytic power will be compared to quantitative experimental results and computer simulations. Influence of the enzymatic environment over species along the reaction coordinate will be analysed. Concepts of transition state stabilisation and reactant destabilisation will be confronted. Divided site model and near-attack conformation hypotheses will also be discussed. Molecular interactions such as covalent catalysis, general acid-base catalysis, electrostatics, entropic effects, steric hindrance, quantum and dynamical effects will also be analysed as sources of catalysis. Reaction mechanisms, in particular that catalysed by protein tyrosine phosphatases, illustrate the concepts.

Keywords: enzymatic catalysis; computer simulation; reaction mechanism.

INTRODUÇÃO

Enzimas são catalisadores de extraordinária eficiência. Os aumentos de velocidade de reações alcançados por enzimas podem chegar a vinte ordens de magnitude¹! Por outro lado, enzimas possuem uma grande especificidade pelos substratos, sua atividade é controlável e totalmente seletiva.

Sob uma perspectiva teórica, a sugestão de Fischer², conhecida como a hipótese da “chave e fechadura”, foi a primeira proposta para explicar o poder catalítico das enzimas. Com o advento da teoria do estado de transição³ nos anos 1930, Pauling⁴ propôs que esta espécie seria preferencialmente ligada pelo sítio ativo enzimático. Já em 1969, Jencks⁵ escreveu que “o estudo dos mecanismos moleculares de catálise enzimática é necessariamente empírico e qualitativo”. No entanto, nos últimos 30 anos, simulações computacionais que permitem determinações quantitativas de propriedades termodinâmicas têm alterado este panorama, apontando e quantificando os mecanismos catalíticos empregados por enzimas.

Cabem aqui duas definições. O termo “mecanismo catalítico” é usado para descrever as forças ou interações microscópicas utilizadas pelas enzimas para amplificar a velocidade de reações. Já “mecanismo de reação” é a sequência de transformações (quebra e formação de ligações químicas) e mudanças de estrutura do complexo enzimático observadas ao longo do progresso da reação.

Diversos mecanismos catalíticos já foram propostos para explicar as acelerações enzimáticas em nível microscópico. Por exemplo, Menger contabilizou 21 teorias sobre catálise enzimática⁶. Este número é elevado em consequência de diferentes interpretações semânticas e imprecisas definições de quantidades termodinâmicas nas teorias propostas, e também porque a maior parte destas teorias não foi (ou não pode ser) testada quantitativamente. Portanto, há muita controvérsia sobre a importância de cada um dos diferentes mecanismos e teorias propostos para explicar o poder catalítico das enzimas⁷⁻⁹.

Sem dúvida, cada enzima possui sua própria “receita”⁶ para catálise, em que uma combinação das interações propostas pode estar atuando. O principal objetivo deste texto é analisar e

quantificar a contribuição de cada uma destas interações para o aumento da velocidade de reações no ambiente enzimático. Inicialmente apresentaremos o modelo básico e suas equações macroscópicas para a atividade enzimática, e a definição de reação de referência em solução aquosa, que deve ser usada na comparação com a reação enzimática. Em seguida, discutiremos qual espécie ao longo da coordenada de reação sofre influência determinante do ambiente enzimático. A hipótese de desestabilização do estado reagente (complexo enzima-substrato ou complexo de Michaelis) será confrontada com a hipótese de estabilização do estado de transição (TS)³. Propostas relacionadas como as conformações próximas ao ataque¹⁰ e do sítio dividido⁶ também serão analisadas.

As interações moleculares usadas para explicar os mecanismos catalíticos, incluindo catálise covalente, catálise ácido-base geral, interação eletrostática, efeitos entrópicos, impedimento estérico e, efeitos quânticos e dinâmicos serão discutidas. Esta lista de interações pode ser incompleta, dada a limitação de tamanho para este texto, e parcial, pois reflete as propostas que foram testadas e quantificadas por simulações computacionais. Mecanismos de reação de algumas enzimas, em particular as proteínas tirosinafosfatases, serão usados para ilustrar os conceitos discutidos¹¹.

DESCRIÇÃO MACROSCÓPICA

As Equações de Michaelis-Mentem^{12,13} são o modelo básico usado para descrever cinética enzimática para um único substrato, S:



$$K_S = \frac{[E][S]}{[E \cdot S]} = \frac{k_{-1}}{k_1} \quad (2)$$

$$K_M = \frac{k_{-1} + k_{cat}}{k_1} \quad (3)$$

onde E denota enzima, P é o produto, K_S é a constante de equilíbrio para dissociação do complexo E.S, k_{cat} é a constante de velocidade de reação observada e K_M é a constante de Michaelis. Na maioria dos casos, $k_{-1} \gg k_{cat}$, portanto, $K_S \approx K_M$. Neste modelo, a velocidade

*e-mail: garantes@iq.usp.br

de inicial de formação do produto, $v = k_{cat} [E.S]$, também pode ser escrita como:

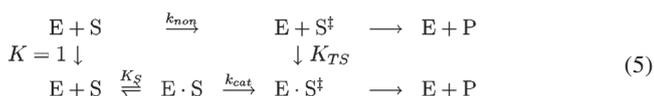
$$v = \frac{k_{cat}[E]_T[S]}{K_M + [S]} \quad (4)$$

onde definimos a concentração total de enzima, $[E]_T = [E] + [E.S]$. Equações semelhantes também são válidas para sistemas mais complexos como, por exemplo, na presença de dois ou mais substratos diferentes. Nestes casos, a definição da constante K_S é alterada.

Considerando este modelo, existem três maneiras da evolução maximizar a constante de velocidade de catálise aparente, definida como k_{cat}/K_M . Alteração somente de K_S para um determinado substrato pode amplificar a velocidade de catálise aparente pelo aumento da $[E.S]$, além de modificar a seletividade por este substrato. Um dos possíveis mecanismos empregados neste caso é o encaminhamento, em que o volume livre para difusão do substrato é reduzido ou por mudanças estruturais na enzima (como em complexos enzimáticos multi-unidades) ou por interações eletrostáticas que guiam o substrato para o sítio ativo¹⁴. A velocidade de catálise também pode ser amplificada pelo aumento da k_{cat} sem alteração da seletividade pelo substrato, ou seja, apenas aumentando a velocidade da etapa de transformação química na enzima. Finalmente, a velocidade de catálise pode ser aumentada pela modificação simultânea de k_{cat} e K_S . Não discutiremos aqui processos que alterem apenas a associação entre enzima e substrato (K_S). Este texto focaliza os processos que aumentam a velocidade de catálise principalmente pelo aumento de k_{cat} .

REAÇÃO DE REFERÊNCIA

Para identificação das fontes de aceleração enzimática, a reação catalisada deve ser comparada com uma reação de referência em solução aquosa sem catálise. Caso a reação enzimática e a reação observada em solução ocorrem segundo o mesmo mecanismo, ou seja, possuam um estado de transição semelhante, o seguinte ciclo termodinâmico pode ser usado para comparação:



A velocidade de reação na enzima será maior que a reação não catalisada em solução se $k_{cat}/K_S > k_{non}$. Neste caso, a afinidade da enzima pelo TS ou complexo ativado, S^\ddagger , é $K_{TS} = (k_{cat}/K_S)/k_{non} > 0$, ou seja, o complexo ativado é mais estável quando ligado à enzima que quando em solução.

No entanto, o mecanismo de reação atuante no sítio enzimático pode ser diferente do mecanismo de reação observado em solução aquosa na ausência do catalisador. Por exemplo, a hidrólise de ésteres de fosfato em solução aquosa ocorre diretamente num ataque nucleofílico por uma molécula de água. Porém, a mesma reação catalisada pelas proteínas tirosina-fosfatases (PTPs) ocorre em duas etapas: uma tiólise do éster de fosfato, seguida de uma hidrólise do intermediário tiosforilado (Figura 1)^{15,16}.

Nestes casos, o mecanismo de reação na enzima não deve ser

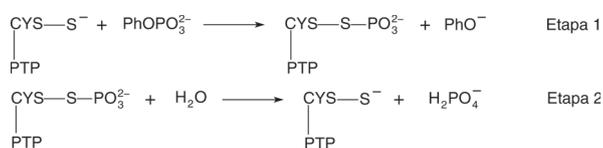


Figura 1. Esquema da reação de hidrólise catalisada pelas proteínas tirosina-fosfatases (PTP). Ph indica um grupo fenil e CYS, o resíduo de cisteína

diretamente comparado com o mecanismo observado em solução via um ciclo termodinâmico como o da Equação 5. A análise do efeito enzimático deve ser feita em comparação com uma reação de referência em solução aquosa (caracterizada por k_{non}), constituída pelos mesmos grupos reativos (R e S na Figura 2) e o mesmo mecanismo de reação na enzima, mas solvatados por água. (Figura 2)^{9,17}. Simulações computacionais podem ser usadas para calcular o perfil energético das reações de referência que são inacessíveis experimentalmente¹⁸⁻²⁰.

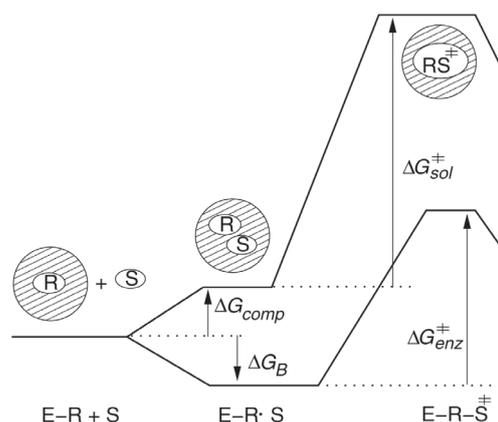


Figura 2. Esquema do perfil energético de uma reação enzimática e sua referência em solução. ΔG_B , ΔG_{comp} , ΔG_{enz}^\ddagger e ΔG_{sol}^\ddagger são, respectivamente, a energia livre de formação do complexo entre a enzima E-R e o substrato S, de complexação entre os reagentes R e S em solução, de ativação da reação no ambiente enzimático e de ativação da reação no complexo de encontro R.S dos reagentes em solução. A região com linhas diagonais indica o solvente aquoso

ESPÉCIE CATALISADA

Estabilização do TS ou desestabilização do reagente?

A hipótese da “chave e fechadura” proposta inicialmente por Fischer para racionalizar o poder catalítico das enzimas foi estendida pela proposta de Haldane², em que o substrato não “encaixa” exatamente na enzima, mas sofre uma distorção após a formação do complexo E.S, e pela proposta de Pauling⁴, em que a enzima se liga mais fortemente ao TS que ao substrato.

A hipótese de Pauling apóia-se no conceito de estado de transição³ e diz que este é mais estabilizado no sítio ativo da enzima em relação a mesma espécie numa reação de referência em solução aquosa. Esta hipótese é amplamente aceita como principal fonte de catálise e parece válida em todas enzimas já analisadas quantitativamente^{2,10,18,19,21,22}.

Centenas de estruturas tridimensionais determinadas para enzimas complexadas a análogos de TS apóiam a noção de complementaridade entre as estruturas do sítio ativo e do TS. Estes complexos são de grande utilidade prática no desenho de drogas que funcionam como inibidores enzimáticos competitivos¹³. No entanto, comprovar a hipótese de Pauling por experimentos não é diretamente possível, já que um TS é um objeto transiente. Simulações computacionais de reações enzimáticas mostram repetidamente^{8,21-24} que a interação eletrostática entre o ambiente enzimático e o TS é a maior responsável pela diminuição de ΔG_{enz}^\ddagger em relação a ΔG_{sol}^\ddagger (Figura 2), demonstrando que o TS é a principal espécie que sofre catálise ao longo da coordenada de reação na enzima.

A afirmação de “maior ligação ou afinidade da enzima pelo TS” é por vezes interpretada literalmente, como se o TS fosse atin-

gido em solução e, depois, ligado e estabilizado pela enzima. Esta interpretação é equivocada porque o TS tem um tempo de vida muito pequeno em solução, da ordem de 10^{-15} s e, portanto, não existe por um tempo suficiente para que difusão e ligação à enzima ocorram. Na interpretação correta, a enzima complexa o substrato no estado fundamental e a ativação ocorre a partir deste complexo. O TS formado possui melhores contatos com a enzima que o substrato, maximizando a estabilização apenas quando o complexo é ativado. Assim, o TS tem maior estabilidade no sítio ativo que em solução.

A proposta de Haldane equivale à desestabilização do estado reagente em relação à mesma espécie livre em solução. A desestabilização contribui para amplificar a velocidade da reação enzimática ao reduzir a barreira de ativação ($\Delta G_{\text{enz}}^{\ddagger}$). No perfil de energia livre de ligação e reação representado pela curva A na Figura 3, ΔG_B indica a energia livre de ligação do substrato intrínseca, ou seja, sem nenhuma desestabilização do estado reagente. Uma alteração estrutural da proteína anterior à ligação do substrato, por exemplo, em consequência da ligação de um regulador alostérico ou de uma mutação no sítio ativo, pode resultar em aumento da K_M (perfil B, Figura 3). Uma alteração estrutural expressiva resultaria em grande desestabilização do complexo E.S e $K_M > 1$ (perfil C, Figura 3). A desestabilização também pode ser consequência de uma mudança estrutural na enzima⁵ após a formação de E.S, gerando o complexo E.S' (perfil D, Figura 3).

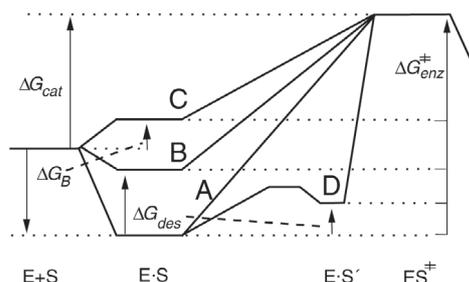


Figura 3. Esquema do perfil de energia livre de reações enzimáticas para complexos E.S com diferentes estabilidades. ΔG_B , ΔG_{des} , $\Delta G_{\text{enz}}^{\ddagger}$ e ΔG_{cat} são, respectivamente, a energia livre de formação do complexo E.S, de desestabilização de E.S, de ativação na enzima ($\Delta G_{\text{enz}}^{\ddagger} = -RT \ln[k_{\text{cat}}/A]$, onde A é o fator pré-exponencial) e de ativação aparente ($\Delta G_{\text{cat}} = -RT \ln[(k_{\text{cat}}/A)/K_M]$)

Jencks propôs que a energia de ligação do substrato observada nos processos representados pelos perfis B e D é $\Delta G_B - \Delta G_{\text{des}}$, ou seja, a diferença entre a energia de ligação intrínseca e a energia usada para desestabilizar o substrato. A desestabilização não (ou pouco) altera a energia do TS⁵ e ΔG_{des} seria usado como fonte de catálise, reduzindo a barreira para reação na enzima. Assim, a energia de ativação na enzima é modificada para cada perfil e $\Delta G_{\text{enz}}^{\ddagger}(A) > \Delta G_{\text{enz}}^{\ddagger}(D) > \Delta G_{\text{enz}}^{\ddagger}(B) > \Delta G_{\text{enz}}^{\ddagger}(C)$ (Figura 3). Esta proposta implica que enzimas têm alta afinidade intrínseca (ΔG_B) pelo substrato.

No entanto, a importância da desestabilização do substrato para catálise deve ser limitada, já que o aumento da energia do complexo E.S acima da energia de E+S (perfil C, Figura 3) não é catalítico. A velocidade máxima de uma reação reflete a diferença entre as energias livres do ponto mais alto (ES[‡]) e do mais baixo (E+S, no caso do perfil C) ao longo da coordenada de reação. A quantidade de energia livre disponível para catálise pela desestabilização do reagente (ΔG_{des}) também é pequena em comparação com a diferença entre as barreiras da reação catalisada e em solução ($\Delta G_{\text{sol}}^{\ddagger} - \Delta G_{\text{enz}}^{\ddagger}$, Figura 2)¹⁷.

Nenhum experimento quantificou diretamente o efeito de desestabilização para reações no sítio enzimático, mas Jencks e

Menger defendem esta proposta baseados principalmente na interpretação da reatividade de compostos orgânicos em solução^{5,25}.

A análise do efeito da evolução separadamente sobre k_{cat} e K_M de diversas enzimas dentro do modelo de Michaelis-Menten (Equação 1) com a relação k_{cat}/K_M fixada indica que enzimas evoluíram para ligar o TS fortemente (baixo $\Delta G_{\text{enz}}^{\ddagger}$ ou alto k_{cat})¹² e os substratos fracamente (alto K_M). Esta conclusão está em desacordo com a proposta de Jencks (vide supra). A desestabilização do complexo E.S (perfis B e D, Figura 3) não muda $\Delta G_{\text{cat}} = -RT \ln[(k_{\text{cat}}/A)/K_M]$, portanto não é claro como esta proposta pode ser usada pela evolução para variar ou maximizar a velocidade de catálise aparente (k_{cat}/K_M)¹⁷.

Enzimas mutantes cujo ΔG_{cat} é significativamente diferente da enzima selvagem podem ser divididas em três classes: mutantes que sofrem alteração da energia do TS, resultando em diminuição apenas de k_{cat} e manutenção de K_M ²⁶ (ou seja, $\Delta G_{\text{enz}}^{\ddagger}$ é aumentado, pois o TS é desestabilizado); mutantes que sofrem modificação de um sítio de ligação distante dos grupos reativos do substrato, resultando em aumento do K_M , mas mantendo o valor de k_{cat} (ou seja, o TS e o complexo E.S têm suas energias aumentadas pela mesma quantidade, logo, $\Delta G_{\text{enz}}^{\ddagger}$ não muda)¹⁷. Não foram encontrados²⁷ mutantes em que ambos K_M e k_{cat} seriam modificados (ou seja, $\Delta G_{\text{enz}}^{\ddagger}$ e ΔG_B apresentariam aumentos da mesma magnitude). Este seria o resultado esperado para uma mutação que anule ou diminua a capacidade de desestabilização do substrato.

A desestabilização dos reagentes é frequentemente associada à seleção pelo sítio ativo de uma conformação do substrato instável em relação à conformação mais populada em solução aquosa. A corismato mutase²⁸ e a lisozima são dois exemplos em que o substrato ligado à enzima é forçado a adotar uma conformação diferente daquela mais estável em solução aquosa. A diferença de energia livre entre os conformêros pode chegar a 5 kcal/mol, mas esta energia não é necessariamente usada na diminuição de $\Delta G_{\text{enz}}^{\ddagger}$ ¹⁸. A seleção de conformações pela enzima está relacionada ao conceito de conformações próximas ao ataque, discutido na próxima seção.

Conformação próxima ao ataque

Bruice²⁹ introduziu o conceito de conformação próxima ao ataque (NAC), definido como a configuração estrutural que os reagentes precisam assumir para chegar ao TS. A análise conformacional de uma série de reações intramoleculares de ciclização de ácidos dicarboxílicos indicou que a fração molar dos reagentes em NACs (definidas de acordo com critérios geométricos como distância e ângulo da ligação em formação) era diretamente proporcional à velocidade da ciclização intramolecular e formação de anidrido (Figura 4)²⁹. A distribuição de conformações dos substratos complexados aos sítios ativos em algumas reações enzimáticas que não envolvem a formação de intermediários covalentes (por exemplo, o rearranjo de corismato para perfenato, catalisado pela enzima corismato mutase) indicou que a proporção de NACs no sítio ativo enzimático era maior que na reação de referência em solução¹⁰.

No entanto, uma NAC não é uma espécie termodinamicamente estável e, portanto, sua definição não é única e depende dos critérios geométricos escolhidos. Assim, a presença de NACs não pode

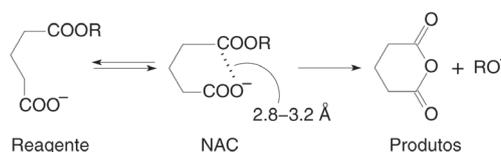


Figura 4. Esquema da reação de ciclização intramolecular passando por uma conformação próxima ao ataque (NAC), definida pela distância de 2,8 a 3,2 Å entre o oxigênio nucleofílico e o carbono esterificado

ser medida experimentalmente e sua contribuição energética para catálise é difícil de ser avaliada com precisão. Estimativas computacionais sugerem que contribuições entre 1 a 3 kcal/mol estão disponíveis para catálise através da estabilização ou seleção de NACs por enzimas, em relação à população normal de NACs observada em solução aquosa^{7,30}.

Também inspirado pela idéia de controle estereo-populacional em reações intramoleculares, Menger propôs sua hipótese espaço-temporal, postulando que a “velocidade da reação entre os grupos X e Y é proporcional ao tempo que X e Y residem a uma distância crítica”²⁵. Esta hipótese é equivalente ao aumento da fração molar de NACs.

Modelo do sítio dividido

Menger também propôs uma função catalítica para o complexo E.S°. No seu modelo de sítio dividido, a interação entre o substrato e a enzima pode ser dividida em sítios ou regiões de ligação (*lig*) e sítios ou regiões de reação (*reac*). A energia livre de cada espécie é descrita pela soma das energias de cada região, $\Delta G = \Delta G(\text{lig}) + \Delta G(\text{reac})$. As seguintes hipóteses estão no cerne do modelo:

1. a energia de formação do complexo E.S proveniente da região ligante, $\Delta G_B(\text{lig})$, é conservada quando o TS é alcançado, já o $\Delta G_B(\text{reac})$ não é conservado no TS, porque a estrutura molecular nesta região é modificada;

2. as forças atuantes são sempre estabilizantes (atrativas) na região de ligação, e desestabilizantes na região reativa.

Nesta proposta, a desestabilização da parte reativa é compensada pela estabilização na região ligante. Assim, a energia de ligação total (ΔG_B) é maior que num complexo E.S com apenas uma região (reativa) desestabilizada (como os exemplos da Figura 3), e a barreira de ativação para reação é reduzida. Este modelo é semelhante à proposta de ligação uniforme feita por Alberty e Knowles³¹.

No entanto, o ganho em energia livre neste modelo é limitado pela diferença entre a estabilização da região ligante e a desestabilização da região reativa. Estimativas computacionais¹⁷ e medidas cinéticas para catálise de substratos com diferentes regiões ligantes mas com a mesma região reativa¹² sugerem que aumentos de velocidade de apenas três ordens de magnitude podem ser obtidos pelo modelo de sítio dividido.

Por exemplo, este modelo pode explicar porque enzimas CDC25 da família das PTPs desfosforilam o substrato natural (a proteína CDK2) 10^3 vezes mais rápido que substratos artificiais que possuem apenas uma região reativa e nenhuma região de ligação³².

Interações e mecanismos catalíticos microscópicos

Nesta seção, revisaremos possíveis forças e interações microscópicas que podem explicar como enzimas estabilizam o TS e aumentam a velocidade das reações catalisadas. Existe uma dificuldade intrínseca em particionar estas contribuições energéticas, porque muitas delas não são aditivas e independentes³³. Por exemplo, experimentos de mutação sítio-dirigida podem avaliar a contribuição de um determinado resíduo para abaixamento da energia de ativação. No entanto, estes experimentos são sempre cooperativos, pois alteram uma série de interações, como eletrostática, impedimento estérico, dinâmica, liberdade configuracional, entre outras. Simulações computacionais são bastante poderosas para lidar com esta limitação, pois permitem que as contribuições sejam separadas e avaliadas independentemente.

Cabe aqui uma explicação sobre as contribuições discutidas a seguir e sua relação com os resultados obtidos em uma simulação. A maioria das simulações computacionais de reações enzimáticas

discutidas aqui são obtidas por potenciais híbridos de mecânica quântica e mecânica molecular^{24,28}.

Nestas simulações, a parte de mecânica quântica descreve a quebra e a formação de ligações químicas dos grupos reativos (átomos do substrato e do sítio ativo que participam da reação química) e, portanto, descreve as contribuições discutidas nas seções “Catálise covalente” e “Catálise ácido-base geral”.

A parte de mecânica molecular é responsável pela descrição das forças intermoleculares, como as forças eletrostáticas e de van der Waals, e das interações ligantes, como estiramento de ligações e torções de ângulos diedrais. As contribuições da “Interação eletrostática” estão diretamente ligadas à força eletrostática presente nesta parte do potencial. Da mesma forma, “Tensão e impedimento estérico” são descritos pelas interações ligantes e de van der Waals (ver abaixo). Contribuições termodinâmicas para catálise, como descrito na seção “Efeitos entrópicos”, são obtidas através da amostragem do espaço configuracional disponível ao sistema enzimático durante a simulação^{18,22}. Finalmente, “Efeitos quânticos e dinâmicos” são observados numa simulação ou através de correções à dinâmica clássica, ou explicitamente se a dinâmica da parte reativa for tratada quanto-mecanicamente⁸.

Catálise covalente

Catálise covalente pode ser descrita como uma modificação química transiente na enzima que ativa o substrato ou transfere um grupo reativo do substrato para outro aceptor. Para a catálise covalente ser efetiva, as seguintes condições devem ser atingidas⁵: o catalisador deve ter uma maior reatividade frente ao substrato que o aceptor final; o intermediário formado entre o catalisador e o substrato deve ser mais reativo que o substrato; o intermediário deve ser termodinamicamente instável (maior energia livre) em relação ao produto final, de maneira que o intermediário não acumule.

Catalisadores como proteases, esterasas, descarboxilases, fosfatases (incluindo as PTPs), entre outras enzimas, formam intermediários covalentes e parecem seguir as condições acima. Por exemplo, fosfatases que utilizam um resíduo de cisteína para ataque nucleofílico sobre fósforo formam um intermediário tiofosforilado¹⁵ (Figura 1). A energia de ativação para alcoólise³⁴ ou hidrólise³⁵ de tioésteres de fosfato (quebra da ligação P-S) é 6 e 10 kcal/mol mais baixa que a respectiva reação de oxidólise de fosfato (quebra da ligação P-O). Assim, a vantagem energética obtida pela catálise covalente é desta magnitude (5-10 kcal/mol).

Houk propôs recentemente³⁶ que a catálise covalente é um mecanismo fundamental para todas as enzimas de alta capacidade catalítica [com $(k_{cat}/K_M)/k_{non} > 10^{11}$]. Contudo, este argumento pode ser refutado já que algumas das enzimas de maior capacidade catalítica não apresentam intermediários covalentes⁷.

A formação de intermediários resulta na divisão do processo total em diversas etapas ou reações³⁷, portanto, o mecanismo de reação na enzima pode ser diferente do mecanismo de reação observado em solução. A análise do mecanismo catalítico deve ser feita entre cada etapa e a respectiva reação de referência não catalisada em solução (ver seção Reação de referência).

Catálise ácido-base geral

A catálise ácido-base geral, ou seja, doação (ou aceitação) de H⁺ por um grupo enzimático que funciona como um ácido (ou uma base), é uma das etapas envolvidas em diversas reações bioquímicas. A transferência pode ocorrer de três formas: um pré-equilíbrio rápido e anterior à etapa limitante da velocidade de reação; simul-

taneamente à reação ou, constituir a etapa limitante (por exemplo, na transferência de H^+ de carbonos alifáticos)⁵. No primeiro caso, as barreiras intrínsecas de transferência de H^+ são pequenas. Porém, a transferência pode formar intermediários de maior energia livre e, conseqüentemente, elevar a barreira total da reação. Estes intermediários são evitados nas reações enzimáticas⁵.

A transferência também pode neutralizar ou alterar a carga dos grupos reagentes, facilitando outras etapas do mecanismo de reação. Por exemplo, o ataque nucleofílico das PTPs sobre o substrato é facilitado pela doação de H^+ ao grupo de saída. Caso esta transferência de H^+ não aconteça (após uma mutação do resíduo que funciona como ácido geral), outras etapas do mecanismo de reação são energeticamente desestabilizadas ou até impedidas^{15,38}.

Avaliar a contribuição da catálise ácido-base geral para a diminuição da barreira de reação não é trivial, porque a transferência de H^+ pode influenciar diferentes etapas ou ocorrer em etapas não limitantes. Em solução aquosa, a catálise ácido-base geral propicia aumentos na constante de velocidade de cerca de 10^2 - 10^3 vezes⁵. Para as PTPs, cerca de 5 kcal/mol de estabilização do TS são atribuídos à catálise ácido-base geral, tanto por simulações computacionais¹⁵ como por mutações do resíduo envolvido na transferência de H^+ ³⁹.

Interação eletrostática

A importância das interações eletrostáticas foi rapidamente identificada para a ligação de substratos e para estabilidade e enovelamento de proteínas, através da formação de pares iônicos ou pontes salinas⁵. Porém, sua importância como mecanismo catalítico não foi comprovada até o advento de experimentos de mutação sítio-dirigida e de simulações computacionais^{26,40}.

Mutações em resíduos que estabilizam cargas formais ou parciais do TS resultam em diminuição de k_{cat} ⁴¹. Alterações no microambiente dielétrico de uma reação podem diminuir a altura da barreira em mais de 20 kcal/mol⁴². Assim, enzimas fornecem um microambiente altamente polar cujo potencial eletrostático complementa as mudanças de cargas observadas ao longo de uma reação¹⁸ (Figura 5).

A coordenação (não covalente) com íons metálicos e ligações de hidrogênio são normalmente consideradas interações eletrostáticas, pois são modeladas com razoável sucesso por esta força^{18,23}.

Simulações computacionais mostram repetidamente^{8,21,23,24} que a interação eletrostática entre o ambiente enzimático e o TS das reações catalisadas é a maior responsável pela diminuição de ΔG_{enz}^\ddagger em relação a ΔG_{sol}^\ddagger (Figura 2). Mesmo nos casos onde o TS é menos polar que o estado reagente, a estabilização eletrostática do TS na enzima é maior que em solução²⁰. O aumento médio de velocidade atribuído à interação eletrostática é calculado em 10^{11} vezes (equivalente a um abaixamento de 17 kcal/mol na energia de ativação)⁸.

Solventes polares, como a água, podem reorientar seus dipolos para estabilizar a distribuição de carga atingida no TS. Em uma enzima, no entanto, os dipolos que estabilizam o TS já estão *pré-orientados* no complexo E.S para estabilizar as cargas do TS. Em solução, a energia para *reorientar* os dipolos do solvente complementarmente às cargas do TS é elevada. Na enzima, a energia de orientação dos dipolos no sítio ativo já foi gasta durante o enovelamento e a formação da estrutura nativa. Portanto, enzimas são “supersolventes”, que funcionam por substituição de solvatação do TS, com pequena energia de reorganização¹⁸.

Uma proposta relacionada, em que enzimas propiciariam microambientes apolares que desestabilizam os reagentes por dessolvatação^{5,28} (vide supra), é baseada na observação que algumas reações em fase gasosa ou em solventes apolares ocorrem mais rapidamente que as mesmas reações em soluções polares⁴². Energias de

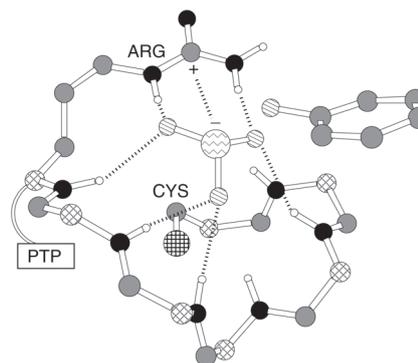


Figura 5. Exemplo de estabilização eletrostática do TS: estrutura do sítio ativo coordenado ao TS da primeira etapa de reação catalisada pelas PTPs (Figura 1). Interações eletrostáticas entre o grupo fosfato (PO_3^-) e a carga positiva da arginina (ARG) e as ligações de hidrogênio estão indicadas por linhas tracejadas. A estrutura foi obtida a partir de uma simulação de dinâmica molecular¹⁵ e parte dos átomos foram apagados para facilitar a visualização. Legenda para os tipos atômicos: H(branco), C(cinza), C_α (linhas diagonais cruzadas), N(preto), O(linhas diagonais), P(ondulado) e S(linhas cruzadas)

solvatação do reagente por água e pelo sítio enzimático (solvatação assume um sentido generalizado aqui) podem ser calculadas em comparação com a reação em fase gasosa, ou seja, na ausência de qualquer solvente. Por exemplo, a energia de solvatação calculada para o estado reagente na enzima haloalcano desalogenase é cerca de 30 kcal/mol menor que em solução aquosa e, portanto, supôs-se que a enzima desestabilizaria o substrato ao diminuir sua solvatação²⁸. Entretanto, a energia de pré-organização do sítio ativo é incorretamente descartada neste cálculo (ao contrário da reorganização do solvente aquoso)⁹. A hipótese de dessolvatação também é refutada se um ciclo termodinâmico (como a Equação 5) inclui a transferência dos fragmentos reativos polares do solvente aquoso (referência) para o possível microambiente apolar. Esta transferência envolve uma alta energia, que não está disponível no processo de catálise¹⁸. Sítios ativos são ambientes altamente polares e heterogêneos, bastante diferentes de solventes apolares ou da fase gasosa.

Efeitos entrópicos

O aumento da velocidade de reações por efeitos entrópicos é freqüentemente considerado um importante mecanismo catalítico^{5,24,43}. Esta proposta já foi enunciada de diversas maneiras, mas em essência implica que o espaço configuracional disponível para os reagentes em solução é restringido no sítio ativo enzimático, resultando em significativas diferenças entrópicas entre as energias livres de ativação na enzima e na reação de referência em solução.

A definição do estado de referência em solução é crucial para correta estimativa dos efeitos entrópicos. O efeito de concentração em solução aquosa, referente à aproximação dos reagentes, é dado por $\approx -RT \ln 55$ para cada par de reagentes. Assim para uma reação bimolecular R+S em solução (Figura 2), esta contribuição (ΔG_{comp}) é 2.4 kcal/mol a 300 K. Descontando o efeito de concentração, a diferença entre as barreiras entrópicas (ΔS) para ativação da reação em solução e na enzima pode ser escrita por:

$$\begin{aligned} (\Delta \Delta S^\ddagger)_{sol \rightarrow enz} &= [\Delta S_{enz}^\ddagger - \Delta S_{sol}^\ddagger] \\ &= [(S_{enz}^{TS} - S_{enz}^{E,S}) - (S_{sol}^{TS} - S_{sol}^{R,S})] \\ &= [(S_{enz}^{TS} - S_{sol}^{TS}) - \Delta S_B] \end{aligned} \quad (6)$$

onde $\Delta S_B = S_{enz}^{E,S} - S_{sol}^{R,S}$ é a entropia de ligação do substrato, considerando os estados E.S e R.S (Figura 2).

As propostas que sustentam o efeito entrópico como mecanismo catalítico assumem *implicitamente* que os graus de liberdade do TS em solução e na enzima estão congelados, ou seja $S_{enz}^{TS} = S_{sol}^{TS} \approx 0$ e, portanto, superestimam $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz} = -\Delta S_B^{19,24,41}$. Esta suposição é baseada num raciocínio para reações bimoleculares utilizando reagentes *esféricos*⁵ em que até -35 unidades entrópicas (cerca de 10 kcal/mol a 300 K) seriam obtidas pela restrição translacional no sítio ativo. Porém, este raciocínio não é válido para estruturas *moleculares*, porque parte da movimentação livre nos complexos R.S e E.S também está livre no TS (parte dos graus de liberdade translacionais e rotacionais tidos como “congelados” são, na verdade, transformados em vibrações de baixa frequência dos TS, cuja contribuição entrópica é bastante apreciável)^{18,41,44}.

Simulações computacionais das contribuições entrópicas ainda sofrem de problemas de convergência e, portanto, as estimativas não são totalmente seguras⁴⁵, mas simulações semiquantitativas^{18,45} indicam que $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz} < 1$ kcal/mol.

Nenhum experimento demonstrou diretamente que a catálise enzimática estivesse associada à $(\Delta\Delta S^\ddagger)_{sol \rightarrow enz}$. A análise das contribuições para energia livre de reações enzimáticas e de sua referência em solução determinadas por gráficos de Arrhenius^{10,45} indica que a contribuição entrópica é equivalente nos dois ambientes. Dados experimentais obtidos em solução⁴⁶ sugerem que a água já impõe restrições à movimentação translacional e rotacional de solutos polares e, portanto, o ganho em termos destas contribuições pela transferência dos reagentes para o sítio ativo enzimático seria pequeno.

Os aumentos de velocidade observados em reações intramoleculares (ver seção Espécie catalisada), em que os reagentes estão covalentemente ligados e parcialmente imobilizados, são citados como evidência de catálise por efeitos entrópicos. Porém, a informação obtida destes modelos não é diretamente transferível para reações enzimáticas, porque a restrição das configurações disponíveis na reação intramolecular não está necessariamente ligada às possíveis restrições observadas no sítio ativo, em comparação com a reação intermolecular de referência em solução¹⁸. Simulações computacionais também indicam que o aumento da velocidade de reações intramoleculares não está necessariamente associado a contribuições entrópicas²⁹.

Portanto, embora efeitos entrópicos para catálise não possam ser totalmente ignorados, a magnitude destas contribuições é menor que o tradicionalmente proposto⁵.

Tensão e impedimento estérico

A tensão ou impedimento estérico resultante da aproximação espacial entre moléculas ou partes da mesma molécula foram sugeridos como um possível mecanismo catalítico ligado à desestabilização do substrato (ver seção Espécie catalisada)⁵.

Existem dois casos limites nessa hipótese. Se o sítio ativo for relativamente rígido e sua estrutura complementar ao TS, o reagente sofreria uma distorção geométrica para se acomodar no sítio ativo⁵, diminuindo a energia de ativação como mostrado no perfil B, Figura 3. Caso o sítio ativo seja flexível e o substrato rígido, a formação do complexo E.S seria acompanhada por uma mudança conformacional na enzima que tensionaria e desestabilizaria a geometria do substrato, aumentando a energia livre do sistema (perfil D, Figura 3)⁵.

Nenhuma observação experimental suporta diretamente ambas as propostas⁵ e as evidências existentes são consideradas ambíguas¹². A catálise na lisozima é tradicionalmente citada como evidência da tensão que pode ser exercida sobre o substrato, pois a estrutura do sítio ativo parece complementar à conformação “barco” do substrato (um sacarídeo, cuja conformação “cadeira” é cerca de 5 kcal/mol

mais estável em solução aquosa). No entanto, a diferença de energia livre associada à mudança da conformação cadeira para conformação sofá entre a solução aquosa e o ambiente enzimático é menor que 1 kcal/mol¹⁸. Portanto, esta diferença não é suficiente para explicar o abaixamento da energia de ativação no ambiente enzimático.

Medidas de RMN e simulações computacionais mostram que enzimas são moléculas bastante flexíveis, que podem acomodar mudanças na estrutura do substrato sem aumento significativo da energia livre^{8,37}. Mudanças geométricas do reagente associadas à formação do TS são menores que 1 Å, e deformações desta magnitude são facilmente acomodadas pelos modos vibracionais de baixa energia da proteína (que, pictoricamente, pode ser comparada a um conjunto de molas com baixa constante de força)¹⁸.

Numa simulação computacional, o impedimento estérico pode ser definido como o potencial repulsivo das interações de van der Waals e as contribuições da distorção de ligações, ângulos e diedrais. A contribuição total destas interações para diminuição de ΔG_{enz}^\ddagger foi calculada em menos que 2 kcal/mol para diferentes enzimas^{18,30}.

A proposta de “ajuste induzido”³⁵ é semelhante mas não está relacionada à catálise, porque a energia de ligação é utilizada na modificação da estrutura enzimática para regular a especificidade pelo substrato, ao invés de diminuir ΔG_{enz}^\ddagger .

Efeitos quânticos e dinâmicos

A dinâmica enzimática também foi proposta recentemente como um possível mecanismo catalítico^{2,8}.

Movimentações de larga escala na estrutura enzimática como torção de sub-unidades, modos vibracionais de baixa frequência e reposicionamento de grupos protéicos causados, por exemplo, pela ligação do substrato ou de um regulador alostérico, são alterações conformacionais na enzima responsáveis pelo correto posicionamento dos grupos catalíticos. Se a amostragem conformacional é feita corretamente durante as simulações computacionais, estes efeitos, chamados de reorganização da estrutura enzimática²³, são incorporados normalmente^{18,21,24} e não representam um mecanismo catalítico em particular^{8,41}.

A liberdade conformacional não deve ser confundida com as flutuações dinâmicas da estrutura protéica, principalmente as vibrações que amplificam efeitos quânticos como tunelamento^{47,48}. A teoria generalizada do estado de transição^{3,49} pode ser usada para racionalizar estes efeitos. Segundo a Equação:

$$k(T) = \tau(T) (k_B T/h) \exp [-\Delta G^\ddagger/RT] \quad (7)$$

onde R , h e k_B são constantes, e T é a temperatura, a constante de velocidade da reação, $k(T)$, pode ser elevada pelo aumento do coeficiente de transmissão, $\tau(T)$, ou pela diminuição da barreira de reação, ΔG^\ddagger .

As mudanças conformacionais descritas acima alteram apenas ΔG^\ddagger . Já o fator de transmissão pode ser alterado pela dinâmica do sistema reativo, onde $\tau(T) \leq 1$, ou seja, uma trajetória de reação pode ultrapassar o TS, mas ser improdutiva e retornar aos reagentes.

Além de um recruzamento não-reativo clássico, $\tau(T)$ pode ser alterado por efeitos de solvatação de não-equilíbrio (por exemplo, proveniente de uma distribuição configuracional dos reagentes fora do equilíbrio) e por efeitos quânticos como tunelamento^{2,8,41,43,50}. Os efeitos de recruzamento e de não-equilíbrio podem apenas tornar $\tau(T)$ menor que um e, usualmente, já são pequenos em solução aquosa. Neste meio, $\tau(T)$ é quase unitário². Assim, pouca ou nenhuma vantagem catalítica pode ser explorada por enzimas buscando minimizar estes efeitos, de acordo com resultados observados em simulações computacionais^{2,8}.

Já o tunelamento pode fenomenologicamente resultar em $\tau(T) > 1$, ou seja, pode amplificar a constante de velocidade sem alterar o termo exponencial na Equação 7. Contribuições de até 3 ordens de magnitude para o aumento da velocidade de reações enzimáticas foram atribuídas à dinâmica da proteína, ou seja, às vibrações que amplificam o tunelamento em comparação com a reação em solução, nos sistemas em que uma transferência de hidrogênio (tanto nas formas H^+ , H neutro e H^-) é determinante da velocidade de reação^{8,43,50}. As desidrogenases de álcool⁴⁷ e de lactato⁴⁸ são exemplos de sistemas em que este mecanismo catalítico é observado.

CONCLUSÃO

Diversas teorias e hipóteses já foram propostas para explicar porque enzimas são catalisadores de extrema eficiência. Neste texto apresentamos uma perspectiva computacional destas propostas, principalmente daquelas que foram testadas quantitativamente e comparadas com experimentos.

A utilização de ciclos termodinâmicos e de corretas reações de referência é essencial para racionalização das interações e dos mecanismos catalíticos. Sem dúvida, cada enzima utiliza uma combinação das forças e interações revisadas aqui, dependendo da reação química catalisada. Contudo, a espécie que sofre influência determinante do ambiente enzimático é o estado de transição. Este é estabilizado principalmente por interações eletrostáticas geradas pelo sítio ativo e pela estrutura enzimática pré-organizados. Em média, reduções de cerca de 17 kcal/mol na energia de ativação podem ser obtidas por este mecanismo catalítico.

O aumento da fração molar de conformações próxima ao ataque (NACs) e a desestabilização do substrato contribuem em escala bem menor (no máximo 5 kcal/mol) para reduções na energia de ativação enzimática. A desestabilização é viável principalmente pelo modelo do sítio dividido. A catálise covalente, a catálise ácido-base geral e efeitos quânticos e dinâmicos também podem contribuir na mesma magnitude (5 kcal/mol cada) para o aumento da velocidade das reações enzimáticas em que intermediários são formados ou hidrogênio seja transferido, respectivamente.

Já efeitos entrópicos e impedimento estérico, antes postulados como importantes mecanismos catalíticos, contribuem, na verdade, com menos que 3 kcal/mol para reduções na energia de ativação enzimática e têm pouca influência sobre os aumentos de velocidade observados.

Simulações computacionais são, portanto, ferramentas bastante poderosas na investigação da atividade enzimática e na elucidação do poder catalítico destas macromoléculas biológicas.

AGRADECIMENTOS

O autor gostaria de agradecer as bolsas financiadas pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), nos projetos 99/07688-9 e 2005/03995-7.

REFERÊNCIAS

- Lad, C.; Williams, N. H.; Wolfenden, R.; *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5607.
- Benkovic, S. J.; Hammes-Schiffer, S.; *Science* **2003**, *301*, 1196.
- Billing, G. D.; Mikkelsen, V.; *Introduction to Molecular Dynamics and Chemical Kinetics*, Wiley: New York, 1st ed., 1996.
- Pauling, L.; *Nature* **1948**, *161*, 707.
- Jencks, W. P.; *Catalysis in Chemistry and Enzymology*, Dover: New York, 1st ed., 1987.
- Menger, F. M.; *Biochemistry* **1992**, *31*, 5368.
- Bruice, T. C.; *Chem. Rev.* **2006**, *106*, 3119.
- Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G.; *Science* **2004**, *303*, 186.
- Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M.; *Chem. Rev.* **2006**, *106*, 3210.
- Bruice, T. C.; *Acc. Chem. Res.* **2002**, *35*, 139.
- Arantes, G. M.; *Tese de Doutorado*, Universidade de São Paulo, Brasil, 2004.
- Fersht, A.; *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman: New York, 1st ed., 1999.
- Voet, D.; Voet, J.; *Biochemistry*, Wiley: New York, 2nd ed., 1995.
- Raushel, F. M.; Thoden, J. B.; Holden, H. M.; *Acc. Chem. Res.* **2003**, *36*, 539.
- Arantes, G. M.; *Biochem. J.* **2006**, *399*, 343.
- Arantes, G. M.; Chaimovich, H.; *J. Phys. Chem. A* **2005**, *109*, 5625.
- Warshel, A.; Florián, J.; *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5950.
- Warshel, A.; *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley: New York, 1st ed., 1991.
- Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D.; *Acc. Chem. Res.* **2001**, *34*, 72.
- Roca, M.; Martí, S.; Andrés, J.; Moliner, V.; Tunón, I.; Bertrán, J.; Williams, I. H.; *J. Am. Chem. Soc.* **2003**, *125*, 7726.
- Field, M. J.; *J. Comp. Chem.* **2002**, *23*, 48.
- Dwyer, M. A.; Looger, L. L.; Hellinga, H. W.; *Science* **2004**, *304*, 1967.
- Kollman, P. A.; Kuhn, B.; Perakyla, M.; *J. Phys. Chem. B* **2002**, *106*, 1537.
- Åqvist, J.; Warshel, A.; *Chem. Rev.* **1993**, *93*, 2523.
- Menger, F. M.; *Acc. Chem. Res.* **1985**, *18*, 128.
- Leatherbarrow, R. J.; Fersht, A. R.; Winter, G.; *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7840.
- Warshel, A.; *J. Biol. Chem.* **1998**, *273*, 27035.
- Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G.; *Chem. Rev.* **2006**, *106*, 3188.
- Bruice, T. C.; Lightstone, F. C.; *Acc. Chem. Res.* **1999**, *32*, 127.
- Kraut, A.; Strajbl, M.; Villá, J.; Warshel, A.; *J. Am. Chem. Soc.* **2002**, *124*, 4097.
- Albery, W. J.; Knowles, J. R.; *Biochemistry* **1976**, *15*, 5631.
- Rudolph, J.; *Biochemistry* **2002**, *41*, 14613.
- Kraut, D. A.; Carroll, K. S.; Herschlag, D.; *Annu. Rev. Biochem.* **2003**, *72*, 517.
- Gregersen, B. A.; Lopez, X.; York, D. M.; *J. Am. Chem. Soc.* **2003**, *125*, 7178.
- Thain, E. M.; *J. Chem. Soc.* **1957**, 4694.
- Zhang, X.; Houk, K. N.; *Acc. Chem. Res.* **2005**, *38*, 379.
- Hammes, G. G.; *Biochemistry* **2002**, *41*, 8221.
- Arantes, G. M.; Loos, M.; *Phys. Chem. Chem. Phys.* **2006**, *8*, 347.
- Jackson, M. D.; Denu, J. M.; *Chem. Rev.* **2001**, *101*, 2313.
- Gilson, M. K.; Honig, B. H.; *Nature* **1987**, *330*, 84.
- Villá, J.; Warshel, A.; *J. Phys. Chem. B* **2001**, *105*, 7887.
- Menegon, G.; Loos, M.; Chaimovich, H.; *J. Phys. Chem. A* **2002**, *106*, 9078.
- Kohen, A.; Klinman, J. P.; *Acc. Chem. Res.* **1998**, *31*, 397.
- Ayala, P.; Schlegel, H. B.; *J. Chem. Phys.* **1998**, *108*, 2314.
- Villá, J.; Strajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A.; *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11899.
- Yu, Y. B.; *J. Phys. Chem. B* **2003**, *107*, 1721.
- Nagel, Z. D.; Klinman, J. P.; *Chem. Rev.* **2006**, *106*, 3095.
- Antoniu, D.; Basner, J.; Nunez, S.; Schwartz, S. D.; *Chem. Rev.* **2006**, *106*, 3170.
- Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J.; *J. Phys. Chem.* **1996**, *100*, 12771.
- Bala, P.; Grochowski, P.; Nowinski, K.; Lesyng, B.; McCammon, J. A.; *Biophys. J.* **2000**, *79*, 1253.

ANEXO B – The catalytic acid in the dephosphorylation of the Cdk2-pTpY/CycA protein complex by Cdc25B phosphatase

The Catalytic Acid in the Dephosphorylation of the Cdk2-pTpY/CycA Protein Complex by Cdc25B Phosphatase

Guilherme Menegon Arantes*

Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900 São Paulo, SP, Brasil

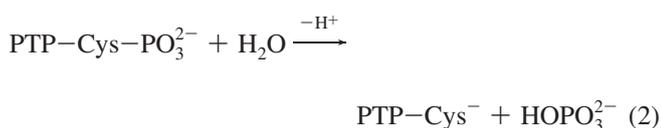
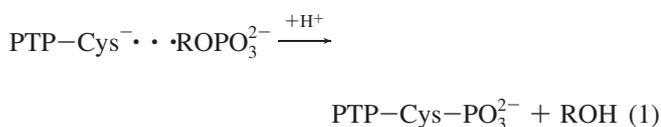
Received: August 5, 2008; Revised Manuscript Received: September 16, 2008

The development of anticancer therapeutics that target Cdc25 phosphatases is now an active area of research. A complete understanding of the Cdc25 catalytic mechanism would certainly allow a more rational inhibitor design. However, the identity of the catalytic acid used by Cdc25 has been debated and not established unambiguously. Results of molecular dynamics simulations with a calibrated hybrid potential for the first reaction step catalyzed by Cdc25B in complex with its natural substrate, the Cdk2-pTpY/CycA protein complex, are presented here. The calculated reaction free-energy profiles are in very good agreement with experimental measurements and are used to discern between different proposals for the general acid. In addition, the simulations give useful insight on interactions that can be explored for the design of inhibitors specific to Cdc25.

1. Introduction

Cdc25 phosphatases are crucial regulators of the eukaryotic cell cycle because they activate cyclin-dependent kinases (Cdk) responsible for cycle check points. Overexpression of Cdc25 phosphatases has been linked to diverse cancers.¹ Hence, these enzymes are now attractive targets for antineoplastic agents.^{2–4} Because most drugs that target enzymes are competitive or irreversible inhibitors of active sites,⁵ it is important that drug designers know the catalytic residues and understand the catalytic mechanism of the target enzyme.

Cdc25 phosphatases are protein tyrosine phosphatases (PTPs) of dual specificity.⁶ They hydrolyze phosphotyrosine, phosphoserine, and phosphothreonine in two consecutive reactions. The first step is the nucleophilic attack from a PTP cysteine side chain toward the phosphate ester substrate, with a possible H⁺ transfer from a general acid to the leaving group. A thiophosphorylated PTP intermediate is formed, and the substrate is dephosphorylated (eq 1). The PTP intermediate is hydrolyzed, and free enzyme is regenerated in the second step (eq 2).^{6–8}



Almost all PTPs use as a general acid, an Asp residue located in a mobile loop (the WPD loop) at a distance from the active site.⁷ Cdc25s are an exception to this observation. Structural studies have demonstrated that Cdc25s lack an Asp residue near the active site or in a flexible loop comparable to the WPD

loop in the other PTPs that could donate a H⁺ to the leaving group.^{9,10} Hence, the identity of the general acid on Cdc25s remains elusive. It has been suggested that the catalytic acid is a Glu residue (Glu474 on Cdc25B) located in the active site just after the catalytic cysteine (Cys473 on Cdc25B).^{9,11} This Glu residue is conserved among all known Cdc25 isoforms.⁷ Another proposal suggested that the substrate could bind as a phosphate monoanion and transfer the proton itself to the leaving group.¹² If well characterized, the difference in the catalytic acid between Cdc25 phosphatases and other PTPs could be explored in the design of inhibitors specific to Cdc25s.

In the following work, both proposals for the catalytic acid were tested via molecular dynamics simulations for the first reaction step (eq 1) of Cdc25B docked with its natural substrate, the protein complex Cdk2-pTpY/CycA, cyclin-dependent kinase 2 bis-phosphorylated and complexed with cyclin A.¹³ This reaction corresponds to dephosphorylation of pThr14 on Cdk2 via a nucleophilic attack by Cys473 on Cdc25B. The next section gives details about the modeled structures and potentials used in the simulations. Calculated free-energy profiles shown in the Results and Discussion section are compared to the experimental rate constant and used to discern which proposal for the general acid is energetically feasible. In addition, interactions containing Glu474 are suggested as a new target for the design of inhibitors specific to Cdc25s.

2. Materials and Methods

Initial coordinates of a Cdc25B/Cdk2-pTpY/CycA model complex were kindly provided by J. Rudolph (Duke University). This model has been extensively validated via kinetic, mutagenic, X-ray crystallographic, and calorimetric studies.¹³ The protein complex was solvated in a cubic water box of side 108.8 Å. The resulting system has a total of 129793 atoms, including 39263 water molecules, 4 Na⁺ to neutralize the total charge, and 735 amino acids from the three proteins. This is a very large model structure as well as the first computer simulation of a phosphatase reaction with its natural substrate. The protonation state of each residue was the same as that adopted

* E-mail: garantes@iq.usp.br.

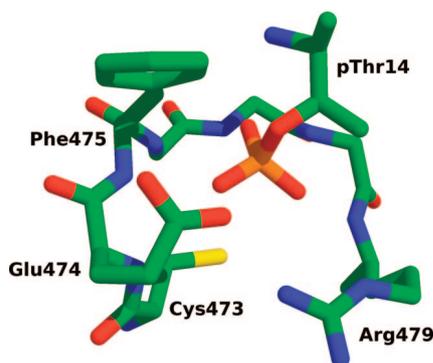


Figure 1. Snapshot of the simulated Michaelis complex. Only part of the P-loop heavy atoms are shown. Some side chains and parts of the backbone were erased to ease visualization.

by Rudolph et al.,¹³ but the side chain of Glu474 was protonated, Ser473 was changed to Cys, and its side chain was ionized. This system was equilibrated for 1 ns of molecular dynamics simulation¹⁴ at 300K and 1 atm using the Berendsen temperature and pressure couplings.¹⁵ The protein complex and the solvent were represented with the OPLS-AA¹⁶ and TIP-3P¹⁷ force fields, respectively. The LINCS algorithm was used to constrain hydrogen bonds, and a time step of 2 fs was used. Long-range electrostatics were truncated with the force-switching method starting at $r_{\text{on}} = 10 \text{ \AA}$ until $r_{\text{off}} = 13 \text{ \AA}$.¹⁵

During this initial equilibration phase, the distance between the pThr14 O_{γ} on the Cdk2 leaving group and the acidic hydrogen from Glu474 on Cdc25B, $d(\text{OH})$, was noted to be too large for H^+ transfer, $d(\text{OH}) \approx 7 \text{ \AA}$. The free energy associated with decreasing such a distance was determined. An umbrella potential was used to guide the $d(\text{OH})$ coordinate toward lower values in a series of short simulations. The effect of such a potential was removed, and the short simulations were pieced together using the weighted histogram analysis method (WHAM).¹⁸ Reference values ξ_i were equally spaced at 0.10 \AA , and a constant $k_{\text{umb}} = 1500 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ was used. Molecular dynamics runs with at least 5 ps of equilibration and 20 ps of data collection were used for each reference window ξ_i . Further details are the same as those previously reported.^{8,19}

Another 5 ns of molecular dynamics were run for equilibration of the solvated Michaelis complex. Harmonic potentials were used to tether the pThr14 phosphorus to the Cys473 sulfur and the pThr14 O_{γ} oxygen to the Glu474 acidic hydrogen. Both harmonic potentials had minima set to 4.0 \AA with $k_f = 50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. The final configuration has a root-mean-squared deviation (rmsd) of 0.8 \AA for all protein C_{α} atoms in comparison with the initial structure provided by Rudolph.¹³ This final equilibrated structure (Figure 1) represents a better model to study the possible phosphate and H^+ transfers catalyzed by Cdc25B in complex with Cdk2-pTpY/CycA.

Free-energy profiles for the chemical reactions were obtained with a hybrid potential of quantum mechanics and molecular mechanics (QM/MM)²⁰ specifically parametrized for phosphatases.²¹ This potential has already been tested extensively and shown to provide excellent agreement with experiment for another dual-specificity phosphatase, VHR, in wild-type and mutated forms, reacting either with aryl or alkyl small-molecule substrates.^{8,22} In the hybrid potential simulations, the side chains of Cys473, Glu474, and pThr14 were represented in the QM region with the parametrized QM/MM potential.²¹ The remainder of the protein complex and the solvent were represented with the OPLS-AA and TIP-3P force fields, respectively. A combination of an umbrella potential and the WHAM method was used

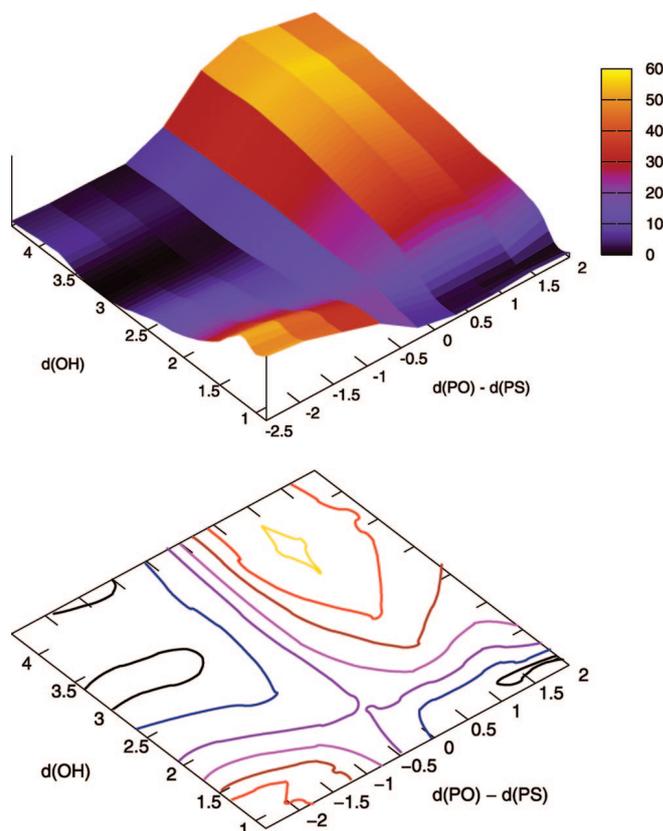


Figure 2. Two-dimensional free-energy profile (color coded in kcal/mol) for dephosphorylation of the Cdk2-pTpY/CycA complex by Cdc25B. The phosphate in pThr14 was modeled as a dianion. H^+ transfer is represented by $d(\text{OH})$, the distance between the pThr14 O_{γ} and the acidic hydrogen from Glu474. Phosphate transfer is described by $d(\text{PO}) - d(\text{PS})$, the difference between forming P-S and breaking P-O bonds to phosphorus. In the Michaelis complex, $d(\text{PO}) - d(\text{PS}) < -1.5$. In the product, $d(\text{PO}) - d(\text{PS}) > +1.0$. The upper panel is a 3D projection. The lower panel is a contour plot for the same data with discrete energy levels of 2, 7, 17, 27, 37, 47, and 57 kcal/mol, following the color code given in the upper panel.

to obtain the reaction free-energy profiles as described above. A velocity Verlet-Langevin scheme was used to integrate the equations of motion²³ during the simulation of reaction profiles. A time step of 1 fs and a friction coefficient of $\gamma = 10 \text{ ps}^{-1}$ were used. Additional details are the same as those used in previous phosphatase simulations.^{8,21}

Equilibration molecular dynamics were carried out with the GROMACS program (version 3.3.2).¹⁵ Computations of free-energy profiles were done with the DYNAMO library.²⁴

3. Results and Discussion

In the simulated Michaelis complex shown in Figure 1, the dianionic phosphate group of pThr14 is coordinated to the Cdc25B P-loop. The ionized phosphate equatorial oxygens are hydrogen bonded with the P-loop backbone and with the Arg479 side chain on Cdc25B. Such an arrangement is equivalent to the one observed for reaction of VHR phosphatase with small-molecule substrates.⁸ The pThr14 O_{γ} forms a hydrogen bond to the side chain of Arg36 on Cdk2. This last side chain together with the side chain in Phe475 on Cdc25B creates a pathway for proton transfer from Glu474. The average distance between Arg36 and Phe475 side chains observed during the last 1 ns of molecular dynamics in the equilibration phase is 4.6 \AA . It is enough space for the transferred H^+ to travel from Glu474 to pThr14. Glu474 O_{ϵ} also forms a hydrogen bond to the Arg36

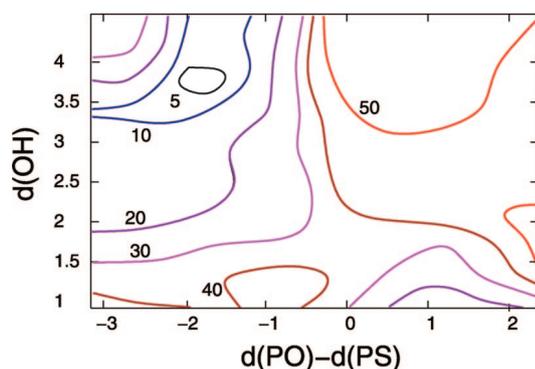


Figure 3. Two-dimensional free-energy contour plot (color coded in kcal/mol) for dephosphorylation of the Cdk2-pTpY/CycA complex catalyzed by Cdc25B with the phosphate group in pThr14 modeled as a monoanion. The $d(\text{PO}) - d(\text{PS})$ is defined in Figure 2, and $d(\text{OH})$ is the distance between the pThr14 O_γ and the phosphate hydrogen.

side chain. Hence, the Arg36 side chain on Cdk2 is buried in the Cdc25B active site. Site-directed mutagenesis of this residue has not decreased phosphatase activity nor changed the k_{cat}/K_M versus pH profiles.¹³ Thus, Arg36 on Cdk2 might not be important for the transition-state stabilization, but it seems relevant for complexation between Cdk2 and Cdc25B.

Figure 2 describes the energetics of two reaction coordinates representing H^+ transfer and phosphate transfer. In the Michaelis complex, $d(\text{OH}) < 3.0 \text{ \AA}$ and $d(\text{PO}) - d(\text{PS}) < -1.5$. In this region, the free-energy surface is flat, indicating that there is some conformational freedom for Glu474 and pThr14 side chains. Phosphate transfer from pThr14 would not occur without proton transfer to the leaving group. Figure 2 shows that barriers as high as 60 kcal/mol would have to be traversed if $d(\text{OH}) > 2 \text{ \AA}$. On the other hand, proton transfer to pThr14 O_γ would not occur if the P–O bond in pThr14 was preserved, again because barriers are as high as 60 kcal/mol. Hence, both reactions are concerted. In the transition-state region, $d(\text{OH}) \approx 1.5 \text{ \AA}$ and $d(\text{PO}) - d(\text{PS}) \approx -0.3 \text{ \AA}$. In the product region, the phosphate group is transferred to Cys473, Cdc25B is thiophosphorylated, and Cdk2 is dephosphorylated. The calculated free-energy barrier is 17 kcal/mol. This value is identical to the barrier determined experimentally.²⁵ In fact, H^+ transfer is well-advanced, but not complete, in comparison to phosphate transfer in the transition state. This is in line with the interpretation of kinetic isotope effects measured for the dephosphorylation of an alkyl phosphate substrate by Cdc25A.²⁶

The second proposal for general acid catalysis, namely, that pThr14 could bind as a phosphate monoanion and transfer the proton itself to the leaving group was tested on Figure 3. In the reactant state corresponding to the Michaelis complex, one of the equatorial phosphate oxygens in pThr14 is protonated, and the extra proton is coordinated by the Cys473 thiolate. The other interactions in the active site are similar to those described above for the dianion reaction. Phosphate and H^+ transfer are concerted, with the transition state located in the region with $d(\text{PO}) - d(\text{PS}) \approx -0.2 \text{ \AA}$ and $d(\text{OH}) \approx 1.7 \text{ \AA}$. The calculated free-energy barrier is 38 kcal/mol, indicating that this proposal is energetically unfeasible and that no reaction would be experimentally observed if the pThr14 phosphate group was protonated. Reaction barriers of similar magnitudes were also calculated for reaction of small-molecule substrates in monoanionic (monoprotonated) form with the VHR phosphatase.⁸

The simulation method used and the reaction barriers calculated here have a precision of ± 5 kcal/mol or better, as shown for another PTP.^{8,21} Therefore and also because of the

TABLE 1: Calculated Geometrical Parameters for Cdk2-pTpY/CycA Complexed with Cdc25B^a

	initial structure ¹³	Michaelis complex	
Glu474 Dihedrals			
$\text{HO}-\text{C}_\delta-\text{C}_\gamma-\text{C}_\beta$	173	50	
$\text{C}_\delta-\text{C}_\gamma-\text{C}_\beta-\text{C}_\alpha$	-167	53	
$\text{C}_\gamma-\text{C}_\beta-\text{C}_\alpha-\text{N}$	-178	43	
	initial structure ¹³	Michaelis complex	
Distance Between			
HO (Glu474) ^b	O_γ (pThr14) ^c	7.2	4.1
O_γ (pThr14) ^c	H_ω (Arg36) ^c	2.4	2.5
O_ϵ (Glu474) ^b	$\text{H}_{\{\omega'\}}$ (Arg36) ^c	2.6	3.0
O_ϵ (Glu474) ^b	H_ω (Arg427) ^b	4.4	6.0
O_ϵ (Glu474) ^b	H^4 (Tyr528) ^b	1.6	3.4
O_ϵ (Glu474) ^b	C^4 (Phe475) ^b	4.2	3.2
C_δ (Glu474) ^b	C^4 (Phe475) ^b	4.5	4.2

^a Distances given in \AA , and angles given in degrees. ^b On Cdc25B. ^c On Cdk2.

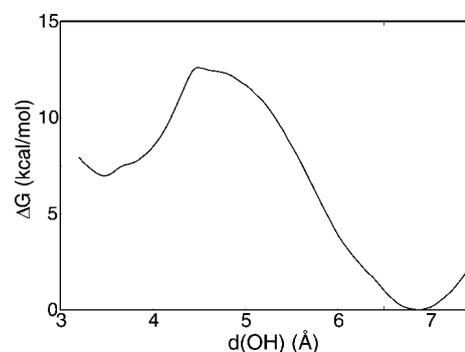


Figure 4. Free-energy profile (kcal/mol) for the conformational transition that decreases the distance between the Cdk2 pThr14 O_γ and the acidic hydrogen from Glu474 on Cdc25B, $d(\text{OH})$.

agreement between the calculated barrier for reaction of pThr14 in dianionic form with experimental measurements (see above), it can be concluded that Glu474 on Cdc25B may well function as the catalytic general acid in dephosphorylation of its natural substrate, pThr14 on the Cdk2-pTpY/CycA complex.

As mentioned in the Materials and Methods section, the distance between pThr14 O_γ and the acidic hydrogen from Glu474 observed in the initial structure¹³ was too large for H^+ transfer (see Table 1). It is shown in Figure 4 that the conformational transition associated with this distance reduction is accessible. Simulations show that this transition corresponds to torsions of the Glu474 side chain dihedrals, with the remainder of the protein complex showing no appreciable structural deviations. Relevant geometrical parameters are given in Table 1. The Glu474 side chain switches from an extended trans conformation at $d(\text{OH}) \approx 7 \text{ \AA}$ to a gauche conformation at $d(\text{OH}) \approx 4 \text{ \AA}$. The calculated free-energy profile shown in Figure 4 has a free-energy barrier of 12 kcal/mol for reaction from right to left. Thus, the Glu474 side chain is flexible, and the transition that approximates the acidic hydrogen to the leaving group is thermally accessible at room temperature. It has been proposed that a related conformational transition could be rate-limiting in the Cdc25-catalyzed dephosphorylation of small-molecule substrates containing leaving groups with high $\text{p}K_a$'s.¹¹ However, the simulations presented here indicate that the chemical step is slower than the Glu474 conformational transition. Nevertheless, the free-energy difference between the right and left minima, 7 kcal/mol, should be added to the

reaction barrier calculated for the dianion reaction (Figure 2). Measured reaction rates are related to the lowest accessible free-energy minima and the top of the barrier.

Table 1 also shows interactions of Glu474 that may be explored in the rational design of inhibitors to Cdc25s. For instance, a possible inhibitor could push the Glu474 side chain away from the Cdc25B active site in the direction of Arg427. Then, Glu474 O_ε could hydrogen bond to the side chains of either Arg427 and Tyr528. In fact, a hydrogen bond is already formed between Glu474 and Tyr528 side chains in the initial structure. A putative inhibitor could prevent disruption of this last hydrogen bond, preserving a favorable energetic contact and, hence, block the conformational transition that decreases *d*(OH) and allows general acid catalysis. Such an inhibitor could also benefit from π -stacking interactions with Phe475.

The present study gives compelling evidence that Glu474 on Cdc25B, or the equivalent amino acid in other Cdc25 isoforms, performs as the general acid for dephosphorylation of its natural substrate. The different identity and location of this general acid on Cdc25 from other PTPs can be explored in the design of specific inhibitors. Actually, we believe this is a new and important strategy to design specific inhibitors for Cdc25 phosphatases. This inhibitory mechanism would involve obstruction of Glu474 side chain flexibility, avoiding the general acid to get close enough to the substrate leaving group for H⁺ transfer. In fact, a related possibility has already been explored with success for phosphatase Ptp1B. A conformational transition of the WPD loop in Ptp1B that correctly places the catalytic general acid was blocked by small molecules.²⁷ We are now exploring similar inhibitory strategies in our laboratory.

Acknowledgment. The author would like to acknowledge funding from FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo, Projects 07/52772-6 and 07/59345-6).

References and Notes

- (1) Lyon, M. A.; Ducruet, A. P.; Wipf, P.; Lazo, J. S. *Nat. Rev. Drug Discovery* **2002**, *1*, 961–976.
- (2) Ducruet, A. P.; Vogt, A.; Wipf, P.; Lazo, J. S. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 725.

- (3) Jiang, Z.-X.; Zhang, Z.-Y. *Cancer Metastasis Rev.* **2008**, *27*, 263–272.
- (4) Contour-Galcera, M.-O.; Sidhu, A.; Prvost, G.; Bigg, D.; Ducommun, B. *Pharmacol. Ther.* **2007**, *115*, 1–12.
- (5) Robertson, J. G. *Biochemistry* **2005**, *33*, 5561–5571.
- (6) Rudolph, J. *Biochemistry* **2007**, *46*, 3595–3604.
- (7) Jackson, M. D.; Denu, J. M. *Chem. Rev.* **2001**, *101*, 2313–2340.
- (8) Arantes, G. M. *Biochem. J.* **2006**, *399*, 343–350.
- (9) Fauman, E. B.; Cogswell, J. P.; Lovejoy, B.; Rocque, W. J.; Holmes, W.; Montana, V. G.; Piwnica-Worms, H.; Rink, M. J.; Saper, M. A. *Cell* **1998**, *93*, 617–625.
- (10) Reynolds, R. A.; Yem, A. W.; Wolfe, C. L.; Deibel, M. R.; Chidester, C. G.; Watenpaugh, K. D. *J. Mol. Biol.* **1999**, *293*, 559–568.
- (11) McCain, D. F.; Catrina, I. E.; Hengge, A. C.; Zhang, Z.-Y. *J. Biol. Chem.* **2002**, *277*, 11190–11200.
- (12) Rudolph, J. *Biochemistry* **2002**, *41*, 14613–14623.
- (13) Sohn, J.; Parks, J. M.; Buhman, G.; Brown, P.; Kristjansdottir, K.; Safi, A.; Edelsbrunner, H.; Yang, W.; Rudolph, J. *Biochemistry* **2005**, *44*, 16563–16573.
- (14) Field, M. J. *A Practical Introduction to the Simulation of Molecular Systems*, 1st ed.; Cambridge University Press: Cambridge, U.K., 1999.
- (15) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (16) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (18) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
- (19) Thomas, A.; Jourand, D.; Bret, C.; Amara, P.; Field, M. J. *J. Am. Chem. Soc.* **1999**, *121*, 9693–9702.
- (20) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (21) Arantes, G. M.; Loos, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347–353.
- (22) Arantes, G. M. On the Protein Tyrosine Phosphatases. Phosphate Esters Intrinsic Reactivity and Computer Simulation of the Mechanisms of Enzymatic Reaction. Thesis, Institute of Chemistry, University of São Paulo, 2004.
- (23) Allen, M.; Tildesley, D. *Computer Simulation of Liquids*, 1st ed.; Oxford University Press: New York, 1987.
- (24) Field, M. J.; Albe, M.; Bret, C.; Martin, F. P.-D.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088–1100.
- (25) Sohn, J.; Rudolph, J. *Biophys. Chem.* **2007**, *125*, 549–555.
- (26) McCain, D. F.; Grzyska, P. K.; Wu, L.; Hengge, A. C.; Zhang, Z.-Y. *Biochemistry* **2004**, *43*, 8256–8264.
- (27) Wiesmann, C.; Barr, K. J.; Kung, J.; Zhu, J.; Erlanson, D. A.; Shen, W.; Fahr, B. J.; Zhong, M.; Taylor, L.; Randal, M.; McDowell, R. S.; Hansen, S. K. *Nat. Struct. Mol. Biol.* **2004**, *11*, 730–737.

JP8070019

ANEXO C – Approximate
Multiconfigurational Treatment of
Spin-Coupled Metal Complexes

Approximate Multiconfigurational Treatment of Spin-Coupled Metal Complexes

 Guilherme Menegon Arantes*[†] and Peter R. Taylor[‡]

Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900, São Paulo, SP, Brazil and Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom

Received March 8, 2010

Abstract: The weak interaction between unpaired electrons in polynuclear transition-metal complexes is often described by exchange and spin polarization mechanisms. The resulting intrinsic multiconfigurational electronic structure for such complexes may be calculated with wave function-based methods (e.g., complete active space configuration interaction and complete active space self-consistent field), but computations become extremely demanding and even unfeasible for polynuclear complexes with a large number of open-shells. Here, several levels of selection of configurations and symmetry considerations that still capture the essential physics of exchange and spin polarization mechanisms are presented. The proposed approximations result in significantly smaller configuration interaction expansions and are equally valid for ab initio and semiempirical methods. Tests are performed in simple molecular systems and in small transition-metal complexes that cover a range of valence and charge states. In particular, superexchange contributions can be calculated to good accuracy using only single ionic excitations. Further reduction in the size of the configuration expansions is possible but restricts the description to low-lying spin ladders. The proposed configuration interaction schemes may be used to resolve space and spin symmetries in the calculation of electronic structures, exchange coupling constants, and other properties pertinent to polynuclear transition-metal complexes.

1. Introduction

Polynuclear transition-metal (TM) compounds with weakly coupled open-shell electrons have interesting magnetic properties as a consequence of the population at thermal energies of low-lying excited states with different total spins. The underlying interactions are traditionally mapped to a spin–spin coupling between momenta \mathbf{S} localized in neighboring magnetic sites and are often described by the Heisenberg–Dirac–van Vleck spin Hamiltonian:¹

$$\hat{H}_{\text{HDVV}} = - \sum_{A < B} J_{AB} \mathbf{S}_A \cdot \mathbf{S}_B \quad (1)$$

where J_{AB} is the isotropic Heisenberg coupling constant between spins on sites A and B. Since $[\hat{H}_{\text{HDVV}}, \hat{S}^2] = 0$, the two operators share a common set of eigenstates. The eigenvalues correspond to a spin ladder, and the energy gaps between low-lying spin states depend linearly on the J coupling constant. For the simplest case of a pair of magnetic sites with spins \mathbf{S}_A and \mathbf{S}_B , the coupling is ferromagnetic, and $J > 0$ in the sign convention assumed on eq 1, if the ground state is high-spin $S = S_A + S_B$. The coupling is antiferromagnetic, and $J < 0$, if the ground state is low-spin $S = |S_A - S_B|$. The spin–spin interaction modeled by eq 1 is in fact an effective one. As proposed by Heisenberg² and Dirac,³ the interactions arise due to spin-independent Coulomb electron–electron repulsion and exchange symmetry.

First-principles calculations with the spin-free electronic Hamiltonian should then be able to predict spin eigenstates and J constants for TM compounds. Anderson^{4,5} was seminal

* Corresponding author. Email: garantes@iq.usp.br.

[†] Universidade de São Paulo.

[‡] University of Warwick.

in realizing how to extract the main contributions to the effective spin coupling from the electronic structure. His model can be understood by considering the following simple valence-bond (VB) scheme: A pair of magnetic sites A and B contain two weakly interacting electrons occupying two orthogonal orbitals (constructed by a suitable rotation of the molecular orbitals) labeled a and b localized on centers A and B, respectively. By weakly interacting it should mean that the two electrons do not form a covalent bond. This situation corresponds to a dihydrogen molecule at stretched bond distance or a spin-coupled Cu(II, d^9) dimer. Four Slater determinants with $M_S = 0$ can be constructed: $|a\bar{b}|$, $|\bar{a}b|$, $|a\bar{a}|$, and $|\bar{b}b|$. The first two are “neutral” configurations, and the last two are charge-transfer “ionic” configurations. Their combination results in the following configuration state functions:

$$\begin{aligned} |^1\Psi_{\text{neu}}\rangle &= 2^{-1/2}[|a\bar{b}| - |\bar{a}b|] \\ |^1\Psi_{\text{ion}}^A\rangle &= |a\bar{a}| \\ |^1\Psi_{\text{ion}}^B\rangle &= |b\bar{b}| \\ |^3\Psi_{\text{neu}}\rangle &= 2^{-1/2}[|a\bar{b}| + |\bar{a}b|] \end{aligned} \quad (2)$$

The energy difference between the triplet $|^3\Psi_{\text{neu}}\rangle$ and the singlet $|^1\Psi_{\text{neu}}\rangle$ will be proportional to K_{ab} , the exchange integral between orbitals a and b .⁶ This direct exchange interaction is ferromagnetic because the high-spin state (triplet) is favored. Configuration mixing between neutral and ionic states will lower the singlet energy and lead to the ground state.⁶

$$|^1\Psi_{\text{Cl}}\rangle = (1 - \alpha)^{1/2}|^1\Psi_{\text{neu}}\rangle + \alpha^{1/2}|^1\Psi_{\text{ion}}\rangle \quad (3)$$

where $|^1\Psi_{\text{ion}}\rangle$ is a superposition of the two ionic configuration state functions shown above and α gives the degree of mixing between the neutral and ionic states. This mixing is equivalent to a virtual hopping of the electron from one magnetic site to the other (the charge-transfer ionic configurations), and it gives an antiferromagnetic contribution to spin-coupling because the low spin (singlet) is favored. In general, for weakly coupled open-shell compounds with several unpaired electrons, neutral configurations will appear in the wave function expansion for all spin states. Their contribution to spin coupling is ferromagnetic, i.e., stabilize the high-spin state, and is known as the *direct exchange* effect or mechanism. Ionic configurations will appear in expansions of all but the highest spin state and give antiferromagnetic contributions known as the through-space *superexchange* mechanism.

This simple VB model can be expanded to explicitly include an occupied valence closed shell of diamagnetic ligand bridges that coordinate metal ions in TM complexes. Ligand-to-metal charge-transfer (LMCT) excitations built out of a set of neutral and ionic configurations, equivalent to those on eq 3, will have either anti- or ferromagnetic contributions to spin coupling. This issue is discussed in more detail below. To make a connection with the jargon of previous perturbative treatments,^{7–9} it should be noted that single LMCT excitations out of neutral configurations are usually called ligand spin polarization (LSP) because an

effective spin density appears on the bridge.⁸ Double LMCT excitations are termed dynamic or double spin polarization (DSP). Excitations from core orbitals or to unoccupied orbitals have been suggested to account for dynamic correlation and orbital relaxation effects^{9,10} and, hence, do not comprise additional spin-coupling mechanisms.

Another modification of the two electrons in two localized orbitals scheme presented above is the addition of a third electron resulting in a mixed valence compound such as the stretched H_2^- molecule. Delocalization or “resonance” of the excess electron between the magnetic sites A and B stabilizes the system and occurs favorably when the local spins \mathbf{S}_A and \mathbf{S}_B are aligned in parallel. This *double exchange* effect may then give effective ferromagnetic contributions to the spin coupling in mixed valence TM complexes.^{11,12}

The method most widely used today to predict J coupling constants for polynuclear complexes is the broken-symmetry approach proposed by Noodleman.^{13,12} In this single configuration description, the solution for the low-spin state (the BS state, corresponding to $M_S = |S_A - S_B|$ in the above example with two magnetic centers) has space and spin symmetries broken. Such state is not a spin eigenstate but a superposition of spin states weighted by Clebsch–Gordan coefficients. A value for J can be estimated¹⁴ by using spin-projection techniques and by also computing the highest spin (HS) state, which usually is well described by a single configuration:

$$J = -\frac{E_{\text{HS}} - E_{\text{BS}}}{\langle \hat{S}^2 \rangle_{\text{HS}} - \langle \hat{S}^2 \rangle_{\text{BS}}} \quad (4)$$

where E is the state energy and $\langle \hat{S}^2 \rangle$ is the expectation value of the total spin operator. The success of the broken-symmetry approach can be traced to appropriate descriptions of direct exchange, superexchange, and LSP mechanisms discussed above.⁸ However, its accuracy obviously depends on the electronic structure method employed for the single configuration calculations, which is often spin-polarized density functional theory. Because eigenfunctions for the lower spin states are not obtained explicitly, the broken-symmetry approach is not suited to study state specific properties. Nevertheless, mapping and spin-projection techniques may also be applied to estimate \mathbf{g} tensors and hyperfine coupling constants¹⁵ and to optimize geometries¹⁶ approximately. Along the same line, an extended broken-symmetry approach has been introduced recently that allows the calculation of energy derivatives for homovalent binuclear complexes.¹⁷

From the VB discussion in the previous paragraphs, it seems evident to employ configuration interaction (CI) of Slater determinants to compute wave functions for low-spin eigenstates. All spin-coupling mechanisms and electronic effects cited above can be naturally accounted for if an appropriate configuration space is used. However, the exponential scaling of the size of the CI space puts serious limitations on the range of TM complexes and properties that can be calculated with CI. For instance, the configurational space generated in full excitation level for about 18 unpaired electrons already exceeds the capacity of modern

CI code implementations and computer hardware. At this point, some heroic CI computations on low-spin states of binuclear TM complexes by Malrieu and collaborators should be mentioned.^{18,9} Their dedicated difference CI method has been used to compute energy differences between spin multiplets in very good agreement with experimental data. Together with perturbative analysis, this CI method has also been used to identify contributions to spin coupling.^{9,10} Even so, the dedicated difference CI also suffers from an exponential scaling of the CI space and thus is limited to binuclear complexes with a small number of unpaired electrons.

In this paper approximate levels of CI selection are proposed in trying to find short CI expansions that still capture the essential physics of spin coupling for the low-spin eigenstates. Determinants are built with localized molecular orbitals. But instead of specifying a given level of excitation from a single reference as in canonical CI, the configurational space is built by completing the spin manifold for neutral (or covalent), ionic, and ligand-to-metal charge-transfer VB-like structures. It is important to note that all approximations proposed here concern only the selection of configurations that enter in the CI. Thus, all the conclusions obtained should be equally valid irrespective of the method, semiempirical or *ab initio*, used to calculate the molecular integrals and configuration energies. A semiempirical Hamiltonian was employed here because future applications of the proposed approximations will use a hybrid quantum/classical potential based on semiempirical methods. Tests are performed in several simple systems so that full CI calculations can be carried out as references. Details of the computational methods are given in the next section. The results show that single ionic excitations between magnetic sites are enough to obtain an accurate superexchange contribution. Further reduction in the size of the CI space is possible but restricts the description to ground spin ladders. For iron–sulfur clusters, spin coupling can be correctly described by rather small CI expansions, paving the way for simulation studies of magnetic and electronic properties of these prosthetic groups in the condensed phase.

2. Computational Methods

Test calculations were performed on simple spin-coupled molecular systems. Two homonuclear diatomics, N_2 and Cr_2 , two bridged triatomics, N_2F^- and Fe_2S^{4+} , and the ring cluster $Fe_2S_2^{2+}$, were studied. Dinitrogen bond distance was set to 4.5 bohr (~ 2.86 Å), and the dichromium bond distance was set to 4.4 bohr (~ 2.33 Å). At such separations, covalent bonding is not significant, and energy splittings between the total spin eigenstates have magnitudes similar to those observed in polynuclear TM complexes. The equilibrium bond lengths for dinitrogen and dichromium are ~ 1.11 and ~ 1.68 Å, respectively. Each atom in the stretched diatomic molecule plays the role of an open-shell metal center or magnetic site. The unpaired electrons are weakly interacting, in a suitable model to the direct exchange and through-space superexchange mechanisms. Yet, dinitrogen is simple enough to allow complete expansions of the electronic wave function as well as several levels of CI selection. Neutral, dipositive, and mononegative total molecular charges were assigned for

dinitrogen as models of magnetic compounds with half-full open shell, less than half-full, and mixed valence, respectively. Triatomic molecules composed of two magnetic centers separated by a diamagnetic ligand are the simplest systems to probe the effect of the proposed approximations on interactions via the ligand spin polarization mechanism. Since bridge ligands found in TM complexes are usually diamagnetic anions, stretched dinitrogen was bridged with fluoride in an angular geometry with C_{2v} symmetry, $\angle = 75^\circ$, $d(N-F) = 1.80$ Å, and $d(N-N) = 2.19$ Å. In the TM compound Fe_2S^{4+} , two Fe(III) are bridged by a sulfide ligand. A symmetric linear geometry was adopted with $d(Fe-S) = 1.271$ Å. The binuclear iron–sulfur cluster $Fe_2S_2^{+1/+2}$ is the prosthetic group found in many electron-transfer proteins, such as ferredoxin. Each iron is also attached to the protein by two cysteine sulfur atoms, with a total tetrahedral coordination. By contrast, the bare $Fe_2S_2^{2+}$ cluster studied here, a D_{2h} geometry was used,¹⁹ with $d(Fe-Fe) = 2.543$ and $d(Fe-S) = 2.251$ Å. The z axis contains the two magnetic sites in all molecules studied.

Calculations were carried out with a semiempirical neglect of diatomic differential overlap (NDDO) Hamiltonian.^{20,21} A slightly modified version of the MOPAC2000^{22,23} code that allowed CI calculations using localized molecular orbitals was employed. Standard AM1 parameters were used for nitrogen and fluoride²⁴ and modified neglect of differential overlap (MNDO)-d parameters were used for sulfur.²⁵ MNDO-d parameters were not available for chromium and iron, so a quick parametrization had to be done. See details and the parameter values in the Supporting Information. Molecular orbitals (MOs) were obtained from high-spin restricted open-shell Hartree–Fock (ROHF) calculations and were localized using an equivalent Pipek–Mezey procedure.²⁶ Although MOPAC does not work with symmetry-adapted basis, all resulting wave functions were checked for the correct space and spin symmetries. Active spaces defined for the CASCI (full CI on the given active space)²⁷ calculations contained all open-shell MOs as well as outer valence unoccupied and double-occupied MOs in N_2^{+2} and N_2^- , respectively. All unpaired electrons were included in the active spaces. Full details of the active spaces used are given for each tested molecule in the Results and Discussion Section. Approximate CI expansions were based on the VB arguments presented in the Introduction. Hence, instead of specifying a given level of excitation from the ROHF solution, the selected CI expansions included all determinants needed to complete the spin manifold for a given level of approximation for the mechanisms of effective spin-coupling discussed. Only $M_S = 0$ (or $M_S = 0.5$, for N_2^-) determinants were used in the selected CI expansions.

For the larger active spaces, CASCI calculations were not feasible for the low-spin states (singlet and triplet). MOPAC generates and diagonalizes the CI matrix (or secular determinant) explicitly, and the code could not be compiled to use more than 2 GB of memory. Thus, the size of the CI expansions were limited to about 9000 configurations, which is less than the number of configurations necessary to expand the singlet and triplet states for the molecules formed by Cr and Fe. All the CASCI calculations were done with the

Table 1. Relative Energies (eV) and Number of Configurations (size) Included in the Wavefunction Expansions for Electronic Eigenstates of Dinitrogen in Neutral, Dipositive, and Negative Total Molecular Charge

		N_2^0			
CI expansion	size	$^1\Sigma_g$	$^3\Sigma_u$	$^5\Sigma_g$	$^7\Sigma_u$
CASCI	400	0.0000	0.0281	0.0876	0.1872
neu + single ion	80	0.0000	0.0278	0.0868	0.1864
neu + p_x, p_y, p_z ion	56	0.0000	0.0278	0.0868	0.1864
neu + p_z ion	32	0.0000	0.0262	0.0822	0.1778
neu + p_x, p_y ion	44	0.0000	0.0014	0.0043	0.0087
		N_2^{2+}			
CI expansion	size	$^1\Sigma_u$	$^3\Sigma_g$	$^5\Sigma_u$	
CASCI	225	0.0000	0.0515	0.1953	
neu + single ion	162	0.0000	0.0514	0.1939	
neu + unpair, p_z ion	114	0.0000	0.0514	0.1939	
neu p_z + unpair, p_z ion	72	0.0000	0.0514	0.1939	
neu + p_z ion	78	0.0000	0.0567	0.2028	
neu + unpair ion	90	0.0000	-0.0029	-0.0087	
		N_2^-			
CI expansion	size	$^2\Sigma_u$	$^4\Sigma_g$	$^6\Sigma_u$	
CASCI	300	0.3125	0.1458	0.0000	
neu + single ion	240	0.3126	0.1458	0.0000	
neu p_z + p_z ion	44	0.3130	0.1460	0.0000	
neu	60	0.3298	0.1568	0.0000	
neu p_z	20	0.3298	0.1568	0.0000	

semieempirical NDDO Hamiltonian. The CASSCF method²⁷ within the MOLCAS 6.2 program system²⁸ was used to compute a reference value for the $Fe_2S_2^{2+}$ cluster. This calculation was conducted with basis symmetry adapted to the D_{2h} point group, using the ANO-RCC²⁹ set with quadruple- ζ contraction (e.g., 7s6p4d3f2g for iron).

3. Results and Discussion

Results of several approximate levels of CI selection on the electronic structure of simple molecules are presented in this section. For the diatomic systems and the linear $Fe_2S_2^{4+}$, all spin ladders shown are Σ states. For NFN^- , the lowest energy spin states are alternating A_1 and B_2 states, and for the ring Fe_2S_2 , the spin ladder shown has alternating A_g and B_{1u} states. For example, the correct energy ordering for the total spin eigenstates of neutral N_2 is $^1\Sigma < ^3\Sigma < ^5\Sigma < ^7\Sigma$.

3.1. Neutral N_2 . For neutral N_2 , the following configuration is obtained after localizing the high-spin ROHF MOs: $[\text{core}]2s^A 2s^A 2s^B 2s^B 2p_z^A 2p_z^B 2p_x^A 2p_x^B 2p_y^A 2p_y^B$, where the over bar assigns spin down and the superscripts A and B are used to label each nitrogen atom. Localized MOs have large contributions by only one atomic function which is then used as a label. The six unpaired electrons in the six 2p MOs are responsible for the spin coupling and form the active space for generation of configurations used in the wave function expansion. Because of localization, the MOs will have a $C_{\infty v}$ symmetry, which is lower than the nuclear point group.

The relative energies obtained for the lowest energy spin eigenstates are shown in Table 1. The CASCI has a total of 400 configurations with $M_S = 0$. There are 20 unpaired neutral configurations, i.e., with 1 electron in each of the 6

active MO. The septet wave function is composed only by these 20 configurations, with equal CI weights. The largest CI weights (~ 0.24 in the singlet state) in the expansions for all other spin states come from two configurations, $|p_z^A p_z^B p_x^A p_x^B p_y^A p_y^B|$ (only the active space is represented on this and the following determinant configurations) and the respective A to B spin inversion. These two configurations correspond to a 4S high-spin state on each N atom. The second largest contributions come from the other 18 unpaired neutral configurations, such as $|p_z^A p_z^B p_x^A p_x^B p_y^A p_y^B|$, which corresponds to combinations of atomic excited states or non-Hund states.³⁰ Ionic configurations have rather smaller contributions (CI weight ≤ 0.03 in the singlet). The next-lying excited state above the septet shown in Table 1 is at least 2 eV higher in energy.

Judgement from the weights in the CASCI expansion would suggest that only the 20 unpaired neutral configurations could be used in the wave function expansion for all spin eigenstates. However, this approximation results in a flat spin ladder, with the same energy for all states. As described in the Introduction Section, neutral configurations are not able to account for the effective antiferromagnetic interactions between the open shells. The ladder is flat because MOs are strictly localized so that the direct exchange (K_{ab}) ferromagnetic contribution is very small, actually null in the precision used. The first reasonable level of approximation, named neu + single ion in Table 1, is an expansion containing 20 neutral configurations plus all the 60 symmetry-allowed "metal-to-metal" (or nitrogen-to-nitrogen) ionic single excitations that can be constructed from the set of neutral configurations, e.g., $|p_z^B p_z^A p_x^B p_x^A p_y^B p_y^A|$. The energy values obtained with this expansion are within 0.001 eV of the CASCI reference, and the number of configurations used is five-fold smaller. Since localized MOs are used, excitations between MOs that belong to the same irrep of $C_{\infty v}$ are symmetry allowed. A second approximation can be made by including neutral and single ionic excitations only between localized MOs composed by the same atomic functions (neu + p_x, p_y, p_z ion, Table 1). This results in identical energies showing that symmetry-allowed "crossed" ionic excitations (e.g., $p_x^B \rightarrow p_y^A$) do not interact with the wave function for the low-lying states of neutral N_2 . An expansion including neutral and the 12 single ionic excitations between the 2p_z MOs (neu + p_z ion) results in energies within 0.01 eV of the CASCI reference. This suggests a third level of approximation in which the only ionic excitations included are those between MOs composed of atomic functions with large overlap (the z axis is the intermolecular axis). As a counter example, an expansion including neutral and ionic excitations between MOs composed of atomic functions with small overlap (neu + p_x, p_y ion) results in almost no antiferromagnetic contributions and a spin ladder in large disagreement with the CASCI reference. It should be noted that, by progressively removing from the CI space the excitations between 2p_x and 2p_y MOs (as in neu + p_x, p_y, p_z ion and in neu + p_z ion), spin ladders of higher energy and different space symmetry will not be correctly described. This is not a problem for neutral N_2 because the next-lying state above the $^7\Sigma$ state is much higher in energy, but it might

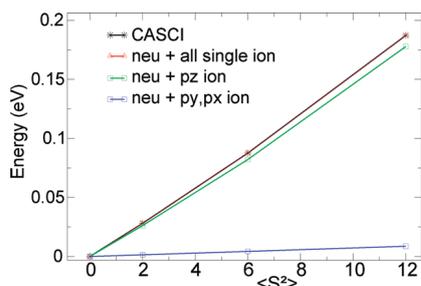


Figure 1. Spin ladders for the lowest energy total spin eigenstates of N_2^0 calculated with different wave function expansions. See text for details.

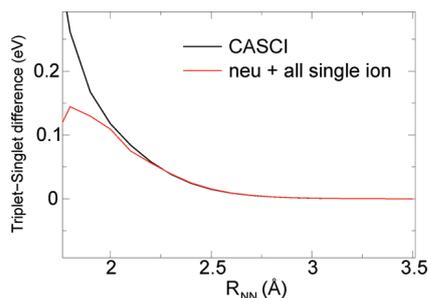


Figure 2. Triplet-singlet energy gap for varying N_2^0 bond distances.

introduce errors when the ground spin ladder is near degenerate to other ladders.

Double ionic ($N^{2-}-N^{2+}$), triple ionic ($N^{3-}-N^{3+}$), and internal paired neutral configurations, e.g. $|p_z^A p_z^B p_x^A p_x^B\rangle$, which also corresponds to non-Hund atomic states, have very small contributions and can be safely neglected. Removing the two neutral configurations corresponding to the 4S high-spin state on each N atom from the expansion neu + single ion or using only these two neutral configurations plus all single ionic ones results in an incomplete spin manifold and, consequently, bogus spin ladders.

Linear spin ladders, i.e., ladders that follow a regular Landé splitting, are obtained within the CASCI, and the levels of approximation suggested above are shown in Figure 1. The CASCI ladder and the expansion named neu + single ion have both correlation coefficients to a straight line of 0.9994 and a F variance quality of 1662. The expansion neu + p_z ion has a correlation of 0.9992 and a F variance quality of 1187. In conclusion, the CI expansion neu + single ion captures the essential physics of exchange interactions for the ground spin ladder (Table 1) as well as for higher energy ladders (not shown) of the stretched dinitrogen molecule.

To test the limits of the proposed configuration selection, the singlet-triplet energy gap was calculated with varying bond distances. Figure 2 shows that the expansion neu + single ion results in energy gaps in very good agreement with the CASCI wave function down to bond distances of ~ 2.0 Å. Below this distance, the interaction between the unpaired electrons is strong, and covalent bonding becomes appreciable. The system is not only spin coupled, and the proposed approximate CI selections do not apply.

3.2. N_2^{2+} . For N_2^{2+} , the configuration obtained after localizing the high-spin ROHF MOs is equivalent to the neutral N_2 configuration (see above) but with two previous

highest occupied molecular orbitals (HOMOs) now unoccupied. The relative energies obtained for the lowest energy spin eigenstates are shown in Table 1. The expansion neu + single ion results in energy values in excellent agreement (within 0.002 eV) with the CASCI reference. For less than half-filled open shells, there are ionic configurations which still have all electrons unpaired. There are 36 of such unpaired ionic configurations for N_2^{2+} . An expansion including all neutral configurations, unpaired ionic and single ionic excitations between the $2p_z$ MOs (neu + unpair, p_z ion) result in energies identical to the neu + single ion expansion. Single ionic excitations between MOs composed by atomic functions with small overlap (e.g., $p_y^B \rightarrow p_y^A$) and crossed single excitations do not interact with the wave function for the low-lying states of N_2^{2+} . A selection of the neutral configurations included in the expansions is possible for the open-shell systems without exactly half-full shells, i.e., more or less than half-filled and mixed valence. An expansion including only neutral configurations with one electron in each $2p_z$ MOs, unpaired ionic and single ionic excitations between the $2p_z$ MOs (neu p_z + unpair, p_z ion) also result in energies identical to the neu + single ion expansion. An expansion including all neutral configurations and the 24 single ionic excitations between the $2p_z$ MOs (neu + p_z ion) results in energies within 0.01 eV of the CASCI reference. But, contrary to the equivalent neu + p_z ion expansion for the neutral N_2 , an excess antiferromagnetic character is observed. This is a consequence of neglecting the ferromagnetic contribution of unpaired ionic configurations, easily seen in the results for the neu + unpair ion expansion in Table 1. Thus, not all metal-to-metal ionic excitations give an antiferromagnetic contribution to spin coupling, but only those that alter the number of unpaired electrons.

Considering a particle-hole symmetry, an equivalent behavior would be observed for the more than half-filled case. For example, in N_2^{2-} , ionic configurations without an empty MO give ferromagnetic contributions, equivalent to the ionic unpaired configurations in the less than half-filled case.

3.3. N_2^- . For N_2^- , the localized high-spin ROHF MOs used in the CI expansions were obtained for the neutral dinitrogen to avoid an artificial polarization of the occupied MOs and thereof biased CI results. Similar results were obtained if a fractional occupation of the MOs was allowed in the ROHF solution. The relative energies obtained for the lowest energy spin eigenstates are shown in Table 1. Delocalization of the excess electron stabilizes the “neutral” configurations resulting in a ferromagnetic CASCI spin ladder. This is the double-exchange effect.¹¹ Antiferromagnetic contributions by the superexchange mechanism are an order of magnitude smaller. Thus, an expansion including only neutral configurations (neu, Table 1) accounts for the double-exchange effect and results in energies within 0.02 eV of the CASCI reference. In fact, an expansion (neu p_z) in which the excess electron occupies only the $2p_z$ orbitals has identical results. However, by removing from the CI space configurations in which the excess electron occupies the $2p_x$ and $2p_y$ MOs, spin ladders

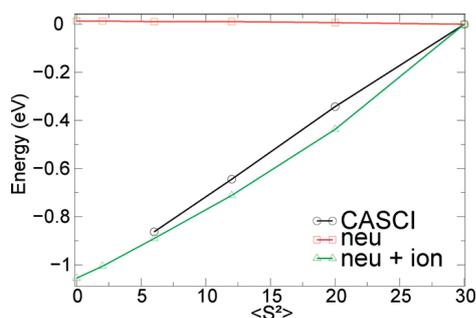


Figure 3. Spin ladders for the lowest energy total spin eigenstates of Cr_2 at 4.4 bohr separation calculated with different wave function expansions.

with higher energy and different space symmetry will not be correctly described, as observed for similar CI selections in $\text{N}_2^{0/+2}$.

The antiferromagnetic contribution can be retrieved in an expansion including all symmetry-allowed single ionic excitations (neu + single ion) resulting in energies within 0.0001 eV of the CASCI reference. The expansion including the interacting neutral and the 24 single ionic excitations between the $2p_z$ MOs (neu $p_z + p_z$ ion) contains five-fold less configurations than the CASCI and results in energies within 0.001 eV of this reference.

3.4. Cr_2 . For the stretched dichromium molecule, covalent bonding between the 3d orbitals is not significant. However, there is still a σ bond formed mostly between the diffuse 4s chromium orbitals.³¹ The correct energy ordering for the total spin states should have the antiferromagnetic singlet as the ground state and the ferromagnetic undecaplet as the highest energy state of the ground spin ladder.

The canonical high-spin ROHF solution has 10 singly occupied MOs formed by antisymmetric and symmetric combinations of the atomic 3d functions. The HOMO-1 and HOMO are formed, respectively, by antisymmetric and symmetric combinations of the 4s functions. After full orbital localization, each Cr atom contains six electrons and a configuration corresponding to a 7S atomic state. The active space was composed of the 12 electrons in 10 MOs formed by 3d functions and the 2 MOs formed by 4s functions. All the configurations used in the expansions were formed out of the two possible combinations of the localized 4s orbitals consistent with a σ bonding MO. The CASCI solution was only computed down to the quintet state. The secular determinants necessary to obtain states with $S < 2$ were too large and could not be built due to memory limitations (see Computational Methods Section).

Figure 3 shows spin ladders calculated for Cr_2 under different CI selections. The undecaplet state was chosen as zero of energy. An expansion including only unpaired neutral configurations (252 configurations in total), e.g., $|\text{core}\rangle d_{xz}^A d_{x^2-y^2}^A d_{xy}^A d_{yz}^A d_{yz}^B d_{xz}^B d_{x^2-y^2}^B d_{xy}^B$, yields an incorrect spin ladder with a high-spin ground state, as expected from the direct-exchange contribution to spin coupling. The expansion neu + single ion, including all 252 unpaired neutral configurations plus 1260 ionic configurations, results in fair agreement with the CASCI result (within 0.1 eV). Ionic configurations were built from the set of unpaired

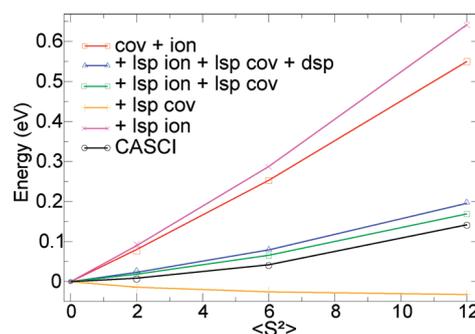


Figure 4. Spin ladders for the lowest energy total spin eigenstates of NFN^- in angular geometry calculated with different wave function expansions.

neutral configurations by metal-to-metal single excitations between MOs belonging to the same irreps of the $C_{\infty v}$ group. It should be noted that the neu + single ion CI expansion contains only 1512 configurations, instead of the 63 504 configurations that would be necessary to expand the singlet state in the CASCI wave function. Energies within 0.01 eV of the neu + single ion expansion are obtained by a smaller expansion with 952 configurations that does not contain the crossed ionic excitations between MOs belonging to the same $C_{\infty v}$ irrep but formed by different atomic functions, e.g., $d_{xz}^A \rightarrow d_{yz}^B$.

The spin ladders obtained with CASCI and neu + single ion approximation have, respectively, correlation coefficients to a straight line of 0.9998 and 0.996 and a F variance quality of 4032 and 465. The approximations proposed for the model stretched dinitrogen are equally valid for the stretched dichromium and result in a reduction of at least two orders of magnitude in the size of the CI space.

3.5. NFN^- in C_{2v} Symmetry. On the following sections, the proposed approximations are tested on compounds containing diamagnetic bridges. Localized MOs were obtained for angular NFN^- from a high-spin ROHF solution. Each nitrogen has a double-occupied 2s-like shell and 3 unpaired electrons in orbitals composed by the 2p functions. Fluoride has 4 double-occupied orbitals composed by 2s and 2p functions. All 9 MOs formed by p functions and 12 electrons are included in the active space.

Figure 4 shows spin ladders calculated under different CI selections. The singlet was chosen as zero of energy. An expansion including only unpaired neutral and ionic single excitations between the magnetic (nitrogen) centers with the bridge (fluoride) MOs left double occupied (neu + single ion, 80 configurations in total) yields a largely antiferromagnetic ladder, in large disagreement with the CASCI reference. The ligand spin polarization has to be included for a qualitatively correct description of the NFN^- wave function.

LSP configurations are obtained by LMCT *single* excitations built from the set of neutral and ionic determinants. An expansion including the neu + single ion set and all configurations generated from the (20) neutral determinants by LMCT single excitations (+ lsp neu, 280 configurations in total) results in an overestimation of the ferromagnetic interactions. On the other hand, a similar expansion (+ lsp ion, 406 configurations in total) but with LSP configurations

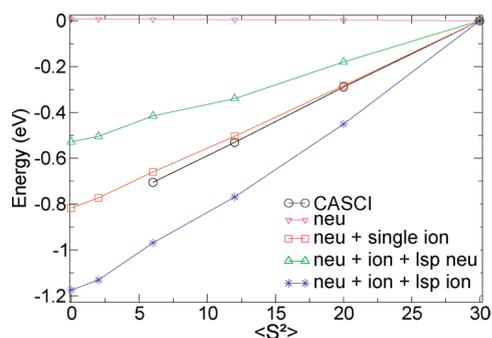


Figure 5. Spin ladders for the lowest energy total spin eigenstates of linear $\text{Fe}_2\text{S}_4^{4+}$ calculated with different wave function expansions.

generated by single excitation from the (60) ionic determinants results in an overestimation of the antiferromagnetic contributions. A proper balance is obtained by an expansion including the neu + single ion set plus both neutral and ionic ligand spin polarization (+ lsp ion + lsp neu), resulting in energies within 0.03 eV of the CASCI reference. Double LMCT excitations can also be constructed from the neutral and ionic configurations set by either exciting the same bridge orbital or two different ones. The resulting contributions are anti- and ferromagnetic but with much smaller magnitude (~ 0.01 eV), as found for other TM bridged systems.^{8,7}

In other words, the set of unpaired neutral and single ionic configurations might be considered a zero-order reference set. LSP excitations built out of this multireference set result in anti- and ferromagnetic contributions to spin coupling if the excitation originates from an ionic or a neutral configuration, respectively. For NFN^- , the LSP contributions are larger than the through-space superexchange contributions and have to be included for a qualitatively correct description of the spin coupling. This is not generally true, as shown below for the TM compounds. DSP contributions are relatively small and can be removed from the CI space without affecting the results significantly.

3.6. Linear $\text{Fe}_2\text{S}_4^{4+}$. Localized MOs were obtained for linear $\text{Fe}_2\text{S}_4^{4+}$ from a high-spin ROHF solution. Each metal center has a half-filled valence shell with 5 unpaired electrons in 5 orbitals composed by 3d functions, corresponding to an atomic ^6S state. The sulfur bridge has four double-occupied 2s- and 2p-like orbitals. The 3 outer-valence bridge MOs and the 10 MOs formed by iron 3d functions were included in the active space, with the respective 16 electrons.

A CASCI solution with such a large active space is not feasible within the memory limitations found here (see Computational Methods Section). Instead, Figure 5 shows a CASCI result obtained with only 10 electrons in the 10 MOs formed by iron 3d functions. The undecaplet state was chosen as zero of energy. The neu expansion includes only unpaired neutral configurations with double-occupied ligand MOs (252 configurations in total, Figure 5) and results in small ferromagnetic coupling, as observed above for the chromium dimer. The neu + single ion expansion includes the ionic configurations (1512 configurations in total) and results in very good agreement with the 10 electron in 10 orbitals CASCI. The effect of neutral LSP configurations (neu + ion + lsp neu) is ferromagnetic, and the ionic LSP (neu + ion

Table 2. Relative Energies (eV) for $\text{Fe}_2\text{S}_2^{2+}$ Lowest Energy Spin Eigenstates Calculated with Two Different CI Expansions

$\langle \hat{S}^2 \rangle$	neu + single ion	CASSCF
0	0.000	0.000
2	0.041	0.046
6	0.135	0.136
12	0.283	0.267
20	0.475	0.427
30	0.717	0.679

+ lsp ion) is antiferromagnetic. Neutral and ionic LSP configurations are obtained by LMCT *single* excitations built from the set of neutral (neu) and ionic (ion) determinants, respectively. However, contrary to the NFN^- example above, through-space superexchange dominates, and the ligand spin polarization is relatively smaller in $\text{Fe}_2\text{S}_4^{4+}$. For example, the ladder obtained with the expansion neu + ion + lsp neu is antiferromagnetic. In fact, inclusion of both neutral and ionic LSP configurations practically cancels out the polarization effect and results in a spin ladder very close to the neu + single ion expansion. Even if LMCT excitations are not explicitly included in the CI space, the effect of bridges and ligands is at least partially included when MOs are generated and when energies of the zero-order multireference configurations are calculated. Results similar to those shown in Figure 5 for the neu + single ion expansion are obtained by removing the crossed ionic excitations, as observed above for stretched N_2 and Cr_2 , leading an expansion with only 952 configurations.

3.7. $\text{Fe}_2\text{S}_2^{2+}$ Ring. The final example is the iron–sulfur cluster $\text{Fe}_2\text{S}_2^{2+}$. Localized MOs obtained from a high-spin ROHF solution show a half-filled valence 3d shell with 5 unpaired electrons in each iron center. Each sulfur bridge has 3 outer-valence double-occupied localized MOs composed by 2p functions. An active space containing all 16 valence MOs and the respective 22 electrons is only feasible using modern direct CI procedures. An ab initio CASSCF computation using such large active space is taken as reference in Table 2. The CASSCF singlet wave function is expanded in almost two million determinants in comparison to the approximate and much shorter expansion neu + single ion that includes only 1512 determinants corresponding to the unpaired neutral and ionic single excited states. The agreement between the CASSCF and the selected neu + single ion expansion is very good (within 0.05 eV) and suggests that this level of approximation captures the essential physics of spin coupling in transition-metal complexes. In fact, energies within 0.003 eV of the neu + single ion expansion were obtained with an even smaller expansion containing 952 configurations, by removing the crossed ionic excitations.

4. Conclusions

Approximate configuration interaction expansions were introduced for the calculation of wave functions with correct spin and space symmetries of weakly coupled transition-metal compounds with many open shells. The selection of configurations included in the CI space was based on physical

arguments for the mechanisms of spin coupling, namely direct exchange, superexchange, double exchange, and ligand spin polarization. In the spirit of valence-bond calculations, localized (molecular) orbitals were used in the construction of Slater determinants. But, instead of specifying a level of excitation as in the normal CI terminology, the expansions included all determinants needed to complete the spin manifold compatible with the exchange mechanisms depicted in the Introduction Section.

A zero-order multireference set was identified as the set of neutral and single ionic configurations. The neutral set accounts for the direct-exchange ferromagnetic mechanism and corresponds to configurations with an equivalent number of unpaired electrons in each magnetic site (excluding the excess electron in mixed valence systems). The ionic set is built by symmetry-allowed metal-to-metal single excitations from the neutral set that alter the total number of unpaired electrons. For all the spin-coupled compounds tested here and, we believe, for any spin-coupled system, single ionic excitations are enough to account for the through-space superexchange antiferromagnetic mechanism.

Symmetry-allowed excitations involve molecular orbitals that belong to the same irrep of the *localized* MOs point group. The contribution of symmetry-allowed crossed ionic excitations, i.e., excitations between MOs formed mainly by different atomic functions, was very small or null for the ground spin ladder in all molecules studied. For other systems, this result will depend on the localization method employed and on whether the localized MOs resemble pure atomic orbitals or combinations thereof. Even smaller expansions are possible by selectively removing from the CI space other ionic configurations or neutral configurations for the more or less than half-filled and mixed valence systems that have very small or null CI weights in the expansions of the low-lying spin states. For instance, removing excitations between MOs formed by atomic functions with a small overlap in N₂ resulted in energies close to those obtained with the full zero-order set for the ground spin ladder. However, spin ladders of higher energy and different space symmetry might not be correctly described by CI spaces smaller than the zero-order multireference set.

Ligand-to-metal charge-transfer configurations constructed from the zero-order reference set account for ligand spin polarization and double spin polarization. Single LMCT out of the neutral set always give a ferromagnetic contribution. On the other hand, single LMCT out of the ionic set always give an antiferromagnetic contribution, sometimes called "through-bond" superexchange. The LSP configurations should be included in the CI space whenever this contribution is comparable in magnitude to through-space direct and superexchange. For the iron-sulfur compounds studied here, the LSP contribution is small and approximately cancels out when both ionic and neutral single LMCT excitations are included. This is not a general result,^{9,10,30} but it is a valuable one in reducing the size of the CI expansions. A related argument is valid for the mixed valence system tested. The double-exchange effect in N₂⁻ is much larger

than the superexchange so that ionic configurations can be excluded from the CI space without affecting the energy splittings significantly.

Comparisons with experimental *J* coupling constants are not given here. Such comparisons would not be fair at this stage because the calculations presented do not include the effect of dynamic correlation, which is essential for quantitative results.^{27,9} Dynamic correlation can be added on top of the zero-order set by either multireference CI or perturbative corrections.²⁷ If a semiempirical method is employed, then correlation can be implicitly included in the parametrization of electron-repulsion integrals.

The proposed approximations result in much shorter CI expansions. For example, the CASSCF result obtained with 2×10^6 configurations for Fe₂S₂²⁺ is reproduced with about 10³ configurations. However, the exponential scaling of the CI space size is not entirely ameliorated. Polynuclear compounds with a larger number of magnetic centers and unpaired electrons will still require large configurational spaces that may exceed the available computational resources even if including only neutral and single ionic excitations between neighboring sites. Nevertheless, identifying the spin-coupling mechanisms with valence-bond structures and including controlled approximations in the CI expansion may open the way to treat these more challenging systems.

Acknowledgment. Jeppe Olsen (University of Århus) is acknowledged for helpful discussions. G.M.A. acknowledges funding from FAPESP, projects 07/52772-6 and 07/59345-6.

Supporting Information Available: Semiempirical parameters and the procedure used for their calibration. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Blondin, G.; Girerd, J.-J. *Chem. Rev.* **1990**, *90*, 1359–1376.
- (2) Heisenberg, W. *Z. Phys.* **1928**, *49*, 619–636.
- (3) Dirac, P. *Proc. R. Soc. London, Ser. A* **1929**, *123*, 714–733.
- (4) Anderson, P. W. *Phys. Rev.* **1950**, *79*, 350–356.
- (5) Anderson, P. W. *Phys. Rev.* **1959**, *115*, 2–13.
- (6) Hay, P. J.; Thibault, J. C.; Hoffmann, R. *J. Am. Chem. Soc.* **1975**, *97*, 4884–4899.
- (7) de Loth, P.; Cassoux, P.; Daudey, J. P.; Malrieu, J. P. *J. Am. Chem. Soc.* **1981**, *103*, 4007–4016.
- (8) Noodleman, L.; Davidson, E. R. *Chem. Phys.* **1986**, *109*, 131–143.
- (9) Calzado, C. J.; Cabrero, J.; Malrieu, J. P.; Caballol, R. *J. Chem. Phys.* **2002**, *116*, 2728–2747.
- (10) Calzado, C. J.; Angeli, C.; Taratiel, D.; Caballol, R.; Malrieu, J.-P. *J. Chem. Phys.* **2009**, *131*, 044327.
- (11) Zener, C. *Phys. Rev.* **1951**, *82*, 403–405.
- (12) Noodleman, L.; Han, W.-G. *J. Biol. Inorg. Chem.* **2006**, *11*, 674–694.
- (13) Noodleman, L. *J. Chem. Phys.* **1981**, *74*, 5737–5743.

- (14) Soda, T.; Kitagawa, Y.; Onishi, T.; Takano, Y.; Shigeta, Y.; Nagao, H.; Yoshioka, Y.; Yamaguchi, K. *Chem. Phys. Lett.* **2000**, *319*, 223–230.
- (15) Sinnecker, S.; Neese, F.; Noodleman, L.; Lubitz, W. *J. Am. Chem. Soc.* **2004**, *126*, 2613–2622.
- (16) Li, J.; Noodleman, L. In *Spectroscopic Methods in Bioinorganic Chemistry*; Solomon, E. I., Hodgson, K. O., Eds.; ACS Symposium Series: Washington, DC, 1998; Vol. 692.
- (17) Nair, N. N.; Schreiner, E.; Pollet, R.; Staemmler, V.; Marx, D. *J. Chem. Theory Comput.* **2008**, *4*, 1174–1188.
- (18) Cabrero, J.; Amor, N. B.; de Graaf, C.; Illas, F.; Caballol, R. *J. Phys. Chem. A* **2000**, *104*, 9983–9989.
- (19) Hubner, O.; Sauer, J. *J. Chem. Phys.* **2002**, *116*, 617–628.
- (20) Dewar, M. J.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- (21) Thiel, W. Semiempirical Methods. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2001; Vol. 3, pp 1–24.
- (22) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–45.
- (23) Stewart, J. J. P. *MOPAC 2000*; Fujitsu Limited, Tokyo, Japan, 1999.
- (24) Dewar, M. J.; Zoebisch, E. G.; Healy, H. F.; Stewart, J. P. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (25) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616–626.
- (26) Pipek, J.; Mezey, P. *J. Chem. Phys.* **1989**, *90*, 4916–4926.
- (27) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*, 1st ed.; Wiley: New York, 2000; pp 523–645.
- (28) Karlstrom, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222.
- (29) Roos, B. O.; Lindh, R.; Malmqvist, P.-A.; Veryazov, V.; Widmark, P.-O. *J. Phys. Chem. A* **2005**, *109*, 6575–6579.
- (30) Bastardis, R.; Guihéry, N.; de Graaf, C. *J. Chem. Phys.* **2008**, *129*, 104102.
- (31) Goodgame, M. M.; Goddard, W. A., III *Phys. Rev. Lett.* **1985**, *54*, 661–664.

CT1001279

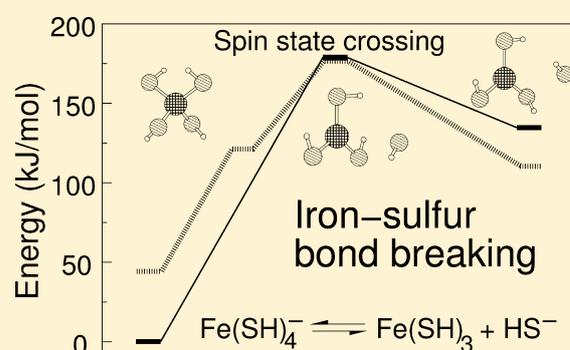
ANEXO D – Ferric-Thiolate Bond
Dissociation Studied with Electronic Structure
Calculations

Ferric–Thiolate Bond Dissociation Studied with Electronic Structure Calculations

Guilherme Menegon Arantes^{*,†} and Martin J. Field[‡][†]Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900, São Paulo, SP Brazil[‡]Institut de Biologie Structurale (IBS), CEA/CNRS/Université Joseph Fourier, 71 Avenue des Martyrs, CS 10090, 38044 Grenoble Cedex 9, France

Supporting Information

ABSTRACT: The stability and reactivity of iron–sulfur clusters are fundamental properties for the biological function of these prosthetic groups. Here, we investigate the ferric–thiolate bond dissociation of model iron–sulfur tetrahedral complexes with high-level *ab initio* multiconfigurational electronic structure calculations. We find that the reaction mechanism is homolytic with a spin-crossing from the sextet state in the reactant to quartet state in the product. We also compare several density functionals and semiempirical configuration interaction with the high-level *ab initio* results to find an accurate but computationally more efficient method to describe the reaction. The functionals M06 and those based on the OPTX exchange functional show the best performance and may reasonably describe the various electron correlation effects involved in ferric–thiolate bond dissociation.



1. INTRODUCTION

Iron under various oxidation states is commonly found in biomolecules tetrahedrally coordinated to thiolate or inorganic sulfide ligands. These iron–sulfur clusters are involved in many complex and essential biochemical processes that depend on redox reactions,^{1–3} such as photosynthesis, cellular respiration, and signaling.

Recently,^{4–6} atomic force microscopy (AFM) and molecular modeling with quantum chemical calculations were used together to probe the stability and reactivity of ferric–thiolate bonds in a simple iron–sulfur protein, rubredoxin. This protein contains one iron center coordinated by the side-chains of four cysteines.⁷ Our aims in the study presented here are to investigate the intrinsic properties of ferric–thiolate bond dissociation, free from protein or solvent interactions, and to find a good balance between accuracy and efficiency in its description by comparing several different electronic structure methods.

Modeling ferric–thiolate dissociation by quantum chemical methods is a difficult problem. The multiplet structure of iron (d^5 in the ferric redox state) and its spin manifold, as well as the bond-breaking process, result in near-degenerate electronic configurations that should, in principle, be described by multiconfigurational methods.⁸ One approach that has been able to correctly describe near-degenerate effects in transition-metal compounds is the complete active space (CAS) self-consistent field (SCF) method,⁹ especially when some dynamic electron correlation is also perturbatively recovered, as in its

CASPT2¹⁰ variant. This method has been successfully applied in a number of studies of iron complexes.^{11–14}

Density functional theory (DFT)¹⁵ has been used extensively to calculate the electronic structure of transition-metal compounds.¹⁶ Given the approximate nature of the available exchange-correlation functionals it is not evident whether the aforementioned electron correlation effects, nondynamic from near-degeneracies and dynamic from Coulomb interactions, are included in a balanced way. As a result, the performance of several functionals has been tested for different iron complexes.^{14,16–27}

It was recognized early that the inclusion of Hartree–Fock (HF) exchange into hybrid DFT functionals increases the relative stability of configurations with unpaired electrons.¹⁷ Consequently, the energy splittings between different spin states of iron–sulfur complexes could be tuned by optimizing the amount of exact exchange added in, for example, the B3LYP functional.¹⁷ It was also shown that the addition of exact exchange underestimates metal–ligand bond energies and that pure-GGA (generalized gradient approximation) functionals overestimate them.¹⁹ A comparison of several metal–ligand diatomics and models of metalloprotein active-sites containing iron centers²⁰ suggested that a good balance could be obtained with the TPSSH functional.²⁸ Another comparison that included

Received: June 13, 2015

Revised: September 8, 2015

Published: September 9, 2015

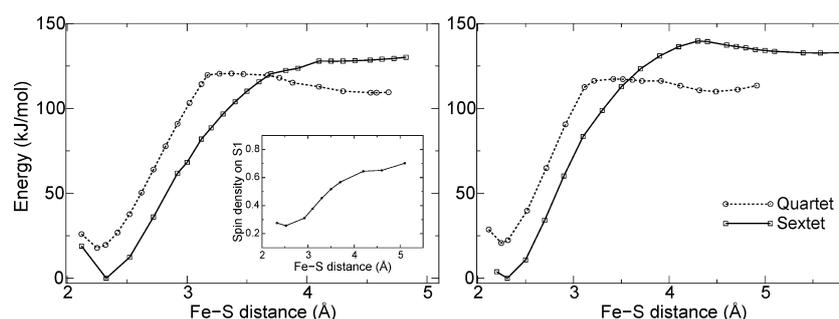
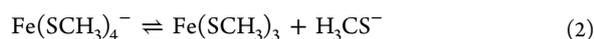


Figure 1. Relative energy profiles for ferric–thiolate bond dissociation obtained at the DFT level. Calculations were carried out at the B3LYP/6-31+G(2df,p) level in the quartet (dashed line) and sextet (solid line) spin states for $\text{Fe}(\text{SH})_4^-$ (left panel) and $\text{Fe}(\text{SCH}_3)_4^-$ (right panel). The inset shows the spin density over the dissociating sulfur atom (S1) for the sextet state.

ferrous and ferric complexes with tetrahedral coordination^{14,18} suggested that functionals such as OLYP and OPBE, based on Handy and Cohen’s OPTX pure-GGA exchange functional,²⁹ are the most computationally cost-effective in describing iron compounds.

Here we investigate ferric–thiolate bond dissociation in the following two isolated model reactions



using several DFT functionals and multiconfigurational *ab initio* (CASPT2) and semiempirical³⁰ methods. Details of the reaction mechanisms, spin states and densities, electronic configurations, and relative energetics are given in the following sections. These results are used to help identify an accurate and efficient treatment of the dissociation reaction.

2. COMPUTATIONAL METHODS

Molecular geometries for species along reactions 1 and 2 were optimized at the DFT level with the B3LYP functional^{31,32} and the 6-31+G(2df,p)³³ basis set for all possible electronic spin states ($S = 0.5, 1.5, \text{ and } 2.5$). Spin-polarized orbitals were adopted. Optimizations conducted with the TZVP basis set³⁴ resulted in equivalent structures. Ferric–thiolate bond dissociation was obtained by optimizing the geometry with the dissociative Fe–S distance fixed at a given value. The GAUSSIAN 09 program (rev. A1)³⁵ was used for geometry optimizations. It should be noted that the calculated Fe–S bond length for the reactant ground state (2.32 Å) is in good agreement with the experimental bond lengths observed in crystallographic structures of proteins that contain tetrahedral Fe–S centers. As an example, in rubredoxin (PDB ID 8RXN), the bond lengths are in the range of 2.27–2.30 Å.

A minimum-energy crossing point (MECP) between sextet and quartet electronic states was optimized with the pDynamo library³⁶ as described previously⁵ with the B3LYP functional and the TZVP basis set.

Single-point DFT energies for geometries corresponding to the CASPT2 low-energy pathway described below were obtained with the ORCA program version 3.0.1³⁷ and the following functionals: OLYP, O3LYP,^{29,32} OPBE,^{29,38} PBE0,³⁹ BLYP,^{32,40} B3LYP, B3LYP-D3, BP86,^{40,41} TPSS, TPSSH,²⁸ M06,⁴² and B2PLYP.⁴³ Spin-polarized orbitals and standard integration grids were adopted. The Def2-TZVP basis set⁴⁴ and resolution of identity with the TZV/J⁴⁵ auxiliary basis were

used. A second-order SCF optimization had to be activated to obtain convergence in several cases.

Although the $\text{Fe}(\text{SH})_4^-$ ground state (reactant in sextet electronic spin) belongs to the S_4 molecular point group, the tetrahedral symmetry around the iron center is broken in all other geometries studied here. Thus, all calculations were carried out in the C_1 point-group symmetry.

Single-point multiconfigurational calculations were carried out with the geometries optimized by DFT along the ferric–thiolate bond dissociation. The active space was not built using an analysis of the corresponding irreducible representations (irreps) and symmetry adapted linear combinations of ligand molecular orbitals. Instead, it was chosen with five molecular orbitals for the metal 3d open-shell, as these are necessary to describe the near-degenerate configurations arising from the ferric ion multiplet structure, and two extra pairs of correlating molecular orbitals (MOs) describing the bond dissociation process. This active space contained nine electrons in nine orbitals. Similar choices, such as seven electrons in seven MOs, were also tried and yielded equivalent results. The *double-shell* effect¹¹ was tested for the ferric ion complexes studied here with an active space of nine electrons in fourteen orbitals and found to be negligible (Table S3).

All multiconfigurational CASSCF and CASPT2^{8,10} calculations were performed with the MOLCAS program, version 7.4.⁴⁶ The ANO-RCC basis-set⁴⁷ with contractions Fe-[7s6p5d2f1g], S[6s5p3d2f], and H[3s2p1d] was employed as this was shown to be appropriate in previous benchmark calculations for iron compounds.¹² CASSCF orbitals were optimized for the average of the first three roots with equal weights. Scalar relativistic effects were included using a Douglas–Kroll–Hess Hamiltonian. In all CASPT2 calculations the core electrons were frozen (26 frozen MOs), and only the valence space was correlated. The standard IPEA shift of 0.25 au and an extra denominator shift of 0.3 au were used to avoid intruder states.^{10,48}

To test if semiempirical potentials could be used to model ferric–thiolate bond dissociation, the performance of the modern PM6 parametrization³⁰ that contains d-orbitals in its minimal basis set was investigated with complete active space configuration interaction (CASCI) calculations.⁸ Several active spaces were tested for these semiempirical calculations, and the best results were obtained with seven electrons in seven MOs space. Orbitals were determined from restricted-open shell (ROHF) calculations with fractional occupation.^{49,50}

The performance of the original PM6 parametrization was very poor (see Results and Discussion); therefore, a specific

reparametrization for iron was carried out. CASPT2 energies reported here for Fe–S bond dissociation, iron atomic multiplet and ionization energies taken from spectroscopical data,⁵¹ and DFT energies for different $\text{Fe}(\text{SH})_4^-$ and $\text{Fe}(\text{SCH}_3)_4^-$ geometries (not shown) were used as target or reference properties. Optimized parameters are given in the Supporting Information (Table S1). Full details of the reparametrization procedure are not relevant for the conclusions drawn here and will be given in a future publication. All semiempirical calculations were done with the pDynamo library.³⁶

3. RESULTS AND DISCUSSION

3.1. Energy Profiles for Bond Dissociation. Our calculation results for reaction 1 show the ground-state reactant is a sextet spin state and has a tetrahedral geometry belonging to the S_4 point group. The quartet and doublet reactants have distorted tetrahedral coordination around the iron center with overall C_2 symmetry. All other geometries along the Fe–S dissociation belong to the C_1 point group.

The DFT energy profiles shown in Figure 1 for both $\text{Fe}(\text{SH})_4^-$ and $\text{Fe}(\text{SCH}_3)_4^-$ dissociation have several noteworthy features. First, the quartet reactant lies 20 kJ/mol higher in energy than the sextet reactant. However, the lowest-energy product of bond dissociation is a quartet. Thus, there is a spin crossing along the reaction. A minimum-energy crossing point was optimized at a ferric–thiolate bond distance $d(\text{Fe}–\text{S}) \approx 3.7$ Å. Products in all spin states are ion–molecule complexes with small (<4 kJ/mol) recombination barriers. Also, the quartet transition-state region is reached before the MECP, whereas the sextet transition-state region is reached after the MECP. The doublet state energy profile is not shown nor discussed in detail here as all doublet stationary geometries along ferric–thiolate bond dissociation lie at least 90 kJ/mol higher than the other spin states.

The ferric–thiolate bond dissociation mechanism can be deduced from the DFT profiles. Although the reactants have a formally ferric center, both sextet and quartet products have a ferrous center with a dissociated thiolate radical. To illustrate this, the inset in Figure 1 shows that the spin density in the dissociative sulfur atom (S1) changes rather smoothly from 0 to 1 during Fe–S bond dissociation. As a result, Fe–S bond dissociation in both sextet and quartet spin states is homolytic.⁵

The energetics, iron coordination, and Fe–S bond distances of stationary points are very similar between the $\text{Fe}(\text{SH})_4^-$ and $\text{Fe}(\text{SCH}_3)_4^-$ dissociation profiles. Thus, the conclusions drawn below for $\text{Fe}(\text{SH})_4^-$ should be equally valid for $\text{Fe}(\text{SCH}_3)_4^-$ when comparing the different CASPT2, DFT, and semiempirical methods. The only caveat concerns the product region as different $d(\text{Fe}–\text{S})$ distances are observed because of the bulkier methyl groups in $\text{Fe}(\text{SCH}_3)_4^-$ (Figure 1). A related observation should also be made for the products of ferric–thiolate bond cleavage in $\text{Fe}(\text{SH})_4^-$ optimized with the OPBE and TPSSH functionals (data not shown). For these two functionals, the products resulted in H_2S instead of HS^- , as this leaving group removes a proton from another thiol group at $d(\text{Fe}–\text{S}) > 4$ Å.

The multiconfigurational CASPT2 energy profile for $\text{Fe}(\text{SH})_4^-$ dissociation is shown in Figure 2. The energies and structures reported in this figure correspond to the CASPT2 minimum-energy pathway for ferric–thiolate bond dissociation with DFT optimized geometries. These six structures, along with the doublet product and reactant structures (not shown in

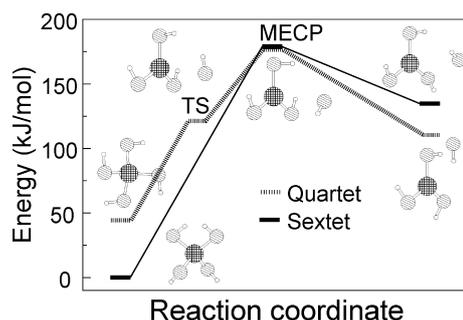


Figure 2. Relative energy profiles for $\text{Fe}(\text{SH})_4^-$ ferric–thiolate bond dissociation obtained at the CASPT2 level. Calculations were carried out with the ANO-RCC basis set for the quartet (dashed line) and sextet (solid line) spin states. Structures are shown in ball–stick representation with iron in crossed lines and sulfur in diagonal lines.

the figure), will be termed in what follows as reactants (R), transition state (TS), MECP, and products (P), together with their respective total spin electronic state, which can have doublet (2), quartet (4), and sextet (6) multiplicities.

The reaction coordinates of the low-energy species along the CASPT2 minimum-energy pathway are similar to those found with DFT. However, we were unable to identify a ^6TS with the CASPT2 single-point calculations because of the rather flat transition region observed in the DFT profile [$d(\text{Fe}–\text{S}) > 4$ Å in the left panel of Figure 1]. The relative energies for the quartet and sextet profiles found with CASPT2 differ at most by 20 kJ/mol from those obtained with B3LYP for all species except the MECP (Table S2). This is the accuracy (~ 20 kJ/mol) often expected of hybrid functionals in isodesmic processes involving metals, such as heterolytic bond dissociations.¹⁹ However, at the crossing point, the difference in relative energy between the B3LYP and CASPT2 energy profiles is higher (50 kJ/mol), which can be partially attributed to a structural relaxation energy not included in our CASPT2 single-point calculations. Although the $^6\text{MECP}$ and $^4\text{MECP}$ CASPT2 energies differ by less than 4 kJ/mol, indicating that they might be near a surface crossing, the structure still could be far from the true minimum-energy crossing of the CASPT2 energy surfaces. Further comparisons between the multiconfigurational results and several DFT functionals will be given in the next section.

A homolytic reaction mechanism with a spin crossing along the dissociation is also observed in the CASPT2 results. The reactant is in a sextet state, and the most stable product is in the quartet state containing a partial thiolate radical and a ferrous center. However, instead of changing smoothly as in the monoconfigurational DFT calculations (inset in Figure 1), CASSCF spin-densities change abruptly during dissociation as a consequence of hopping between surfaces with different electronic configurations. As shown in Table 1 for the reactant states, almost no spin density is seen on sulfur atoms and the spin density on the iron atom equates to the expected number of unpaired electrons in the spin state. For the MECP, spin densities correspond to different electronic configurations, and spin unpairing is already evident on the dissociating sulfur atom (S1). In fact, the crossing between the sextet and quartet spin states is equivalent to a spin flipping over this sulfur atom. Electronic configurations and, consequently, spin densities are similar between the MECP and product states for both sextet and quartet spin states.

Table 1. Mulliken Spin Populations Calculated from the CASSCF Wavefunction for Fe and S Atoms in Structures from Reaction 1^a

	Fe	S1	S2	S3	S4
² R	1.00	-0.03	0.02	-0.02	0.02
² P	1.34	-0.29	-0.01	-0.00	-0.04
⁴ R	3.11	-0.02	-0.03	-0.03	-0.02
⁴ TS	3.50	-0.50	-0.07	0.03	0.03
⁴ MECP	3.52	-0.52	0.03	0.03	-0.05
⁴ P	3.50	-0.53	0.03	0.03	-0.05
⁶ R	4.75	0.06	0.06	0.06	0.06
⁶ MECP	3.91	0.86	0.03	0.03	0.14
⁶ P	3.91	0.84	0.03	0.18	0.02

^aThe spin density moduli were less than 0.02 for all hydrogen atoms in all structures.

In the calculations for ⁶R employing spatially restricted orbitals (ROHF and CASSCF), we observe a typical molecular orbital ordering equivalent to what is usually expected from *d*⁵ metal centers with tetrahedral coordination.⁵² The valence orbitals of the sulfur ligand combine into nonbonding MOs belonging approximately to the *T*₁ irrep in the *T*_d local symmetry around the iron center and into bonding MOs belonging to irreps *A*₁, *E*, and *T*₂. The five unpaired electrons of the metal are distributed in *e*² *t*₂³-like orbitals, with contributions of about 5–10% weight from ligand orbitals from irreps *E* and *T*₂. For the other geometries along the reaction pathway, the local symmetry around the iron center is distorted and it is more difficult to label the symmetry of the MOs. The ferric–thiolate bond that is broken upon dissociation has a σ character.³ All the occupied bonding and nonbonding ligand orbitals lie lower in energy than the metal *d*-shell. This is in contrast to the orbital ordering observed in our DFT unrestricted calculations in which an inverted level-scheme is found due to strong spin polarization of the MOs. The MO ordering in the inverted-level scheme shows the metal *d*-shell below the ligand bonding and nonbonding MOs. This has also been observed in similar iron–sulfur compounds and extensively discussed by Noodleman et al.,^{53–55} Solomon et al.,⁵⁶ and more recently by Ichiye and co-workers.^{22,57}

Wave functions for all species in the sextet state show a single configuration with more than 98% of the weight in the CI expansion. The other two reactant states, ⁴R and ²R, are dominated by a single configuration with \approx 80% weight in the CI expansion, and the remaining configurations account for different occupations of the metal *d*-shell. All other quartet and doublet species have large mixtures of configurations without any surpassing 20% weight. These results suggest that monoconfigurational electronic structure methods may be able to describe the bond-breaking process properly in only the sextet state.

3.2. Comparison with DFT Functionals and Semiempirical Methods. To test the performance of more approximate electronic structure methods in the description of ferric–thiolate bond dissociation, we compared the CASPT2 relative energies with a series of DFT functionals and semiempirical methods. Table 2 shows the deviations and maximum errors obtained, and Table S2 in the Supporting Information gives the relative energy values. We note that the estimated accuracy of CASPT2 to describe spin-splittings is \sim 12 kJ/mol.^{12,14}

Table 2. Comparison of Several DFT Functionals and Semiempirical Methods with the CASPT2 Reference^a

method	STDEV	MAE
B2PLYP	38.4	108.6
B3LYP	39.9	113.7
B3LYP-D3	38.4	105.2
BLYP	54.2	143.5
BP86	49.2	134.7
M06	17.8	31.8
O3LYP	25.4	83.1
OLYP	19.5	75.0
OPBE	26.9	83.8
PBE0	32.6	92.8
PM6 ^b	79.7	319.7
PM6R ^b	25.2	67.5
TPSS	52.8	143.6
TPSSh	46.4	122.5

^aStandard deviations (STDEV) and maximum absolute errors (MAE) of relative energies (in kilojoules per mole) for structures from the Fe(SH)₄ ferric–thiolate bond dissociation pathways. Values obtained from the relative energies (Table S2) of nine single-point calculations, corresponding to the six structures shown in Figure 2 (both quartet and sextet states were calculated for the MECP), ²R and ²P. ^bPM6 is the original semiempirical parametrization and PM6R is the parametrization obtained in this work (Table S1).

An analysis of the data shows that the best performing functional is the M06 hybrid meta-GGA, followed by the simpler OLYP pure-GGA functional, although with a much larger maximum error. The doublet states have the highest relative energies and give the maximum errors for all functionals except B2PLYP and M06. The OLYP functional is followed closely by OPBE and O3LYP, suggesting that Handy and Cohen's OPTX exchange functional²⁹ is the most appropriate to describe Fe–S bond dissociation. This exchange functional includes some nondynamic electron correlation effects (named "left–right correlation")²⁹ that contribute to the correct description of homolytic bond dissociation. On the other hand, Becke's one-parameter exchange functional,⁴⁰ used in BLYP and BP86, performs significantly worse. This is because it stays closer to the uniform electron gas assumption (the Dirac coefficient is not scaled)²⁹ and does not include such left–right nondynamic correlation effects.

Apart from the highly parametrized M06, meta-GGA functionals, such as TPSS, do not perform particularly well. Inclusion of Hartree–Fock exchange does not markedly improve the results for TPSSh in comparison to TPSS. By contrast, hybrid functionals, such as B3LYP and PBE0, do show lower errors, especially for energy splittings between different spin states. It should be noted that the pure GGA functionals not based on the OPTX exchange functional, such as BLYP, BP86, and TPSS, give a qualitatively wrong description of the spin splitting, as ⁴R is more stable than ⁶R (Table S2). Mixing some exact exchange with the OPTX functional as in O3LYP does not improve the spin splittings in comparison to the pure-GGA OLYP. The double-hybrid functional B2PLYP does not perform significantly better, suggesting that the inclusion of dynamic correlation effects via an MP2 contribution is not important for the description of ferric–thiolate bond dissociation. Finally, inclusion of dispersion corrections as in the B3LYP-D3 functional does not change the relative energies, in comparison to the B3LYP functional.

We conclude that the hybrid M06 functional and the functionals based on Handy and Cohen's OPTX exchange give the best results for iron–sulfur bond dissociation. If computational efficiency is important, then the pure-GGA functionals OLYP and OPBE are to be preferred.

Given that we are particularly interested in the simulation of iron–sulfur clusters in complex systems, we wanted to evaluate methods that are computationally more efficient than CASPT2 and DFT. As an initial choice, we tried the recent semiempirical Hamiltonian PM6³⁰ that contains d-orbitals in its minimal basis-set and was parametrized for molecules containing iron. These tests, using unrestricted HF (UHF) and ROHF + CASCI calculations and the original PM6 Hamiltonian, resulted in a very poor description of ferric–thiolate bond dissociation energetics (Table 2 and Table S2, which show the ROHF + CASCI results only). This poor performance is due to an incorrect MO ordering, as the orbital formed mainly by the Fe 4s-shell was occupied and lower in energy than the MOs composed by the Fe 3d open-shell, suggesting an imbalance between the semiempirical ζ_s and ζ_d orbital coefficients. Our previous experience⁵⁸ indicates that this imbalance can be corrected by a specific reparametrization of the iron PM6 parameters. This we did successfully, leading to a parametrization that we denote PM6R (Table S1).

ROHF + CASCI calculations with the new PM6R Hamiltonian perform rather well, with deviations similar to functionals based on OPTX exchange, but with smaller maximum errors (Table 2). However, this is not a fair comparison because the CASPT2 reference energies were actually used to fit the PM6R parameters. The observed MO orbital ordering is indeed similar to that obtained for the ab initio CASSCF wave functions with, in increasing order of energy, doubly occupied ligand MOs, singly occupied MOs corresponding to the Fe 3d shell, and an unoccupied Fe 4s shell. The resulting spin densities are consequently more accurate and close to those shown in Table 1. However, it should be noted that UHF calculations with the new parameters yielded poor results. Thus, although our specifically parametrized semiempirical Hamiltonian, in conjunction with CI calculations, can describe the energetics and electronic configuration of ferric–thiolate bond dissociation rather well, the transferability and applicability of these new parameters in other circumstances needs to be further validated.

4. CONCLUSIONS

We have studied ferric–thiolate bond dissociation in the isolated model iron–sulfur tetrahedral compounds, $\text{Fe}(\text{SH})_4^-$ and $\text{Fe}(\text{SCH}_3)_4^-$, using the multiconfigurational wave function methods, ab initio CASPT2 and semiempirical CASCI, and several DFT functionals. The reaction proceeds via a homolytic mechanism with a spin crossing between the sextet and quartet states. This two-state reactivity is often seen in transition-metal compounds,⁵⁹ and so studies of the stability, reactivity, and biosynthesis of iron–sulfur clusters in proteins need to carefully account for the different accessible spin states.

The high-level CASPT2 method was used as a reference, against which the various other electronic structure approaches were compared. We find that the M06 functional is the best for describing Fe–S bond dissociation when all possible spin states are to be considered. Our DFT comparisons are also in agreement with previous studies¹⁴ that highlight the reliability of the OLYP and OPBE functionals for describing iron

complexes. The B3LYP functional gives satisfactory results, and it should be fairly dependable for geometry optimizations.

Ferric–thiolate bond dissociation is a difficult problem to model as it involves different kinds of electron correlation effects that change unevenly as the reaction progresses. Near-degenerate nondynamic correlation is important for describing the multiplet structure of the ferric center and, consequently, the energy splittings between the various spin states. By contrast, left–right nondynamic correlation is critical for the description of homolytic bond dissociation, whereas angular correlation in the transition metal also needs to be considered appropriately, given the change in metal coordination from tetrahedral reactants to trigonal products. As a result, this reaction constitutes a tough test for electronic structure methods, in particular for the single-determinant, principally DFT, approaches tried here.

We conclude by noting that specifically parametrized semiempirical methods employing d-orbitals and CI calculations can be used to model the Fe–S dissociation reaction, but further studies are necessary to evaluate the general applicability of this approach.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.5b05658.

Tables containing the reoptimized PM6 iron semiempirical parameters and the relative energies calculated with all the electronic structure methods used here (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: garantes@iq.usp.br.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding from FAPESP (12/02501-4 and 14/21900-2) and a joint FAPESP-CNRS project (23277) are gratefully acknowledged.

■ REFERENCES

- (1) Beinert, H.; Holm, R. H.; Munck, E. Iron-Sulfur Clusters: Nature's Modular, Multipurpose Structures. *Science* **1997**, *277*, 653–659.
- (2) Holm, R. H.; Kennepohl, P.; Solomon, E. I. Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev.* **1996**, *96*, 2239–2314.
- (3) Solomon, E. I.; Gorelsky, S. I.; Dey, A. Metal-Thiolate Bonds in Bioinorganic Chemistry. *J. Comput. Chem.* **2006**, *27*, 1415–1428.
- (4) Zheng, P.; Li, H. Highly Covalent Ferric-Thiolate Bonds Exhibit Surprisingly Low Mechanical Stability. *J. Am. Chem. Soc.* **2011**, *133*, 6791–6798.
- (5) Arantes, G. M.; Bhattacharjee, A.; Field, M. J. Homolytic Cleavage of Fe-S Bonds in Rubredoxin Under Mechanical Stress. *Angew. Chem., Int. Ed.* **2013**, *52*, 8144–8146.
- (6) Zheng, P.; Arantes, G. M.; Field, M. J.; Li, H. Force Induced Chemical Reactions on the Metal Center in a Single Metalloprotein Molecule. *Nat. Commun.* **2015**, *6*, 7569.
- (7) Beinert, H. Iron-Sulfur Proteins: Ancient Structures, Still Full of Surprises. *JBIC, J. Biol. Inorg. Chem.* **2000**, *5*, 2–15.
- (8) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*, 1st ed.; Wiley: New York, 2000.

- (9) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A Complete Active Space SCF Method (CASSCF) Using a Density Matrix Formulated Super-CI Approach. *Chem. Phys.* **1980**, *48*, 157–173.
- (10) Andersson, K.; Malmqvist, P. A.; Roos, B. O. Second-Order Perturbation Theory with a Complete Active Space Self-Consistent Field Reference Function. *J. Chem. Phys.* **1992**, *96*, 1218.
- (11) Pierloot, K. *Nondynamic Correlation Effects in Transition Metal Coordination Compounds*, 1st ed.; Marcel Dekker: New York, 2001; pp 123–158.
- (12) Pierloot, K.; Vancoillie, S. Relative Energy of the High- $(^5T_{2g})$ and Low- $(^1A_{1g})$ Spin States of $[\text{Fe}(\text{H}_2\text{O})_6]^{2+}$, $[\text{Fe}(\text{NH}_3)_6]^{2+}$, and $[\text{Fe}(\text{bpy})_3]^{2+}$: CASPT2 Versus Density Functional Theory. *J. Chem. Phys.* **2006**, *125*, 124303.
- (13) Arantes, G. M.; Taylor, P. R. Approximate Multiconfigurational Treatment of Spin-Coupled Metal Complexes. *J. Chem. Theory Comput.* **2010**, *6*, 1981–1989.
- (14) Swart, M. Accurate Spin-State Energies for Iron Complexes. *J. Chem. Theory Comput.* **2008**, *4*, 2057–2066.
- (15) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*, 1st ed.; Oxford University Press: Oxford, 1996.
- (16) Cramer, C. J.; Truhlar, D. G. Density Functional Theory for Transition Metals and Transition Metal Chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757–10816.
- (17) Reiher, M.; Salomon, O.; Hess, B. A. Reparameterization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48–55.
- (18) Swart, M.; Groenhof, A. R.; Ehlers, A. W.; Lammertsma, K. Validation of Exchange-Correlation Functionals for Spin States of Iron Complexes. *J. Phys. Chem. A* **2004**, *108*, 5479–5483.
- (19) Jensen, K. P.; Roos, B. O.; Ryde, U. Performance of Density Functionals for First Row Transition Metal Systems. *J. Chem. Phys.* **2007**, *126*, 014103.
- (20) Jensen, K. P. Bioinorganic Chemistry Modeled with the TPSSH Density Functional. *Inorg. Chem.* **2008**, *47*, 10357–10365.
- (21) Sorkin, A.; Iron, M. A.; Truhlar, D. G. Density Functional Theory in Transition-Metal Chemistry: Relative Energies of Low-Lying States of Iron Compounds and the Effect of Spatial Symmetry Breaking. *J. Chem. Theory Comput.* **2008**, *4*, 307–315.
- (22) Niu, S.; Nichols, J. A.; Ichiye, T. Optimization of Spin-Unrestricted Density Functional Theory for Redox Properties of Rubredoxin Redox Site Analogues. *J. Chem. Theory Comput.* **2009**, *5*, 1361–1368.
- (23) Bergeler, M.; Stiebritz, M. T.; Reiher, M. Structure-Property Relationships of Fe₄S₄ Clusters. *ChemPlusChem* **2013**, *78*, 1082–1098.
- (24) Bruska, M. K.; Stiebritz, M. T.; Reiher, M. Analysis of Differences in Oxygen Sensitivity of Fe-S Clusters. *Dalton Trans.* **2013**, *42*, 8729–8735.
- (25) Ye, S.; Neese, F. Accurate Modeling of Spin-State Energetics in Spin-Crossover Systems with Modern Density Functional Theory. *Inorg. Chem.* **2010**, *49*, 772–774.
- (26) Isley, W. C., III; Zarra, S.; Carlson, R. K.; Bilbeisi, R. A.; Ronson, T. K.; Nitschke, J. R.; Gagliardi, L.; Cramer, C. J. Predicting Paramagnetic ¹H NMR Chemical Shifts and State-Energy Separations in Spin-Crossover Host-Guest Systems. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10620–10628.
- (27) Ferre, N.; Guihery, N.; Malrieu, J.-P. Spin Decontamination of Broken-Symmetry Density Functional Theory Calculations: Deeper Insight and New Formulations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14375–14382.
- (28) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- (29) Handy, N. C.; Cohen, A. J. Left-Right Correlation Energy. *Mol. Phys.* **2001**, *99*, 403–412.
- (30) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.
- (31) Becke, A. D. Density Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.
- (32) Lee, C.; Yang, W.; Parr, R. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (33) Ditchfield, R.; Hehre, W.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (34) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to K. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (36) Field, M. J. The pDynamo Program for Molecular Simulations using Hybrid Quantum Chemical and Molecular Mechanical Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1151–1161.
- (37) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.
- (38) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (39) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (40) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098–3100.
- (41) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822–8824.
- (42) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Non-covalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (43) Schwabe, T.; Grimme, S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.
- (44) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (45) Weigend, F. Accurate Coulomb-Fitting Basis Sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.
- (46) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-A.; Neogrády, P.; Pedersen, T. B.; Pitonak, M.; Reiher, M.; Roos, B. O.; et al. MOLCAS 7: The Next Generation. *J. Comput. Chem.* **2010**, *31*, 224–247.
- (47) Roos, B. O.; Lindh, R.; Malmqvist, P.-A.; Veryazov, V.; Widmark, P.-O. New Relativistic ANO Basis Sets for Transition Metal Atoms. *J. Phys. Chem. A* **2005**, *109*, 6575–6579.
- (48) Roos, B. O.; Andersson, K.; Fülscher, M. P.; Serrano-Andrés, L.; Pierloot, K.; Merchan, M.; Molina, V. Applications of Level Shift Corrected Perturbation Theory in Electronic Spectroscopy. *J. Mol. Struct.: THEOCHEM* **1996**, *388*, 257–276.
- (49) Granucci, G.; Toniolo, A. Molecular Gradients for Semiempirical CI Wavefunctions with Floating Occupation Molecular Orbitals. *Chem. Phys. Lett.* **2000**, *325*, 79–85.
- (50) Slavíček, P.; Martínez, T. J. Ab Initio Floating Occupation Molecular Orbital-Complete Active Space Configuration Interaction: An Efficient Approximation to CASSCF. *J. Chem. Phys.* **2010**, *132*, 234102.
- (51) Ralchenko, Y.; Kramida, A. E.; Reader, J. NIST ASD Team, NIST Atomic Spectra Database (version 3.1.5). National Institute of Standards and Technology, Gaithersburg, MD, <http://physics.nist.gov/asd3> (accessed March 7, 2012).

- (52) Butcher, K. D.; Didziulis, S. V.; Briat, B.; Solomon, E. I. Variable Photon Energy Photoelectron Spectroscopy on FeCl_4^- : An Unusual Electronic Structure for High-Spin d5 Complexes. *J. Am. Chem. Soc.* **1990**, *112*, 2231–2242.
- (53) Noodleman, L.; Norman, J. G., Jr; Osborne, J. H.; Aizman, A.; Case, D. A. Models for Ferredoxins: Electronic Structures of Iron-Sulfur Clusters with One, Two, and Four Iron Atoms. *J. Am. Chem. Soc.* **1985**, *107*, 3418–3426.
- (54) Noodleman, L.; Peng, C.; Case, D.; Mouesca, J.-M. Orbital Interactions, Electron Delocalization and Spin Coupling in Iron-Sulfur Clusters. *Coord. Chem. Rev.* **1995**, *144*, 199–244.
- (55) Noodleman, L.; Han, W.-G. Structure, Redox, pKa, Spin. A Golden Tetrad for Understanding Metalloenzyme Energetics and Reaction Pathways. *JBIC, J. Biol. Inorg. Chem.* **2006**, *11*, 674–694.
- (56) Solomon, E. I.; Randall, D. W.; Glaser, T. Electronic Structures of Active Sites in Electron Transfer Metalloproteins: Contributions to Reactivity. *Coord. Chem. Rev.* **2000**, *200*, 595–632.
- (57) Niu, S.; Wang, X.-B.; Yang, X.; Wang, L.-S.; Ichiye, T. Mechanistic Insight into the Symmetric Fission of $[\text{4Fe-4S}]$ Analogue Complexes and Implications for Cluster Conversions in Iron-Sulfur Proteins. *J. Phys. Chem. A* **2004**, *108*, 6750–6757.
- (58) Arantes, G. M.; Loos, M. Specific Parametrisation of a Hybrid Potential to Simulate Reactions in Phosphatases. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347–353.
- (59) Schroder, D.; Shaik, S.; Schwarz, H. Two-State Reactivity as a New Concept in Organometallic Chemistry. *Acc. Chem. Res.* **2000**, *33*, 139–145.

ANEXO E – Homolytic cleavage of Fe-S bonds in Rubredoxin under mechanical stress

Homolytic Cleavage of Fe–S Bonds in Rubredoxin under Mechanical Stress**

Guilherme M. Arantes,* Anirban Bhattacharjee, and Martin J. Field

The biosynthesis, stability, and folding pathways of the iron–sulfur proteins^[1,2] can be probed by analyzing protein unfolding and Fe–S bond dissociation. Some of the simplest iron–sulfur proteins belong to the rubredoxin family, which participate in electron transfer processes in bacteria and archaea. Rubredoxins lack inorganic sulfide and have only one Fe center coordinated by the side-chains of four cysteine residues^[2] (Figure 1 a and b). Recent rubredoxin unfolding experiments, performed by protein engineering and single-molecule atomic-force microscopy,^[3,4] indicated that their Fe–S bonds had an unexpectedly low stability under mechanical stress. It was suggested that the activation process for bond

dissociation^[3] could occur by heterolytic fission, as observed for disulfide bridges,^[5] or homolytic cleavage, as hypothesized for the rupture of C–Si bonds.^[6] Given that Fe complexes can have near-degenerate levels with different total spin,^[7] an understanding of the reactivity of iron–sulfur clusters requires that their spin states and spin crossovers be characterized during the reaction.

Herein we have adopted a multiscale modeling approach employing quantum chemical (QC), molecular mechanical (MM), and hybrid QC/MM potentials, to address these questions. Full details of the models and programs used are given in the Supporting Information. We started by studying the mechanical unfolding of rubredoxin using an MM potential with an implicit model of solvent. A standard biomolecular force field was employed, except for the Fe–S bonds of the iron–sulfur complex which were represented by specially parameterized Morse potentials that permitted bond dissociation. The starting structures for all simulations were derived from those of the oxidized rubredoxin from *Pyrococcus furiosus* (protein databank (PDB) codes 1BRF and 1CAA^[9,10]). Unfolding was emulated by performing molecular dynamics (MD) simulations of the protein with an added harmonic potential that pulled apart the N- and C-termini at a constant speed.

Although we imposed no bias on the order of dissociation of the Fe–S bonds, we observed that the Fe–S(Cys5) bond ruptured first in most of our simulations. Results of one of these simulations carried out with a pulling speed of 10 nm ns⁻¹ are shown in Figure 2. It is clear from the C_α RMSDs that rubredoxin unfolds during the trajectory. The first contacts disrupted are interchain hydrogen bonds in the anti-parallel β-strand formed between the N- and C-termini. Other local polar contacts that stabilize the secondary structure are approximately maintained throughout the simulation, but fluctuations of the hydrophobic contacts and hydrogen bonds in the protein's core result in progressive disorganization of the globular fold. In the last part of the trajectory, the hydrogen bonds that hold together the N-terminal anti-parallel β-strand are broken. Large fluctuations are observed for both the Fe–S_γ bond distances (± 0.5 Å) and S_γ-Fe–S_γ valence angles (± 20°) along the trajectory, until the Fe–S_γ(Cys5) bond breaks after about 1 ns of simulation (Supporting Information Figure S1).^[8]

A snapshot of one of the unfolded rubredoxin structures obtained just before Fe–S_γ bond disruption is shown in Figure 1 c. The local symmetry around the iron–sulfur center is broken upon mechanical unfolding with the iron tetrahedral coordination (pseudo-*T*_d) changing to C₂. This step is characterized by changes in the S_γ-Fe–S_γ valence angles from approximately 109° in folded rubredoxin to 90–125° in the

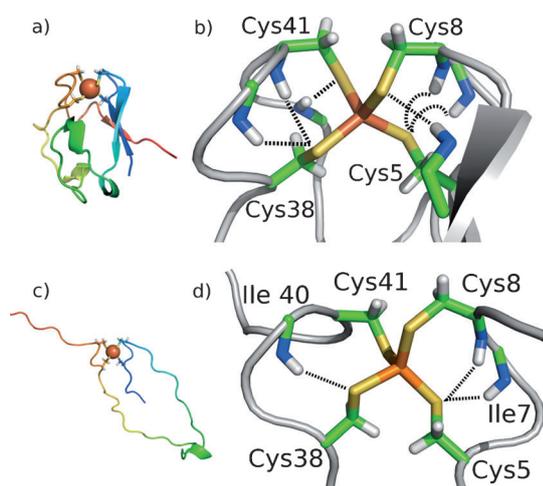


Figure 1. Rubredoxin structures: a) schematic representation of the folded protein; b) close-up of the iron–sulfur cluster from the folded protein; c) schematic representation of the protein from one of the pulling simulations just before Fe–S bond rupture; and d) close-up of the iron–sulfur cluster from the structure shown in (c). In (b) and (d) important hydrogen bonds are indicated by dashed lines and Fe orange, S yellow, C green, N blue.

[*] Dr. G. M. Arantes
Departamento de Bioquímica, Instituto de Química
Universidade de São Paulo
Av. Lineu Prestes 748, 05508-900, São Paulo, SP (Brazil)
E-mail: garantes@iq.usp.br
Dr. A. Bhattacharjee, Dr. M. J. Field
Institut de Biologie Structurale (IBS) Jean-Pierre Ebel
CEA/CNRS/Universite Joseph Fourier
41, rue Jules Horowitz, 38027 Grenoble, Cedex 1 (France)

[**] We acknowledge funding from FAPESP (projects 07/52772-6 and 12/02501-4) and a joint FAPESP-CNRS project (23277).

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201303462>.

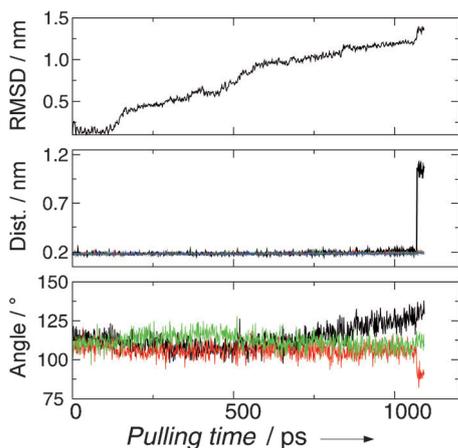


Figure 2. Structural changes during a rubredoxin pulling simulation. Top: the protein C_{α} root mean squared deviation (RMSD) from the initial folded structure; middle: the Fe–S distances for the Cys5 (black), Cys8 (red), Cys38 (green), and Cys41 (blue) side chains; bottom: the $S_{\gamma}(\text{Cys5})\text{--Fe--}S_{\gamma}$ valence angles for the Cys8 (black), Cys41 (red), and Cys38 (green) side chains.

unfolded protein (Tables S3 and S4).^[8] Hydrogen bonds in the vicinity of the iron–sulfur center are also significantly perturbed upon unfolding. The four backbone hydrogen bonds ($\text{O}\cdots\text{HN}$) involving Cys5 and Cys38 in the folded rubredoxin are disrupted as are three of the six hydrogen bonds between the protein backbone and the four S_{γ} s, with only the $S_{\gamma}(\text{Cys5})\cdots\text{NH}(\text{Cys8})$, $S_{\gamma}(\text{Cys5})\cdots\text{NH}(\text{Ile7})$, and $S_{\gamma}(\text{Cys38})\cdots\text{NH}(\text{Ile40})$ interactions maintained in the stretched protein (Figure 1 d and Tables S3 and S4^[8]).

To assess the effect of the rubredoxin environment on Fe–S stretching, we probed the intrinsic stability of Fe^{III} –thiolate bonds using density functional theory (DFT) calculations of the gas-phase model reaction [Eq. (1)].

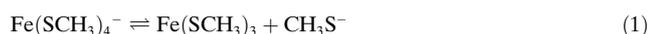


Figure 3a shows the energy profile for this reaction and Table S1^[8] some relevant geometrical parameters. From these it can be seen that the reactant, $\text{Fe}(\text{SCH}_3)_4^-$, has a sextet ground state and S_4 symmetry with local tetrahedral iron coordination, whereas the product, in which one Fe–S bond is broken, has a quartet ground state and C_2 symmetry. Doublet states were also considered but they were at least 50 kJ mol^{-1} higher in energy. The sextet profile has a late transition state (TS) whereas the quartet has an early TS. A minimum energy crossing point (MECP)^[11] between the energy surfaces of the two spin states was found just after the quartet TS along the reaction coordinate. This is the highest energy point along the pathway that connects sextet reactants to quartet products. The energetics of this MECP as well as the efficiency of spin crossover (given mostly by the magnitude of the spin-orbit coupling) are largely responsible for determining the kinetics of the intrinsic Fe–S dissociation reaction.

Spin populations for stationary points along the path are shown in Table S2^[8] which indicates that both quartet and sextet reactants may be described as formal $\text{Fe}^{\text{III}}\cdots\text{S}^{\text{I-}}$ or ferric–thiolate complexes given that the spin populations on

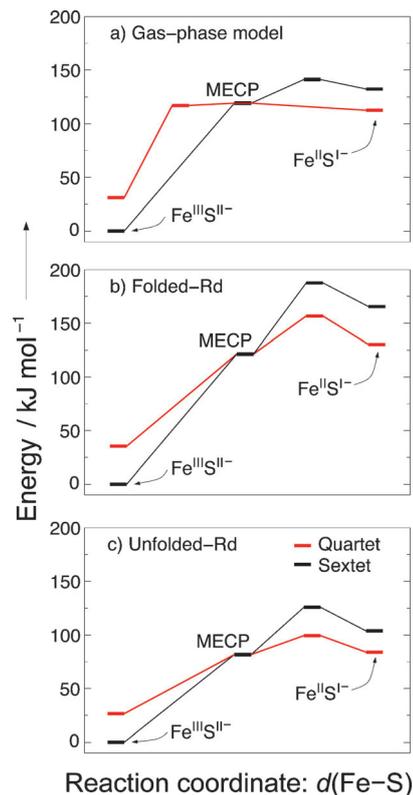


Figure 3. Relative energy profiles for iron–sulfur bond dissociation in the quartet (red) and sextet (black) spin states. Fe–S dissociation proceeds from left to right in all panels with the reaction coordinate taken as the Fe–S distance, $d(\text{Fe–S})$. a) the DFT profile for the gas-phase model Equation (1); b) and c) DFT/MM profiles for folded and unfolded rubredoxin (Rd), respectively.

the sulfur atoms are small. In contrast, the products have approximately one unpaired electron on the dissociated sulfur atom, suggesting a formal $\text{Fe}^{\text{II}}\text{--S}^{\text{I-}}$ or a ferrous–thiolate radical pair. The spin populations of the intermediate points with increasingly higher Fe–S distance interpolate nicely between the $\text{Fe}^{\text{III}}\text{--S}^{\text{II-}}$ and $\text{Fe}^{\text{II}}\cdots\text{S}^{\text{I-}}$ formal descriptions, indicating that Fe–S dissociation in $\text{Fe}(\text{SCH}_3)_4^-$ proceeds through a homolytic bond fission mechanism.

To examine Fe–S bond rupture in the protein, QC/MM calculations of the protein in explicit solvent were performed using structures obtained from the MM/MD simulations. In these models, the Fe center and Cys side-chains were in the QC region and were treated with a DFT potential. Only Fe–S(Cys5) rupture was studied in detail because there is experimental evidence that this and the Fe–S(Cys38) bond are the most labile of the four Cys bonds.^[3,4] Our simple mechanical model under stress also produced Fe–S(Cys5) dissociation in most of the pulling simulations that we conducted, and tests of Fe–S rupture calculated with the hybrid potential gave higher barriers for the other three Cys residues.

Figure 3b and c show that both folded and unfolded rubredoxin are similar to the gas-phase model with ground-state sextet reactants and quartet products. In contrast, the TSs on both surfaces are late, with the MECPs between the

surfaces coming before the TSs and 20–30 kJ mol⁻¹ lower in energy. This situation indicates that the Fe–S dissociation kinetics in rubredoxin will be mainly determined by the ⁴TS–⁶R energy difference. Spin crossover is still part of the reaction mechanism but should play a less-important kinetic role for dissociation in the protein compared to the intrinsic gas-phase reaction. The mechanism of bond fission is homolytic in both rubredoxin models. Spin populations (not shown) are similar to the gas-phase values, with both quartet and sextet products having approximately one unpaired electron on sulfur, indicating a Fe^{II}–thiolate radical dissociation product.

The protein environment imposes significant geometrical constraints on the cluster. The quartet–sextet gap increases by approximately 5 kJ mol⁻¹ in folded rubredoxin and decreases by about the same amount in unfolded rubredoxin compared to the gas-phase gap. The folded protein restricts Fe coordination to a rather symmetrical tetrahedral coordination (Tables S3,S4),^[8] which destabilizes the quartet state. On the other hand, unfolded rubredoxin is more flexible and allows the coordination around iron to relax. Consequently, reaction barriers and energies for both spin-states are significantly higher (ca. 50 kJ mol⁻¹, Figure 3b and c) in folded rubredoxin as a result of constraints in Fe–S bond elongation imposed by the folded protein matrix.

These constraining effects are partially mitigated by hydrogen bonds between the S_γ atoms of the cluster and the protein backbone, especially S_γ(Cys5)⋯NH(Ile7) which is maintained throughout the reaction and stabilizes Fe–S rupture (Table S4).^[8] This type of hydrogen-bonding stabilization has been confirmed by experiments which show that mutation of Ile7 to Pro in rubredoxin results in greater Fe–S bond mechanical stability.^[4] The simulated reaction energy and barrier for the flexible unfolded rubredoxin are 20–30 kJ mol⁻¹ lower than the gas-phase reaction. Both hydrogen-bonding and protein constraints are responsible for shifting ⁴TS to a late TS in unfolded rubredoxin.

The barrier to Fe–S dissociation in unfolded rubredoxin is 99 kJ mol⁻¹ corresponding to the relative energy of ⁴TS (Figure 3c). This compares reasonably well with the barrier of 78 kJ mol⁻¹ derived from the experimental dissociation rate (0.15 s⁻¹).^[3] The low barrier obtained for the reverse reaction, in which the Fe–S bond reforms (right to left on Figure 3c), is also corroborated experimentally as the rubredoxin recovers mechanical stability upon relaxation of the stressed chain.^[3] This qualitative agreement provides support for the simulation methodology that we employ.

A high degree of covalency has been assigned to the Fe–S bond from the interpretation of X-ray absorption spectroscopy on iron–sulfur complexes,^[12] although it does not translate into a high stability for Fe–S bonds in rubredoxin

under mechanical stress.^[3] In the analysis proposed by Solomon et al.,^[12] the stability of a Fe–S bond depends on its covalency as well as on the electrostatic interaction energy between the dissociated fragments. For the homolytic reaction detected in our work, the radical products [Fe^{II}Cys₃⋯S⁻¹Cys] will have neutral formal charges and consequently a much smaller electrostatic interaction than charge separated heterolytic products. Thus, the mechanical stability of Fe–S bonds in stretched rubredoxin is principally determined by their intrinsic bond covalency and interactions with the protein matrix, with the ionic character of the bonds playing only a minor role.

Herein we have shown that iron–sulfur tetrahedral complexes have sextet ground states at their equilibrium geometries but pass through a spin crossing and change to quartet states upon Fe–S dissociation. This two-state reactivity has been extensively characterized in metal complexes^[7] and is relevant for the stability of Fe–S bonds in rubredoxin. For both folded and mechanically unfolded rubredoxin, and for the intrinsic reaction, Fe–S cleavage follows a homolytic mechanism. Such a mechanism, and the insights provided by our calculations, should be pertinent for studies of dissociation of metal–ligand bonds in other bioinorganic complexes and metalloenzymes.

Received: April 23, 2013

Published online: June 18, 2013

Keywords: density functional theory · iron–sulfur clusters · molecular dynamics · single-molecule studies · spin crossover

- [1] H. Beinert, R. H. Holm, E. Munck, *Science* **1997**, *277*, 653–659.
- [2] H. Beinert, *J. Biol. Inorg. Chem.* **2000**, *5*, 2–15.
- [3] P. Zheng, H. Li, *J. Am. Chem. Soc.* **2011**, *133*, 6791–6798.
- [4] P. Zheng, S. J. Takayama, A. G. Mauk, H. Li, *J. Am. Chem. Soc.* **2012**, *134*, 4124–4131.
- [5] S. Garcia-Manyes, J. Liang, R. Szoszkiewicz, T.-L. Kuo, J. M. Fernandez, *Nat. Chem.* **2009**, *1*, 236–242.
- [6] M. Grandbois, M. Beyer, M. Rief, H. Clausen-Schaumann, H. E. Gaub, *Science* **1999**, *283*, 1727–1730.
- [7] D. Schröder, S. Shaik, H. Schwarz, *Acc. Chem. Res.* **2000**, *33*, 139–145.
- [8] Full details can be found in the Supporting Information.
- [9] R. Bau, D. C. Rees, D. M. Kurtz, R. A. Scott, H. Huang, M. W. W. Adams, M. K. Eidsness, *J. Biol. Inorg. Chem.* **1998**, *3*, 484–493.
- [10] M. W. Day, B. T. Hsu, L. Joshua-Tor, J. B. Park, Z. H. Zhou, M. W. Adams, D. C. Rees, *Protein Sci.* **1992**, *1*, 1494–1507.
- [11] S. Madsen, F. Jensen, *Theor. Chem. Acc.* **2009**, *123*, 477–485.
- [12] E. I. Solomon, S. I. Gorelsky, A. Dey, *J. Comput. Chem.* **2006**, *27*, 1415–1428.

ANEXO F – Force induced chemical reactions on the metal center in a single metalloprotein molecule

ARTICLE

Received 16 Jan 2015 | Accepted 19 May 2015 | Published 25 Jun 2015

DOI: 10.1038/ncomms8569

OPEN

Force-induced chemical reactions on the metal centre in a single metalloprotein molecule

Peng Zheng^{1,2}, Guilherme M. Arantes³, Martin J. Field⁴ & Hongbin Li¹

Metalloproteins play indispensable roles in biology owing to the versatile chemical reactivity of metal centres. However, studying their reactivity in many metalloproteins is challenging, as protein three-dimensional structure encloses labile metal centres, thus limiting their access to reactants and impeding direct measurements. Here we demonstrate the use of single-molecule atomic force microscopy to induce partial unfolding to expose metal centres in metalloproteins to aqueous solution, thus allowing for studying their chemical reactivity in aqueous solution for the first time. As a proof-of-principle, we demonstrate two chemical reactions for the FeS₄ centre in rubredoxin: electrophilic protonation and nucleophilic ligand substitution. Our results show that protonation and ligand substitution result in mechanical destabilization of the FeS₄ centre. Quantum chemical calculations corroborated experimental results and revealed detailed reaction mechanisms. We anticipate that this novel approach will provide insights into chemical reactivity of metal centres in metalloproteins under biologically more relevant conditions.

¹Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1. ²School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210063, P. R. China. ³Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, Avenue Lineu Prestes 748, São Paulo SP 05508-900, Brazil. ⁴Institut de Biologie Structurale (IBS) Jean-Pierre Ebel, CEA/CNRS/Université Joseph Fourier, 71 Avenue des Martyrs, CS 10090, Grenoble 9 38044, France. Correspondence and requests for materials should be addressed to G.M.A. (email: garantes@iq.usp.br) or to H.L. (email: Hongbin@chem.ubc.ca).

Metalloproteins are ubiquitous in nature and play indispensable roles in a wide range of cellular processes^{1,2}. Owing to their versatile chemical reactivity, incorporated metal ions add functionality to proteins and help to catalyse some of the most difficult reactions in nature³. Because of this, the investigation of metal centre reactivity is of great importance for understanding metalloprotein function and mechanism of action. In many metalloproteins, labile metal centres are often enclosed and protected by the three-dimensional (3D) structure of the polypeptide chain, limiting access to reactants and making it difficult to study their chemical reactivity in aqueous environment.

To overcome the challenges arising from protein structures, synthetic analogue strategies have been developed. These strategies involve the synthesis of metal complexes that mimic the natural metal centres in proteins, but lack a 3D protein scaffold. For example, FeS clusters from iron–sulfur proteins were some of the earliest types of metal centres that were synthesized, and the study of such metal centres has provided much valuable structural and functional information^{4,5}. The first kinetic study of a ligand substitution reaction on a synthetic cluster $\text{Fe}_4\text{S}_4(\text{SR})_4^{4-}$, which serves as the analogue for the active site of ferredoxin, was accomplished in the early 1970s (refs 6,7). In addition, protonation chemistry on the cluster has also been widely demonstrated⁸. These studies revealed the versatile reactivity of the FeS cluster in iron–sulfur proteins, suggesting that such reactivity may be important for the catalytic functions of some metalloproteins *in vivo*⁹, such as the nitrogenases and hydrogenases.

By contrast, direct demonstration of the reactivity of metalloprotein FeS centres in aqueous solution has been challenging, in part because of the fact that metal centres are often buried and inaccessible to the aqueous environment. The addition of chemical denaturants, which disrupts the protein structure and exposes metal centres to the solution, is often used to investigate the reactivity of FeS centres in metalloproteins in aqueous solution^{10,11}. However, there is growing evidence that the protein structure close to the metal centre can significantly influence and regulate the reactivity of the metal centre, making the use of chemical denaturants potentially problematic^{3,12}. As a result, it is necessary to develop alternative and complementary methods to study the properties of such metal centres by accessing the metal centre while maintaining protein native structure as much as possible.

Single-molecule force spectroscopy has evolved into a powerful technique for investigating force-induced conformational changes in macromolecules and chemical reactions through the application of a stretching force with piconewton precision to individual molecules^{13–20}. In particular, single-molecule force spectroscopy has enabled the investigation of protein unfolding/folding reactions under a stretching force in great detail. The experimental conditions offered by force spectroscopy closely mimic the physiological environment of a wide range of proteins that are subjected to a stretching force in a variety of biological processes, including muscle contraction²¹, protein translocation²² and protein proteasomal degradation²³. Such environments are also relevant for some metalloproteins. For example, metalloprotein superoxide dismutase 1 is known to follow a proteasomal degradation pathway and its forced unfolding is an important step in its degradation and the disassembly of its metal centres²⁴.

Our previous studies have shown that single-molecule atomic force microscopy (AFM) can be used to investigate the unfolding/folding mechanism of metalloproteins and the disassembly of metal centres^{25–29}. In a force spectroscopy experiment, a protein is stretched from two specific residues and can undergo force-induced unfolding. Weak interactions/bonds can be ruptured

along the unfolding pathway^{13,16,17,20}. Rubredoxin is the first metalloprotein that has been studied in detail using single-molecule AFM. Our results showed that the FeS_4 centre is ruptured as the small iron–sulfur protein rubredoxin is unfolded^{26,27}. Subsequent molecular simulations and quantum chemical (QC) calculations provided a detailed description of rubredoxin unfolding and the rupture mechanism of ferric–thiolate bonds³⁰. Our results showed that specific residues (1–5 and 41–53) outside of the metal centre can be first unfolded and extended before rupture of the FeS_4 centre. As a result, the FeS_4 centre can be exposed to the aqueous environment, while the protein structure between residues 5 and 41 remains largely intact^{25,28}. We reasoned that the FeS_4 centre exposed in this way could readily access chemical reactants in solution and participate directly in chemical reactions within the bulk solution, while the rest of the protein structure remains folded.

Here we use the well-characterized rubredoxin as a model system to demonstrate the utility of AFM to probe the chemical reactivity of metal centre in metalloproteins in aqueous environments. We investigate two different chemical reactions at the iron–sulfur centre in rubredoxin: electrophilic protonation chemistry and nucleophilic ligand substitution. We demonstrate that protonation of the FeS_4 centre occurs when the metal centre is exposed as a result of the force-induced partial unfolding of rubredoxin. We observed that the rupture force and bond lifetime of ferric–thiolate bonds in the FeS_4 centre is significantly decreased in acidic solutions as a result of protonation of the FeS_4 centre. In addition, we found that substitution with thiocyanate (SCN^-), a weaker nucleophile compared with thiolate for ferric ion, is facilitated by the application of a mechanical stretching force. The mechanical rupture rate of ferric–thiolate bonds in the presence of SCN^- is linearly dependent on SCN^- concentration. Further details into the molecular mechanisms and distinctive features of these reactions in the FeS_4 centre in rubredoxin were elucidated using QC calculations. By combining single-molecule force spectroscopy with QC calculations, our work represents a new approach towards studying the chemical reactivity of metal centres in metalloproteins, one that we anticipate will provide insights into the chemical reactivity of metal centres in aqueous solution under biologically relevant conditions.

Results

Protonation chemistry on the FeS_4 centre of rubredoxin.

Rubredoxin is the simplest iron–sulfur protein found in nature, with only 53 residues enclosing the FeS_4 centre active site^{31,32}, and its mechanical unfolding and the disassembly of the FeS_4 centre have been characterized in detail using single-molecule AFM techniques and QC calculations^{25–30}. In the FeS_4 centre, ferric–thiolate bonds between the sulfur atoms of four cysteine residues coordinate the ferric ion. As shown in Fig. 1a, the metal ion is buried inside the protein, making it difficult to access by exogenous reactants in solution. Although many iron–sulfur clusters participate in protonation chemistry⁸, such reactions cannot be detected by simply putting the protein in mildly acidic solutions. For example, the FeS_4 centre in rubredoxin is stable even at pH 2 (10 mM H^+) at room temperature³³. However, force-induced partial unfolding of the protein should lead to the exposure of the FeS_4 to aqueous environment (Fig. 1b), making it feasible to study the chemical reactivity of the FeS_4 centre in aqueous environments.

To directly investigate the chemical reactivity of the FeS_4 centre in rubredoxin, we constructed a polyprotein chimera (RD-GB1)_{*n*} to use in single-molecule AFM experiments, where RD represents rubredoxin. It incorporates the well-studied GB1 domain as a

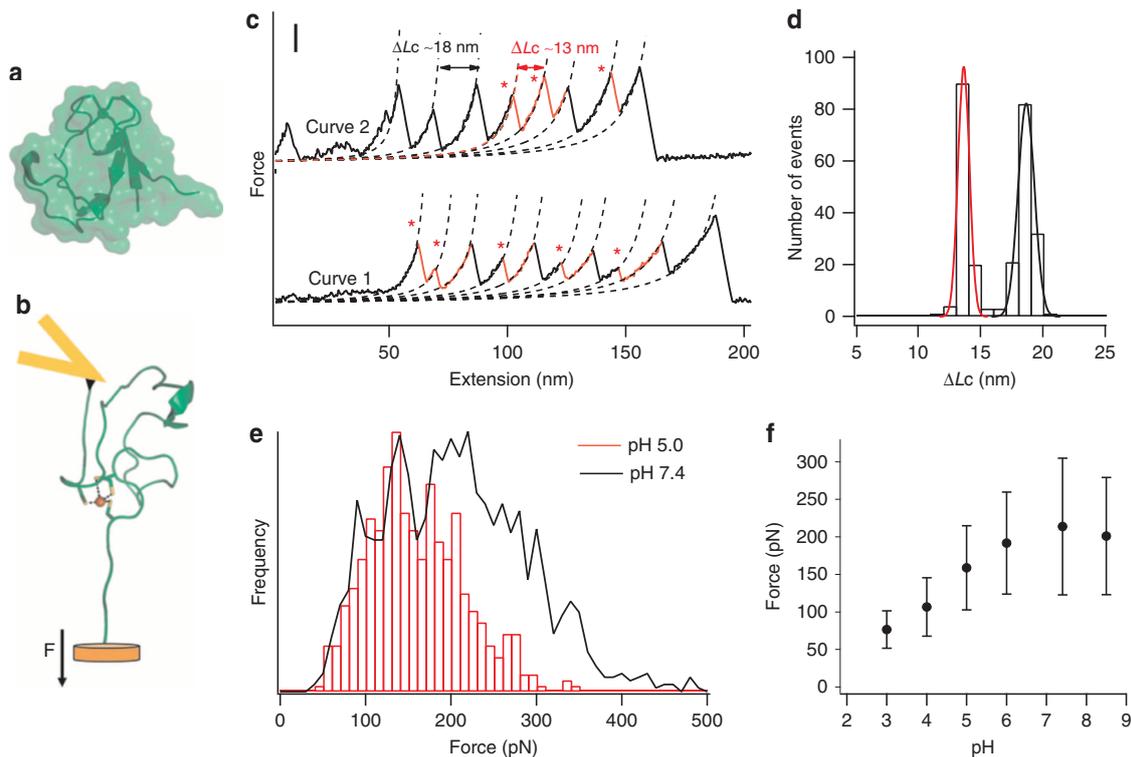


Figure 1 | Partial unfolding of rubredoxin by force exposes the FeS₄ centre and allows for protonation of the metal site. (a) The structure of rubredoxin, depicted in both cartoon and surface models, shows that the FeS₄ centre is buried within the protein structure. (b) A schematic showing how rubredoxin can partially unfold during stretching in the AFM experiments, leading to the exposure of the FeS₄ centre to the aqueous environment. (c) Typical force-extension curves of stretching polyprotein (RD-GB1)_n at a pH of 5.0 (curve 1) and a pH of 7.4 (curve 2). Both curves show force peaks corresponding to the rupture of the FeS₄ centre, which are characterized by ΔLc of 13 nm and is indicated by *, and unfolding of the fingerprint domain GB1 (ΔLc of 18 nm). Scale bar for the y axis, 100 pN. (d) A histogram of ΔLc of (RD-GB1)_n at pH 5 shows two distributions: one is centred at ~13 nm, and the other one is at ~18 nm. Gaussian fits (solid lines) to the experimental data measure a ΔLc of 13.2 ± 0.5 nm for rubredoxin and 18.1 ± 0.6 nm for GB1. (e) Rupture force histograms of rubredoxin at pH 7.4 (in black) and pH 5 (in red). These histograms clearly show that the rupture force of the FeS₄ centre decreases at lower pH. (f) The rupture force of ferric-thiolate bonds decreases as the pH decreases from 7.4 to 3. The error corresponds to the s.d.

fingerprint to help identify single-molecule stretching events³⁴. In addition, the GB1 domain is a non-metalloprotein whose mechanical stability is largely unaffected by pH and ligand substitution, and is thus a perfect control and internal force caliper for investigating the effect of pH on the rupture of FeS₄ centre in rubredoxin in our AFM experiments³⁵. Stretching the polyprotein (RD-GB1)_n results in force-extension curves with a characteristic sawtooth pattern, in which each individual sawtooth corresponds to the mechanical unfolding of an individual domain. Unfolding events with a contour length increment ΔLc of 18 nm correspond to GB1 unfolding events, while those with a ΔLc of 13 nm correspond to the complete unfolding of rubredoxin and the rupture of its FeS₄ centre (Fig. 1c). The unfolding force for rubredoxin corresponds mainly to the force required to rupture its FeS₄ centre, as apo-rubredoxin unfolds at forces that are below the detection limit (~ 20 pN) of our AFM²⁶.

Many residues in rubredoxin, including the sulfur atoms involved in the ferric-thiolate bonds, can easily become protonated. Since apo-rubredoxin unfolds at forces below 20 pN under acidic conditions (data not shown), which is similar to that under neutral conditions, changes of unfolding force of holo-rubredoxin under acidic conditions will reflect the effect of protonation of residues that directly affect the mechanical strength of the FeS₄ centre. In fact, protonation of sulfur atoms will result in Fe-S bond weakening and an acceleration of thiolate bond rupture. Thus, monitoring the force at which the FeS₄

centre ruptures under acidic conditions should provide valuable information about the protonation state of the FeS₄ centre.

Stretching the polyprotein (RD-GB1)_n in the presence of 0.01 mM H⁺ (pH 5) resulted in typical force-extension curves with a sawtooth pattern appearance (Fig. 1c, curve 1). Fitting the experimental data to the worm-like chain model of polymer elasticity³⁶ revealed that the unfolding force peaks show two distinct contour length increments. This is similar to the results of previous experiments performed at neutral pH (Fig. 1c, curve 2)²⁶. The ΔLc histogram clearly shows these two distributions (Fig. 1d). In addition, the average unfolding force of the GB1 domain is similar under the two conditions, specifically 194 ± 48 pN (average \pm s.d., $n = 284$, pH 7.4) and 190 ± 49 pN ($n = 441$, pH 5). This clearly shows that protonation does not affect the unfolding of GB1. In contrast to this, the rupture of the FeS₄ centre was significantly affected by a decrease in pH. The rupture force of the ferric-thiolate bonds at pH 5 was considerably lower (160 ± 60 pN ($n = 500$)) than that measured at neutral pH (211 ± 86 pN, $n = 1421$; Fig. 1c). A clear shift towards lower force is also observed within the rupture force histogram (Fig. 1e). Figure 1f shows the rupture force of ferric-thiolate bonds as a function of pH, from which it is evident that the mechanical bond strength of the ferric-thiolate bond is significantly weakened at an acidic pH when compared with a neutral pH. This change in the ferric-thiolate bond rupture force is likely caused by protonation of the metal site at acidic pH.

To obtain further insights about how protonation affects the ferric–thiolate bond rupture, we performed stretching experiments at different pulling speeds (Supplementary Fig. 1) and measured the pulling-speed dependence of these rupture forces. Using a well-established Monte Carlo procedure³⁷, we estimated the ferric–thiolate bond dissociation rate constant under zero force, and the distance between the bound state and the transition state, Δx_u . An acidic pH significantly increased the Δx_u when compared with the bond rupture of ferric–thiolate bonds at neutral pH (from 0.11 nm at neutral pH to 0.17 nm at acidic pH). This result corroborates our assertion that protonation likely occurs on the thiolate ligand and changes the rupture mechanism of the FeS_4 centre at acidic pH. In addition, the spontaneous dissociation rate, α_0 , increases with a decrease in pH (an increase in $[\text{H}^+]$), indicating how protonation affects the ferric–thiolate bond rupture process. This result indicates that the sulfur atom of the ferric–thiolate bond is indeed protonated during the rubredoxin unfolding process, demonstrating the chemical reactivity of sulfur atoms in the FeS_4 centre (Supplementary Fig. 2).

Ligand substitution on the rubredoxin FeS_4 centre. Force can be used not only to probe the consequence of protonation of the ferric–thiolate bonds, which is a common reaction for iron–sulfur clusters, but may also be exploited to mechanically activate the metal–ligand bond, thus making a suite of otherwise difficult reactions possible. To examine this possibility, we carried out a ligand substitution reaction on the FeS_4 centre in rubredoxin using SCN^- , which is a weaker nucleophilic ligand than thiolate and so does not normally substitute for thiolate at the FeS_4 centre. We carried out force–extension measurements of rubredoxin in the presence of SCN^- (Fig 2). As shown in Figs 2a,b, the ferric–thiolate bond rupture force decreases in the presence of SCN^- , suggesting that SCN^- weakens the ferric–thiolate bond.

Moreover, this weakening effect is dependent on the concentration of SCN^- , where increasing concentrations of SCN^- result in a decrease in bond strength (Fig. 2d). These results clearly indicate that, although SCN^- is a weaker nucleophilic ligand, it still affects the mechanical rupture of the FeS_4 centre. This seeming paradox can be explained if we suppose that stretching forces weaken the FeS_4 centre to such an extent that SCN^- can start to compete with thiolate for the ferric ion. It is worth noting that rubredoxin is much less sensitive to the effect of SCN^- than it is to protonation, as a minimum concentration of 5 mM thiocyanate is necessary for an observable effect (compared with a minimum proton concentration of 0.01 mM).

We then carried out force clamp experiments^{38,39} to obtain quantitative information about how SCN^- affects rupture kinetics. Figure 3a shows a typical extension–time trace of $(\text{RD-GB1})_n$ in the presence of 50 mM KSCN under a constant stretching force of 90 pN. Normalized ensemble averages of length versus time from 51 molecules (Fig. 3b) can be fit using a single exponential relationship; this was used to measure the ferric–thiolate bond rupture rate in the presence of 50 mM SCN^- at a force of 90 pN. By carrying out force clamp experiments at different forces, we found that the logarithm of the rupture rate of ferric–thiolate bonds depends linearly on the pulling force, thereby following force–rupture behaviour predicted by the classic Bell–Evans model (Fig. 3c). Fitting the experimental data to the Bell–Evans model⁴⁰ allowed us to estimate the spontaneous rupture rate of ferric–thiolate bonds at zero force in the presence of a given SCN^- concentration. Substitution rates were obtained from a linear fit of the semi-logarithm plot, specifically 0.15, 0.21, 0.33 and 0.46 s^{-1} at KSCN concentrations of 5, 50, 500 and 800 mM, respectively (Fig. 3c). The distance between bound state and mechanical transition state is ~ 0.14 nm under each concentration. We found that the rupture rate of ferric–thiolate bonds is linearly dependent on $[\text{SCN}^-]$, suggesting that the

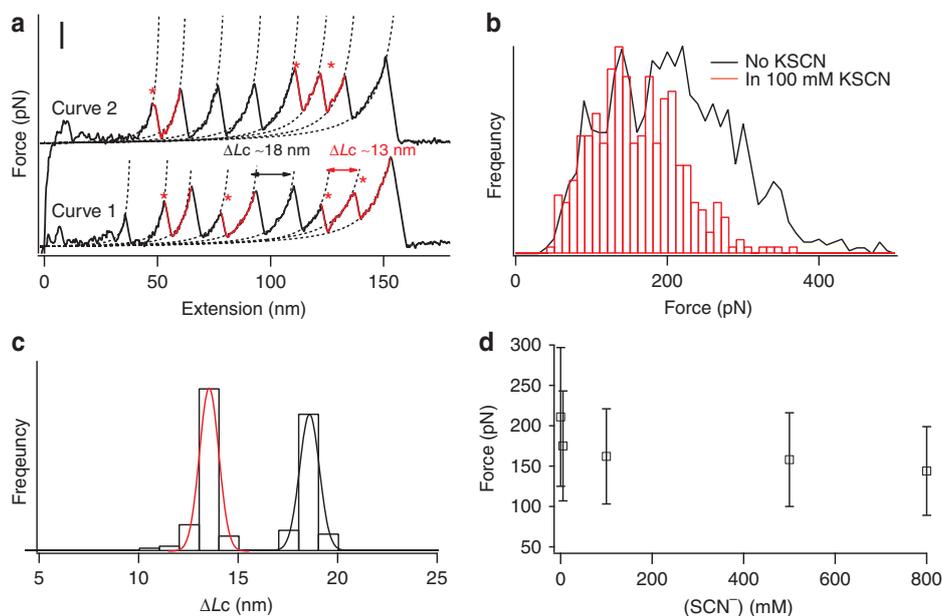


Figure 2 | Mechanical unfolding of rubredoxin at various KSCN concentrations. (a) Typical force–extension curves of stretching polyprotein $(\text{RD-GB1})_n$ in 100 mM KSCN (curve 1) and in the absence of KSCN (curve 2). Both curves show force peaks corresponding to the rupture of the FeS_4 centre, which are characterized by ΔLc of 13 nm and indicated by *, and unfolding of the fingerprotein domain GB1 (ΔLc of 18 nm). Scale bar for the y axis is 100 pN. (b) Rupture force histograms of rubredoxin in the presence of 100 mM KSCN (in black) and in Tris (in red). These histograms clearly show that the rupture force of the FeS_4 centre decreases in the presence of 100 mM KSCN. (c) The histogram of ΔLc of $(\text{RD-GB1})_n$ in the presence of 100 mM KSCN shows two distributions: one is centred at ~ 13 nm and the other one is at ~ 18 nm. Gaussian fits (solid lines) to the experimental data measure a ΔLc of 12.9 ± 0.5 nm for rubredoxin and 18.0 ± 0.5 nm for GB1. (d) The rupture force of ferric–thiolate bonds decreases as the KSCN concentration increases.

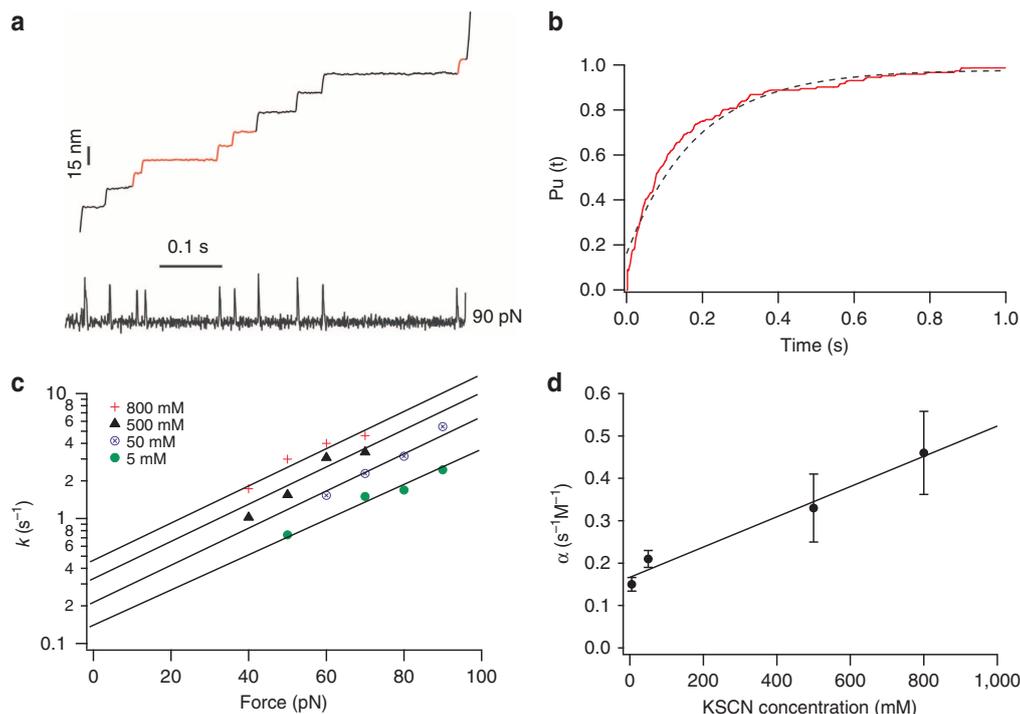


Figure 3 | Force clamp experiments characterize the rupture rate of ferric-thiolate bond in the presence of KSCN. (a) Representative length-time curves of (rubredoxin-GB1)_n under a constant force of 90 pN in the presence of 50 mM SCN⁻. Steps coloured in black correspond to the unfolding events of GB1, while those in red correspond to the unfolding event of rubredoxin. (b) Average time course of unfolding of rubredoxin obtained by summation and normalization of the rubredoxin portion of 51 curves such as the one in a. The unfolding time course of rubredoxin in the presence of 50 mM SCN⁻ can be described by a single exponential (blue dotted line) with a rate constant of 5.39 s⁻¹. (c) Semi-logarithmic plot of FeS₄ centre rupture rate as a function of force under different concentrations of KSCN. The solid lines are fits based on the Bell-Evans model, giving the spontaneous rupture rate of FeS₄ in the presence of KSCN. (d) Rupture rate constant of FeS₄ as a function of KSCN concentration. The solid line is a linear fit to the data. This linear dependence on [KSCN] demonstrates that the SCN⁻ ligand substitution reaction is first order with respect to SCN⁻.

rupture of ferric-thiolate bonds in the presence of SCN⁻ is first order with respect to [SCN⁻]. This linear relationship can be described as $r = k \times [\text{SCN}^-] + 0.15$, where r is the ferric-thiolate bond dissociation rate in the presence of SCN⁻, k (the rate constant) is equal to $0.36 \text{ M}^{-1} \text{ s}^{-1}$ (Fig. 3d) and the intercept of 0.15 indicates the spontaneous bond dissociation rate of Fe-S bond in the absence of SCN⁻. Ligand substitution reactions are common reactions in synthetic analogues of Fe-S clusters from iron-sulfur proteins, and are often second-order reactions, being first order with respect to both the metal cluster and the competing agent⁹. For example, several ligand substitution reactions on the alkylthiolate tetramer dianion Fe₄S₄(SR)₄²⁻, which is an analogue of the active site of ferredoxin, exhibit rates that are linearly dependent on the concentration of the competing agent R'SH (refs 7,41–43). Our results that the SCN⁻ substitution reaction on the ferric-thiolate bond in rubredoxin is first order with respect to SCN⁻ when the pH is kept at 7.4 are in good agreement with studies of similar reactions using inorganic analogues of the FeS cluster from iron-sulfur proteins.

QC calculations outline the FeS bond cleavage mechanism. Our experimental results on protonation and ligand substitution reactions on the FeS₄ centre in rubredoxin demonstrate that chemical reactions can be directly monitored in iron-sulfur proteins in aqueous solution in the absence of chemical denaturants, where the protein is only partially unfolded by a stretching force and the protein structure surrounding the metal centre is still largely present.

Although our AFM experiments allow the observation of ferric-thiolate bond rupture in rubredoxin, they cannot give indications as to the mechanism by which rupture occurs within the protein. This rupture could occur in a heterolytic manner, leading to thiolate anion; it can also be homolytic, resulting in a thiol radical^{44,45}. In our previous QC studies, we found that the rupture of ferric-thiolate bonds without ligand substitution or protonation follows a homolytic mechanism³⁰. To gain insights into the mechanism by which FeS₄ rupture occurs within the protein in the presence of exogenous ligands, we used QC calculations based on density functional theory (DFT) to model Fe-S bond cleavage reactions. The Fe(SCH₃)₄⁻ compound was chosen to model the ferric-thiolate centre in rubredoxin exposed under mechanical tension. In order to unfold rubredoxin completely, two ferric-thiolate bonds have to be broken. These reactions were investigated under three conditions: pure water, acidic conditions and in the presence of SCN⁻, as shown in Fig. 4a–d. Reactions in Fig. 4a–d are referred to as Reaction A–D, respectively. Calculations were performed in an implicit aqueous solvent to account for the aqueous environment in which AFM experiments were conducted. Further details on the calculation procedures are given in the Methods section.

In pure water and in the presence of SCN⁻, we observed that substitution reactions have significantly lowered activation barriers compared with bare FeS bond dissociation without ligand exchange. Under acidic conditions, both ligand substitution and bare dissociation are competitive. In contrast to our previous study in which the rupture of ferric-thiolate bonds follows a homolytic mechanism³⁰, we found that the rupture of

ferric–thiolate bonds proceeds via a heterolytic mechanism in the reactions with ligand substitution and leaving-group protonation; no electronic spin crossing is observed, as all products stay in their initial reactant high-spin (sextet) configuration.

Sequential FeS bond cleavage from $\text{Fe}(\text{SCH}_3)_4^-$ without ligand exchange will proceed by a simple dissociation mechanism, in which products have a three- and a two-coordinated ferric ion centre (Dn only mechanism, Fig. 4d). On the other hand, ligand substitution reactions may proceed through three possible mechanisms: (A) initial thiol(ate) dissociation, which results in a stable intermediate with three-coordinated iron, followed by the new ligand addition (Dn + An mechanism); (B) ligand addition forming a stable penta-coordinated intermediate, followed by thiol(ate) dissociation (An + Dn mechanism); (C) simultaneous ligand addition and thiol(ate) elimination without a stable intermediate (AnDn mechanism).

For reactions occurring in pure water (Fig. 4a), the two ligand substitution reactions proceed via an AnDn mechanism with a concerted proton transfer from the attacking water to the thiolate-leaving group. Two-dimensional potential energy scans were conducted to determine the concerted proton and ferric ion transfer reaction barrier. Figure 5a shows that the second substitution reaction is rate-limiting, with a barrier of 89 kJ mol^{-1} .

For the reaction in the presence of SCN^- , we investigated the sequential substitution of two SCN^- towards iron and the substitution of one SCN^- and one water molecule (Fig. 4b). Figure 5b shows that the reaction profile with the lowest barrier corresponds to an initial reaction with water (barrier of 65 kJ mol^{-1}) followed by the rate-limiting substitution with SCN^- , with a barrier of 70 kJ mol^{-1} , corroborating the first-

order character of the SCN^- reaction. The mechanism for SCN^- substitution towards the ferric ion centre in either $\text{Fe}(\text{SCH}_3)_4^-$ or $\text{Fe}(\text{SCH}_3)_3\text{OH}^-$ is An + Dn, with a shallow and transient penta-coordinated intermediate. The reaction barriers are much higher for two sequential SCN^- reactions, and also for the dissociative mechanisms (Dn + An and Dn only, data not shown).

For the reactions under acidic conditions, protonation of the thiolate ligand leads to considerably lower barriers for FeS bond rupture. Figure 5c shows that the second substitution reaction is rate-limiting, with a barrier of 34 kJ mol^{-1} , corresponding to the water addition step in a Dn + An mechanism. A competitive channel with a barrier of 40 kJ mol^{-1} was found for water substitution without acid catalysis in the second FeS bond cleavage, corresponding to thiol dissociation via an An + Dn mechanism. A purely dissociative mechanism without ligand substitution at iron (Dn only, Fig. 4d) also becomes competitive, with a barrier of 39 kJ mol^{-1} (Fig. 5d), corresponding to the second FeS bond dissociation. It should be noted that for the purely dissociative mechanism (Dn only) shown in Figs 4d and 5d, no transition state was identified, and barriers correspond to the dissociated product energies. Thus, based only on the calculated energy profiles, we are unable to discriminate between reaction C and D or between the associative and dissociative mechanisms under acidic conditions.

In order to resolve this uncertainty and directly probe the effect of mechanical force on the ferric–thiolate bond dissociation, we employed the constrained geometries simulate external force (COGEF) method and determined the rupture or maximum force along the dissociation profiles (Supplementary Fig. 3). Table 1 shows results for the reactions in pure water and acidic

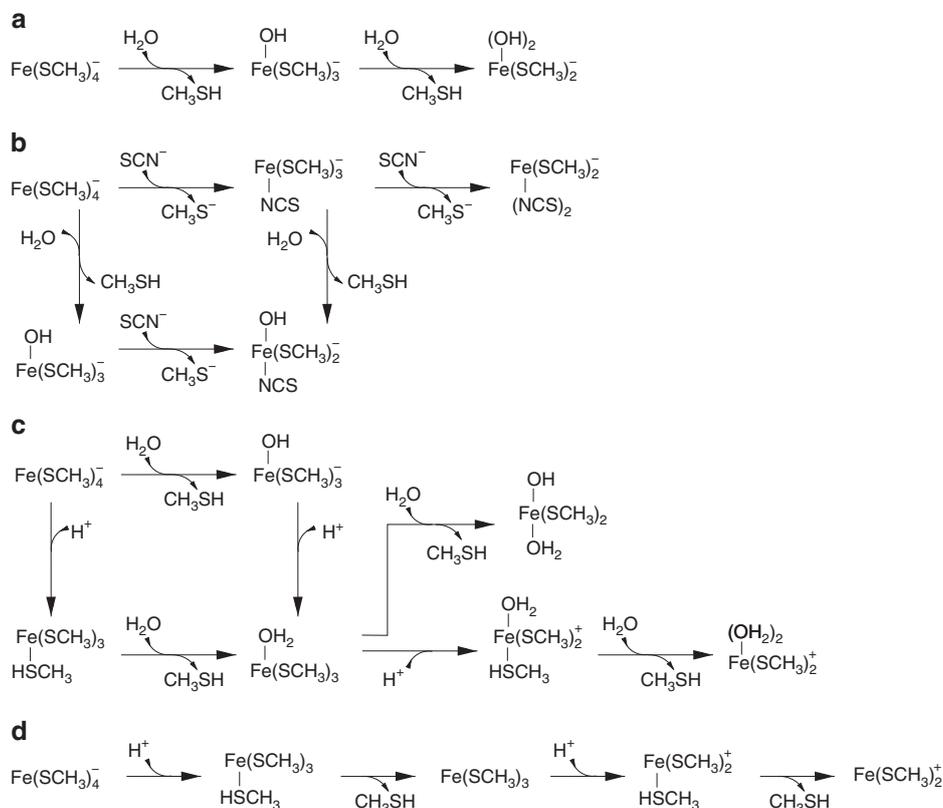


Figure 4 | Scheme of the $\text{Fe}(\text{SCH}_3)_4^-$ model reactions studied. (a) Pure water substitution (reaction A); (b) substitution in the presence of SCN^- (reaction B); (c) water substitution under acid catalysis (reaction C); (d) bond dissociation under acid catalysis (reaction D).

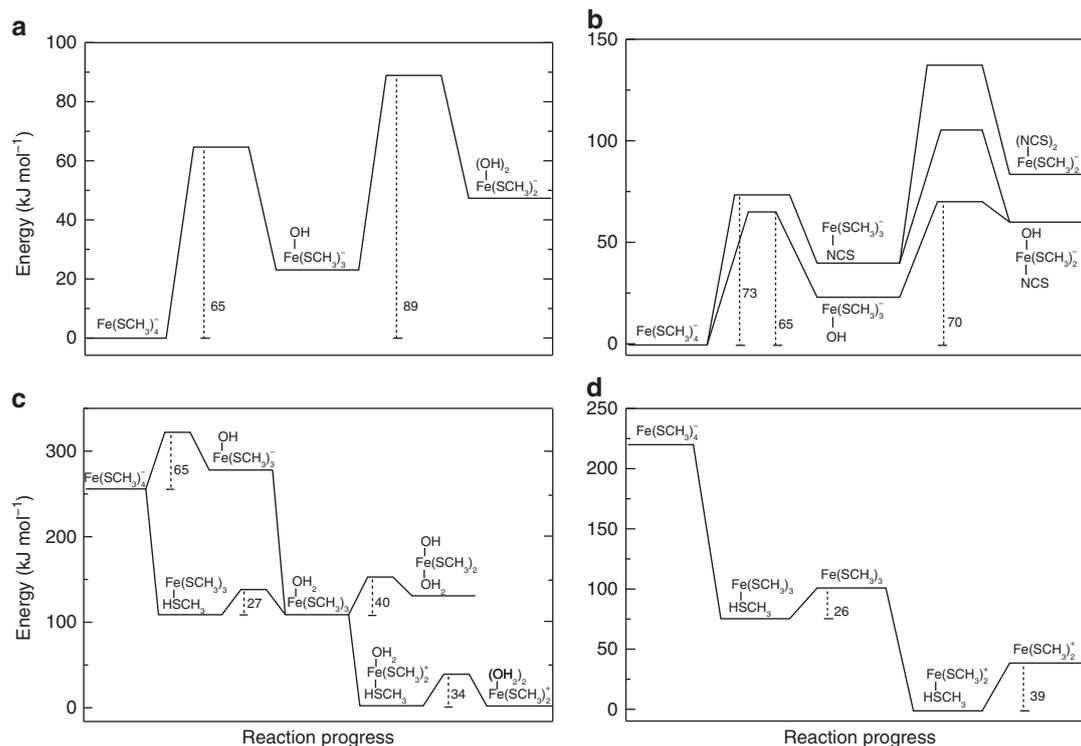


Figure 5 | Relative energy profiles indicate possible mechanisms for the ferric-thiolate bond cleavage. Calculations were performed at the DFT level of theory in implicit aqueous solvent for the $\text{Fe}(\text{SCH}_3)_4^-$ model reactions in their sextet electronic spin states. **(a)** Pure water substitution; **(b)** substitution in the presence of SCN^- ; **(c)** water substitution under acid catalysis; **(d)** bond dissociation under acid catalysis.

Table 1 | Rupture force and difference in bond distance between reactant and transition state.

Reaction	Mechanism	F_{max} (pN)	$\Delta d(\text{FeS})$ (Å)
A, step 1	AnDn	548	0.2
A, step 2	AnDn	651	0.3
C, step 1 protonated	AnDn	298	0.2
C, step 2 neutral	An + Dn	963	0.2
C, step 2 protonated	Dn + An	350	0.6
D, step 1	Dn	579	1.1

Rupture force (F_{max} in pN) and difference in bond distance ($\Delta d(\text{FeS})$ in Å) are calculated with the COGEF method. Steps 1 and 2 correspond to the first and second FeS bond cleavage, respectively (Fig. 4).

conditions. The calculated forces are consistently about twice as high as the average of the measured AFM rupture forces. This shift between calculated COGEF and measured forces has been observed for other mechanochemical activated reactions³⁶.

Most importantly, we observe a clear decrease in the calculated maximum force when going from the pure water ($F_{\text{max}} = 651$ pN) to the acid-catalysed reaction ($F_{\text{max}} = 350$ pN), in agreement with the AFM data. Comparison of the calculated rupture forces helps to clarify the mechanochemical mechanism of ferric-thiolate bond dissociation under acidic conditions. The mechanism with the lowest rupture forces corresponds to reaction C with a concerted AnDn mechanism in the first FeS bond cleavage ($F_{\text{max}} = 298$ pN and barrier of 27 kJ mol^{-1}) and a protonated intermediate that dissociates via a Dn + An mechanism in the second FeS bond cleavage ($F_{\text{max}} = 350$ pN and barrier of 34 kJ mol^{-1}). Reaction D and the two deprotonated steps in

reaction C have significantly higher calculated rupture forces (Table 1).

On going from the pure water to the acid-catalysed reaction, the change in reaction coordinate between bound and transition states estimated from the AFM experiments (Δx_{u}) increases by 0.6 \AA . This is comparable to the 0.3 \AA increase obtained from the calculated differences of ferric-thiolate bond distance between reactant and transition state structures ($\Delta d(\text{FeS})$ in Table 1) for the reactions with the lowest rupture force. Differences between the experimental and calculated values for Δx_{u} and rupture force (F_{max}) may be attributed to approximations in the QC level of theory (DFT) and to a lack of description of the full protein structure in the calculated model.

Discussion

The chemical reactivity of metal centres in metalloproteins plays critical roles in the functioning of metalloproteins in biology. In physiological environments, metal centres of many metalloproteins are often enclosed by the 3D structure of proteins, making it difficult to investigate the reactivity of metal centres in aqueous environments. Synthetic inorganic analogues provide excellent model compounds for probing the structure and function of metal centres in metalloproteins. However, most inorganic analogues, including iron-sulfur clusters, are labile towards water and/or oxygen, in the absence of the protective 3D structure of proteins. As a result, many experiments using inorganic analogues have been conducted in non-aqueous organic solvent, with only a few recent studies performed with partially protic solvents⁴³. Because of this, it remains challenging to investigate the chemical reactivity of metal centres in metalloproteins in aqueous environments in the presence of proteins' 3D structure.

In this paper, using rubredoxin as a model system, we have demonstrated the utility of single-molecule force spectroscopy techniques in probing the chemical reactivity of metal centres in metalloproteins in aqueous solution. Two different chemical reactions, protonation and ligand exchange reaction, were observed to occur at the FeS₄ centre of rubredoxin using AFM. Our results showed that the bond strength of ferric–thiolate bonds in rubredoxin is weakened and their lifetime shortened when competing interactions with exogenous ligands are present. These results are corroborated by QC calculations, and demonstrate that partial unravelling of rubredoxin facilitates the attack of the FeS₄ centre by exogenous ligands in aqueous solution, even though most of the proteins' 3D structure remains largely intact. Moreover, QC calculations reveal that ferric–thiolate bond rupture follows a heterolytic reaction mechanism in the presence of ligands, which is different from the homolytic rupture of ferric–thiolate bonds without ligand substitution or protonation. It is worth noting that heterolytic reaction mechanisms are also common for other reactions in solution. For example, it has been observed in QC calculations that the mechanical rupture of the backbone of polyethylene glycol in solution follows a heterolytic reaction mechanism⁴⁴.

Compared with chemical denaturation methods (which result in complete unfolding of the metalloprotein), the force spectroscopy method demonstrated here allows one to partially unfold a metalloprotein to expose the metal centre while keeping the rest of protein structure largely intact, thus enabling the examination of the chemical reactivity of metal centres in the presence of protein 3D structure in aqueous solution. It is worth noting that how the metal centre becomes exposed is different in mechanical and chemical denaturations. Thus, the chemical reactivity of metal centres may exhibit differences in mechanical and chemical denaturation studies. Nonetheless, the force spectroscopy method demonstrated here provides a new approach to study the chemical reactivity of metal centres in metalloproteins, which will complement other methods (including chemical denaturation and synthetic analogues) and provide additional insights into this important problem. Although this study is carried out on the model protein rubredoxin, the extension of the novel method demonstrated here to other proteins is currently under investigation in our laboratories. In principle, the method proposed should be applicable to the study of a wide variety of other metalloproteins in aqueous solution, including such complex examples as the nitrogenases and hydrogenases. Moreover, the forced unfolding that we employ may be able to mimic the physiological conditions to which some metalloproteins are subject *in vivo*, and thus help elucidate their chemical reactivity and functions in a setting that is close to that *in vivo*.

Methods

Protein engineering. The gene encoding protein chimera RD-GB1 was constructed as previously reported^{26,46}. The protein was expressed using pQE80L vector in *Escherichia coli* strain DH5 α and purified with Co²⁺-affinity chromatography using TALON His-Tag purification resins (Clontech), followed by ion exchange chromatography using a Mono Q 5/50 anion exchange column (GE Healthcare). The resultant ferric-form rubredoxin chimera was reacted with BM(PEO)₃ (1, 8-bis (maleimido)triethylene glycol, Molecular Biosciences) through a thiol–maleimide coupling reaction, forming the polyprotein (RD-GB1)_n (ref. 46). The specific rubredoxin we studied is from the hyperthermophile *Pyrococcus furiosus*. The ferrous rubredoxin (with an oxidation state of +2) is air-sensitive because of oxidation; however, the ferric form we studied here is stable in air. In our experiment, the protein is kept in Tris buffer under ambient conditions.

Single-molecule AFM experiments. Single-molecule AFM experiments were carried out on a custom-built AFM as reported previously³⁴. Before each experiment, each Si₃N₄ cantilever (Bruker Corp) was calibrated in solution using the equipartition theorem (with typical value of ~ 40 pN nm⁻¹). In a typical AFM stretching experiment, the polyprotein sample (2 μ l, 2 mg ml⁻¹) was added on a clean glass coverslip covered by Tris buffer (~ 50 μ l, pH 7.4, 100 mM Tris and

NaCl). The protein was allowed to adsorb on the coverslip for ~ 5 min before the experiment. Concentrations of 5, 50, 500 and 800 mM of KSCN were obtained by adding an appropriate quantity of 1 M KSCN stock solution to the Tris buffer.

QC calculations. DFT with the unrestricted B3LYP functional^{47,48} and the 6-31 + G(2df,p) Pople basis set⁴⁹ were used for quantum chemical calculations. The minimum energy pathway for FeS bond dissociation obtained with this level of theory compares satisfactorily to the energy profile obtained with high-level *ab initio* multiconfigurational calculations (data not shown). Electronic structure calculations were performed with GAUSSIAN 2009 revision A.01 (ref. 50). FeS bond dissociation and FeO (water) or FeN (thiocyanate) bond formation were mimicked by optimizing the complex with a fixed breaking or forming bond distance, respectively. The relevant intermediate structures were then optimized without constraints. Transition states were not fully optimized; however, energy barriers correspond to the highest-energy structures found along the fixed bond scans. Energy differences between these structures and actual transition states were smaller than ~ 2 kJ mol⁻¹. Geometry optimizations of concerted iron and proton transfer reactions were conducted by additional constraints in the hydrogen-acceptor bond distance, using the pDynamo library⁵¹ interfaced with the ORCA programme version 3.01 (ref. 52). Calculations were conducted in aqueous solvent using the Polarizable Continuum Model⁵³ (PCM, in GAUSSIAN), or the Conductor-like Screening Model⁵⁴ (COSMO, in ORCA). The rupture force or the maximum force along the FeS dissociation pathway was calculated with the constrained geometries simulate external force (COGEF) method^{45,55}. The distance between two hydrogen atoms from different methyl groups were constrained (COGEF extension) and the remaining degrees of freedom were optimized. Force versus extension curves were obtained by finite differences of the energy versus extension profiles after cubic spline fitting. All relative energies reported here contain electronic, nuclear repulsion and solvent contributions only.

The pH dependence of the electrostatic free energy between the folded and unfolded states of rubredoxin (Supplementary Fig. 4) was estimated using Poisson–Boltzmann calculations using the standard method proposed in ref. 56.

References

- Holm, R. H., Kennepohl, P. & Solomon, E. I. Structural and functional aspects of metal sites in biology. *Chem. Rev.* **96**, 2239–2314 (1996).
- Lippard, S. J. & Berg, J. M. *Principles of Bioinorganic Chemistry* (University Science Books, 1994).
- Lu, Y., Berry, S. M. & Pfister, T. D. Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. *Chem. Rev.* **101**, 3047–3080 (2001).
- Beinert, H., Holm, R. H. & Münck, E. Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* **277**, 653–659 (1997).
- Rao, P. V. & Holm, R. H. Synthetic analogues of the active sites of iron-sulfur proteins. *Chem. Rev.* **104**, 527–559 (2004).
- Dukes, G. R. & Holm, R. H. Synthetic analogs of the active sites of iron-sulfur proteins. X. Kinetics and mechanism of the ligand substitution reactions of arylthiols with the tetranuclear clusters [Fe₄S₄(SR)₄]²⁻. *J. Am. Chem. Soc.* **97**, 528–533 (1975).
- Que, L., Bobrik, M. A., Ibers, J. A. & Holm, R. H. Synthetic analogs of the active sites of iron-sulfur proteins. VII. Ligand substitution reactions of the tetranuclear clusters [Fe₄S₄(SR)₂]- and the structure of bis(tetramethylammonium) [tetra- μ -sulfide-tetrakis(benzenethiolato)tetrairon]. *J. Am. Chem. Soc.* **96**, 4168–4178 (1974).
- Henderson, R. A. Proton transfer to synthetic Fe–S-based clusters. *Coord. Chem. Rev.* **249**, 1841–1856 (2005).
- Henderson, R. A. Mechanistic studies on synthetic Fe–S-based clusters and their relevance to the action of nitrogenases. *Chem. Rev.* **105**, 2365–2437 (2005).
- Foster, M. W., Bian, S. M., Surerus, K. K. & Cowan, J. A. Elucidation of a 4Fe–4S cluster degradation pathway: rapid kinetic studies of the degradation of Chromatium vinosum HiPIP. *J. Biol. Inorg. Chem.* **6**, 266–274 (2001).
- Higgins, C. L. & Wittung-Stafshede, P. Formation of linear three-iron clusters in Aquifex aeolicus two-iron ferredoxins: effect of protein-unfolding speed. *Arch. Biochem. Biophys.* **427**, 154–163 (2004).
- Lu, Y., Yeung, N., Sieracki, N. & Marshall, N. M. Design of functional metalloproteins. *Nature* **460**, 855–862 (2009).
- Brockwell, D. J. Force denaturation of proteins - an unfolding story. *Curr. Nanosci.* **3**, 3–15 (2007).
- Puchner, E. M. & Gaub, H. E. Single-molecule mechanoenzymatics. *Annu. Rev. Biophys.* **41**, 497–518 (2012).
- Marszalek, P. E. & Dufrene, Y. F. Stretching single polysaccharides and proteins using atomic force microscopy. *Chem. Soc. Rev.* **41**, 3523–3534 (2012).
- Li, H. B. 'Mechanical engineering' of elastomeric proteins: toward designing new protein building blocks for biomaterials. *Adv. Funct. Mater.* **18**, 2643–2657 (2008).
- Carrión-Vázquez, M. *et al.* Mechanical design of proteins-studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **74**, 63–91 (2000).

18. Janshoff, A., Neitzert, M., Oberdorfer, Y. & Fuchs, H. Force spectroscopy of molecular systems - single molecule spectroscopy of polymers and biomolecules. *Angew. Chem. Int. Ed.* **39**, 3213–3237 (2000).
19. Beyer, M. K. & Clausen-Schaumann, H. Mechanochemistry: the mechanical activation of covalent bonds. *Chem. Rev.* **105**, 2921–2948 (2005).
20. Zoldak, G. & Rief, M. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.* **23**, 48–57 (2013).
21. Li, H. *et al.* Reverse engineering of the giant muscle protein titin. *Nature* **418**, 998–1002 (2002).
22. Prakash, S. & Matouschek, A. Protein unfolding in the cell. *Trends Biochem. Sci.* **29**, 593–600 (2004).
23. Sauer, R. T. & Baker, T. A. AAA + proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80**, 587–612 (2011).
24. Kabuta, T., Suzuki, Y. & Wada, K. Degradation of amyotrophic lateral sclerosis-linked mutant Cu,Zn-superoxide dismutase proteins by macroautophagy and the proteasome. *J. Biol. Chem.* **281**, 30524–30533 (2006).
25. Zheng, P., Chou, C. C., Guo, Y., Wang, Y. & Li, H. Single molecule force spectroscopy reveals the molecular mechanical anisotropy of the FeS4 metal center in rubredoxin. *J. Am. Chem. Soc.* **135**, 17783–17792 (2013).
26. Zheng, P. & Li, H. Highly covalent ferric – thiolate bonds exhibit surprisingly low mechanical stability. *J. Am. Chem. Soc.* **133**, 6791–6798 (2011).
27. Zheng, P., Takayama, S.-i. J., Mauk, A. G. & Li, H. Hydrogen bond strength modulates the mechanical strength of ferric-thiolate bonds in rubredoxin. *J. Am. Chem. Soc.* **134**, 4124–4131 (2012).
28. Zheng, P., Takayama, S. J., Mauk, A. G. & Li, H. Single molecule force spectroscopy reveals that iron is released from the active site of rubredoxin by a stochastic mechanism. *J. Am. Chem. Soc.* **135**, 7992–8000 (2013).
29. Zheng, P., Wang, Y. & Li, H. Reversible unfolding-refolding of rubredoxin: a single-molecule force spectroscopy study. *Angew. Chem. Int. Ed. Engl.* **53**, 14060–14063 (2014).
30. Arantes, G. M., Bhattacharjee, A. & Field, M. J. Homolytic cleavage of Fe-S bonds in rubredoxin under mechanical stress. *Angew. Chem. Int. Ed. Engl.* **52**, 8144–8146 (2013).
31. Blake, P. R. *et al.* Determinants of protein hyperthermostability: purification and amino acid sequence of rubredoxin from the hyperthermophilic archaeobacterium *Pyrococcus furiosus* and secondary structure of the zinc adduct by NMR. *Biochemistry* **30**, 10885–10895 (1991).
32. Lovenber, W. & Sobel, B. E. Rubredoxin - a new electron transfer protein from clostridium Pasteurianum. *Proc. Natl Acad. Sci. USA* **54**, 193–199 (1965).
33. Cavagnero, S., Zhou, Z. H., Adams, M. W. & Chan, S. I. Response of rubredoxin from *Pyrococcus furiosus* to environmental changes: implications for the origin of hyperthermostability. *Biochemistry* **34**, 9865–9873 (1995).
34. Cao, Y., Lam, C., Wang, M. & Li, H. Nonmechanical protein can have significant mechanical stability. *Angew. Chem. Int. Ed. Engl.* **45**, 642–645 (2006).
35. Zheng, P. *et al.* Single molecule force spectroscopy reveals that electrostatic interactions affect the mechanical stability of proteins. *Biophys. J.* **100**, 1534–1541 (2011).
36. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
37. Rief, M., Fernandez, J. M. & Gaub, H. E. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett.* **81**, 4764–4767 (1998).
38. Fernandez, J. M. & Li, H. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* **303**, 1674–1678 (2004).
39. Garcia-Manyes, S., Liang, J., Szoszkiewicz, R., Kuo, T. L. & Fernandez, J. M. Force-activated reactivity switch in a bimolecular chemical reaction. *Nat. Chem.* **1**, 236–242 (2009).
40. Evans, E. Probing the relation between force--lifetime--and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 105–128 (2001).
41. Bobrik, M. A., Que, Jr. L. & Holm, R. H. Synthetic analogs of the active sites of iron-sulfur proteins. IV. Ligand substitution reactions of the tetranuclear clusters (Fe4S4(SR)4)²⁻. *J. Am. Chem. Soc.* **96**, 285–287 (1974).
42. Job, R. C. & Bruce, T. C. Iron-sulfur clusters II: Kinetics of ligand exchange studied on a water-soluble Fe4S4(SR)4²⁻ cluster. *Proc. Natl Acad. Sci. USA* **72**, 2478–2482 (1975).
43. Lo, W., Scott, T. A., Zhang, P., Ling, C. C. & Holm, R. H. Stabilities of cubane type [Fe(4)S(4)(SR)(4)](2-) clusters in partially aqueous media. *J. Inorg. Biochem.* **105**, 497–508 (2011).
44. Aktah, D. & Frank, I. Breaking bonds by mechanical stress: when do electrons decide for the other side? *J. Am. Chem. Soc.* **124**, 3402–3406 (2002).
45. Ribas-Arino, J. & Marx, D. Covalent mechanochemistry: theoretical concepts and computational tools with applications to molecular nanomechanics. *Chem. Rev.* **112**, 5412–5487 (2012).
46. Zheng, P., Cao, Y. & Li, H. Facile method of constructing polyproteins for single-molecule force spectroscopy studies. *Langmuir* **27**, 5713–5718 (2011).
47. Becke, A. D. Density-Functional Thermochemistry.3. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
48. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37**, 785–789 (1988).
49. Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-consistent molecular-orbital methods.9. Extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **54**, 724 (1971).
50. Frisch, M. J. *et al.* *Gaussian 09, Revision A.01* (Gaussian Inc., 2009).
51. Field, M. J. The pDynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials. *J. Chem. Theory Comput.* **4**, 1151–1161 (2008).
52. Neese, F. The ORCA program system. *Wires Comput. Mol. Sci.* **2**, 73–78 (2012).
53. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3093 (2005).
54. Klamt, A. Conductor-Like Screening model for real solvents - a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **99**, 2224–2235 (1995).
55. Beyer, M. K. The mechanical strength of a covalent bond calculated by density functional theory. *J. Chem. Phys.* **112**, 7307–7312 (2000).
56. Yang, A. S. & Honig, B. On the pH dependence of protein stability. *J. Mol. Biol.* **231**, 459–474 (1993).

Acknowledgements

This work is supported by the Natural Sciences and Engineering Research Council of Canada, the Canada Foundation for Innovation, the Canada Research Chairs programme. G.M.A. acknowledges funding from FAPESP (projects 12/02501-4 and 14/21900-2).

Author contributions

P.Z. and H.L. conceived the project, P.Z., G.M.A., M.J.F. and H.L. designed the research, P.Z., G.M.A. and M.J.F. performed the experiments/calculations and analysed the data, P.Z., G.M.A., M.J.F. and H.L. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Zheng, P. *et al.* Force-induced chemical reactions on the metal centre in a single metalloprotein molecule. *Nat. Commun.* **6**:7569 doi: 10.1038/ncomms8569 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

ANEXO G – Flexibility and inhibitor binding in Cdc25 phosphatases

Flexibility and inhibitor binding in Cdc25 phosphatases

Guilherme Menegon Arantes*

Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900, São Paulo, SP, Brasil

ABSTRACT

Cdc25 phosphatases involved in cell cycle checkpoints are now active targets for the development of anti-cancer therapies. Rational drug design would certainly benefit from detailed structural information for Cdc25s. However, only apo- or sulfate-bound crystal structures of the Cdc25 catalytic domain have been described so far. Together with previously available crystallographic data, results from molecular dynamics simulations, bioinformatic analysis, and computer-generated conformational ensembles shown here indicate that the last 30–40 residues in the C-terminus of Cdc25B are partially unfolded or disordered in solution. The effect of C-terminal flexibility upon binding of two potent small molecule inhibitors to Cdc25B is then analyzed by using three structural models with variable levels of flexibility, including an equilibrium distributed ensemble of Cdc25B backbone conformations. The three Cdc25B structural models are used in combination with flexible docking, clustering, and calculation of binding free energies by the linear interaction energy approximation to construct and validate Cdc25B-inhibitor complexes. Two binding sites are identified on top and beside the Cdc25B active site. The diversity of interaction modes found increases with receptor flexibility. Backbone flexibility allows the formation of transient cavities or compact hydrophobic units on the surface of the stable, folded protein core that are unexposed or unavailable for ligand binding in rigid and densely packed crystal structures. The present results may help to speculate on the mechanisms of small molecule complexation to partially unfolded or locally disordered proteins.

Proteins 2010; 78:3017–3032.
© 2010 Wiley-Liss, Inc.

Key words: molecular dynamics; conformational ensemble; docking; intrinsic disorder; protein tyrosine phosphatases; transient cavities; computer simulation; receptor flexibility.

INTRODUCTION

Cell division cycle Cdc25 phosphatases have turned into active targets for the development of antineoplastic agents as their overexpression has been detected in several human cancers.^{1–7} Although different classes of Cdc25s inhibitors have been identified so far, the number of potent inhibitors that are active *in vivo* is still limited and restricted to quinoid-based compounds.^{1,2} Thus, rational drug design toward Cdc25 would certainly benefit from a detailed knowledge of this enzyme catalytic mechanism and tertiary structure.

Cdc25s are dual-specificity protein tyrosine phosphatases (PTP) that dephosphorylate pTyr and pThr residues on cyclin-dependent kinases (Cdk) responsible for cell cycle checkpoints.^{8,9} Known as the P-loop, the active site is characterized by the sequence Cys-(Xxx)₅-Arg conserved among all known PTPs (Fig. 1). The conserved Cys is the nucleophile that attacks the substrate phosphate group, which is in turn stabilized by the conserved Arg side chain. Poor selectivity has been observed *in vitro* for the most potent PTP inhibitors because these compounds bind to the conserved P-loop.^{4,5} One important difference between Cdc25 and the other PTPs that could be explored in the design of selective inhibitors is the catalytic general acid. In Cdc25, this residue was identified by the Glu following the nucleophilic Cys in the P-loop sequence.^{9,10} But, in all other PTPs, the general acid is an Asp located in a flexible loop distant from the active site.^{8,11}

The three Cdc25 isoforms found in humans share 70% identity in the primary sequence of their catalytic C-terminal domain, but present larger variance in their N-terminal regulatory domain where several splice variants are possible.¹ All the structural information available for Cdc25s was obtained from X-ray crystallographic data for Cdc25A¹² and Cdc25B¹³ catalytic domains in either apo forms (PDB code 1c25)¹² or with sulfate bound to the P-loop (codes 1qb0, 1cwr, 2a2k, 2ifd).^{13,14} NMR structures and order parameters¹⁵ for Cdc25 in solution, structures for inhibitor-Cdc25 complexes, or for the divergent N-terminal domain have not been described to our knowledge. The most notable difference between the crystal structures of isoforms Cdc25A and B is observed in the backbone of the last 20 aminoacids in the C-terminus. This region contains an α -helix for Cdc25B but it is unstructured for Cdc25A (Fig. 1). Additional 28 (Cdc25A) and 16

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: FAPESP; Grant numbers: 07/52772-6, 07/59345-6

*Correspondence to: Guilherme Menegon Arantes, Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900, São Paulo, SP, Brasil. E-mail: garantes@iq.usp.br

Received 4 April 2010; Revised 7 June 2010; Accepted 23 June 2010

Published online 23 July 2010 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/prot.22826

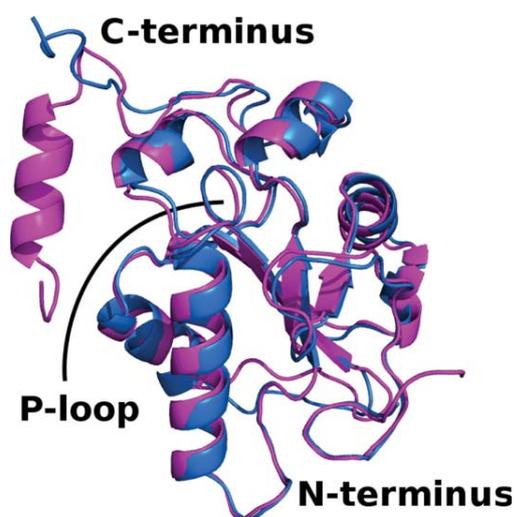


Figure 1

Superposition of the Cdc25A (in blue, PDB code 1c25) and Cdc25B (in magenta, PDB code 1qb0) catalytic domain crystal structures.

(Cdc25B) residues in the protein C-terminal were present in the constructs used for crystallization, but had no observable electron densities suggesting further intrinsic flexibility or disorder in the C-terminus.^{12,13} Interestingly, it has been shown that this C-terminal tail mediates substrate recognition by up to 100-fold in the phosphatase activity of Cdc25B and Cdc25C toward the bisphosphorylated Cdk2-pTpY/CycA natural substrate.¹⁶

Structural disorder plays a role in the molecular recognition between proteins involved in cell signaling.¹⁷ In particular, the interaction between Cdk and several of its partner proteins including p21,¹⁸ but not yet Cdc25, has been shown to proceed through disorder-to-order transitions.¹⁹ Thus, it is possible that the recognition between Cdc25 and Cdk/Cyc complexes is driven by a similar mechanism involving the last 30–40 residues in the Cdc25 C-terminal. Even if a disorder-to-order transition is not involved, C-terminal flexibility will be relevant because this tail occupies the protein-protein contact surface between the Cdc25 catalytic domain and the Cdk2 main body.¹⁴ The Cdc25 C-terminus is also in contact with its own main body and near the active site P-loop (Fig. 1), such that its flexibility will also influence binding of competitive inhibitors. What is then, the effect of this putative C-terminal disorder on the complexation of small molecules to Cdc25? It appears that this question has received little attention for partially unfolded proteins in general, even though it has been suggested that a large fraction of eukariotic proteins (~30%) contains intrinsically disordered regions.¹⁷

Modelling protein-inhibitor complexation, in particular, docking of small molecules to rigid receptor structures, is an established area with several parameterized

and tested methods.^{20,21} Including protein flexibility into such docking protocols is, however, less developed.^{22,23} Most of the approaches currently in use were designed to account for side chain or limited backbone flexibility, for instance, by introducing soft or average grid models of receptor ligand interaction.^{24,25} This is clearly inappropriate for large-scale fluctuations present in partially disordered proteins because average positions will not represent the relevant conformational states.^{23,26} Receptor flexibility may also be included by docking to several different structures from a configurational ensemble.^{27–29} The problem here is how to obtain representative structures for flexible or disordered protein domains. A set of geometries obtained from X-ray crystallography might be perturbed by crystal packing^{30,31} and disordered segments will not have visible electron densities. Ensembles obtained from NMR might be more appropriate models for solution configurations but the fitting procedure traditionally used for structural determination by reproducing distance restraints becomes less reliable for proteins with intrinsic disorder and does not result in an equilibrium distribution of configurations.^{15,30}

Receptor structures may be generated computationally, but large-scale fluctuations and protein disorder are more difficult to simulate accurately. Molecular dynamics (MD) simulations up to the nanosecond timescale under normal temperature probe local fluctuations around the starting geometry.³² High temperature MD may unveil cryptic binding sites or induced fit changes if the protein region that undergoes conformational exchange is known in advance.³³ Otherwise, several MD acceleration^{34,35} and generalized ensemble^{36,37} methods may be used to increase configurational sampling. Computer generation of protein conformational ensembles may be based on geometrical information of reference configurations,^{38–40} on backbone dihedral angle propensities extracted from existing libraries,^{41,42} or on combinations of both approaches.⁴³ Statistical libraries built from geometries of coiled protein fragments are reasonable models for the unfolded or disordered regions.^{41,44} For protein-inhibitor complexation a further challenge is the selection of an ensemble subset that is relevant for binding.²² Although a criteria based on structural variance is often used,^{22,32} a more formal filtering should rely on the Boltzmann distribution of the ensemble.⁴⁵

In this work, Cdc25B C-terminal flexibility is analyzed, and its effect on the complexation of quinone-based small molecules is investigated using computer simulations. Details about the computational methods used are given in the next section. Results for molecular dynamics simulations of Cdc25B strongly support that the C-terminus is flexible and partially unfolded in solution. Three Cdc25B structural models with variable levels of flexibility are compared: A rigid model corresponding to the crystal structure, a semiflexible model corresponding to a set of MD snapshots and a fully flexible model built with

statistical rigor for backbone configurations. Then, Cdc25B-inhibitor complexes using the three structural models are presented and validated by the calculation of binding free energies. Given the lack of experimental data to benchmark the structural models and the generated ligand–receptor complexes, the focus here is on discussing the binding sites and less detail is reported on specific ligand–receptor molecular interactions. Finally, the effect of C-terminal flexibility on the small molecule complexation is described.

METHODS

Molecular dynamics and system set-up

Coordinates of a sulfate-bound Cdc25B crystal structure deposited in the PDB code 1qb0 were used to start the molecular dynamics simulation. This protein chain contains 177 residues from Asp374 to Trp550 of the Cdc25B primary sequence. The initial simulation model was built by removing all crystallographic waters, ions, and β -mercaptoethanol. Missing hydrogen atoms were added, considering all Asp, Glu, Arg, Lys, and His side chains charged. This protonation state is slightly different from the one accepted as the catalytic competent state in which Cys473 is deprotonated and Glu474 protonated.⁹ Such difference should not influence the flexibility in the P-loop active site backbone, which, is rather rigid, but it might play a role for the detailed interactions in small molecule complexation in the active site. The protein chain was solvated by superimposing it to a replicated water box, leaving at least 12 Å between the protein and the box edge. Six chloride anions were added to neutralize the total system charge. Initial coordinates for one of the chloride anions were extracted from the crystal structure (PDB code 1qb0). Although it was suggested that this chloride anion may have a structural role,¹³ it diffuses out of the interaction site within 5 ns of simulation. Periodic boundary conditions were used with a cubic box length of ~ 78 Å. Long-range electrostatics were treated with the particle-mesh Ewald summation with a 1.2 Å Fourier grid spacing and fourth-order interpolation. A switch function with $r_{\text{off}} = 10.0$ Å was used to truncate long-range van der Waals interactions without dispersion corrections.

All model building, molecular dynamics simulations and trajectory analysis were carried out with the GRO-MACS 4.0.2 suite of programs.⁴⁶ The OPLS-AA⁴⁷ force field was used for the protein and ions, and the rigid TIP3P⁴⁸ potential was used for water. The LINCS algorithm was used to constrain all covalent bonds and a time step of 2 fs was used in MD integration.

The solvated protein heavy atoms were restrained to their initial coordinates. The system was first energy-minimized to relax unfavorable contacts, and kinetic energy was gradually injected in the system by running

short (40 ps) MD simulations at 50, 100, 200, and 300 K. Restraints in the protein heavy atoms were gradually removed by reducing the harmonic constant from the initial $1000 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ to 200, 20, and $2 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$ along another set of short (40 ps) MD simulations at 300 K. This careful heating and pre-equilibration phase was used to avoid artifactual flexibility in the apo Cdc25B simulation as well as artifactual decomplexation of inhibitors in the evaluation of interaction energies (see Linear Interaction Energy section below). An equilibration 2 ns MD run was obtained without any position restraints at 300 K and 1 atm. The Berendsen temperature and pressure couplings were used up to this point. Finally, a production 60 ns MD trajectory was obtained at the same temperature and pressure conditions but in a canonical ensemble using the Nose-Hoover extended temperature scheme (time constant $\tau_T = 0.1$ ps) and the Parrinello–Rahman pressure coupling (time constant $\tau_P = 0.5$ ps). No group separation was used.

Bioinformatic analysis of disordered sequences

Four different bioinformatic methods were used to predict the disordered regions of Cdc25 isoforms A and B. Primary sequences were submitted to web servers queries and default algorithm options were used. FoldIndex⁴⁹ builds a score based on residue charge and hydrophobicity for a sliding window of 51 residues. Disopred2⁵⁰ uses a linear support vector machine to analyze the sequence profiles. VL3H⁵¹ and VLXT⁵² both search using neural networks, and the latter is based on residue physicochemical properties. Default server and algorithmic values were used in all queries. Only VL3H gives a continuous N-terminal disordered segment. All other predictors give N-terminal segments containing ordered gaps of up to 30 amino acids, but with a total of more than 75% of disordered residues in the N-terminal domain. The score given for the discontinuous segments is the average of disordered scores (FoldIndex and VLXT) or the range of scores (Disopred2). In FoldIndex, negative scores mean a segment will more likely be disordered. Scores on the other methods give the probability of disordered structure in the segment.

Ensemble generation

Two conformational ensembles based on the Cdc25B catalytic domain crystal structure (code 1qb0) were used here. The MD ensemble was composed of 150 snapshots spaced by 400 ps each from the 60 ns MD trajectory described above. Another Cdc25B conformational ensemble built with a library-based Monte Carlo approach was kindly provided by D. Zuckerman and co-workers.⁴³ The 177 Cdc25B residues deposited on the PDB code 1qb0 are present in this model. The protein is divided into

fragments identified with residue types for which statistical libraries of backbone conformations were built in advance. Fragment libraries have Boltzmann distributions p_i^{frag} , which account for correlations internal to each fragment. The distribution of protein configurations built with a simple product of fragments

$$p_i^{\text{lib}}(\mathbf{r}_i, \dots, \mathbf{r}_M) = \prod_i^M p_i^{\text{frag}}(\mathbf{r}_i) \quad (1)$$

will be biased toward the isolated fragment distributions. Thus, Metropolis Monte Carlo is used to resample and obtain protein configurations in equilibrium distribution.⁵³ The protein internal energy is written as:

$$U^{\text{prot}}(\mathbf{r}_i, \dots, \mathbf{r}_M) = \sum_i^M U_i^{\text{frag}}(\mathbf{r}_i) + U^{\text{rest}}(\mathbf{r}_i, \dots, \mathbf{r}_M) \quad (2)$$

where U_i^{frag} is the energy internal to each fragment obtained from the library and U^{rest} accounts for the interactions between the M fragments. The interaction term is described by a structure-based (or Gō type) potential, which stabilizes the native Cdc25B fold¹³ but allows fluctuations around it in reasonable agreement with experimental data.⁴³ It should be noted that Boltzmann distributed protein backbone configurations are generated with this formal procedure. However, residue side chain conformations are built for each backbone geometry separately, using the backbone-dependent rotamer library SCWRL.⁵⁴ Thus, side chains will not follow the equilibrium distribution. Further details for the generation of the Zuckerman Cdc25B ensemble are given in the original paper.⁴³

Small molecule docking and clustering

The studied ligands were 6-chloro-7-(2-morpholin-4-ylethylamino)quinoline-5,8-dione (NSC 663284), **1** in Figure 2, and 2,5-dihydroxy-3-(7-(2-methyl-benzyl)-1H-indol-3-yl)[1,4]benzoquinone, **2** in Figure 2. For each receptor structure, 200 AutoDock 4.0 runs were performed for each ligand.⁵⁵ Interaction maps were generated in a cubic grid with 60 points and 0.375 Å spacing, centered in the Cys473 S_γ, the catalytic nucleophile. This medium sized grid allowed docking to sites up to 11 Å away from the Cdc25B active site. A distance-dependent dielectric constant was used for electrostatic interactions.⁵⁵ Four torsions for **1** and five torsions for **2** were active. The genetic algorithm used for conformational search was set with 150 individuals in the population for a maximum of 27,000 generations or 2.5×10^6 energy evaluations.

Two clustering steps were used to analyze the docking poses (more than 30,000 per ligand per ensemble). Within each receptor structure, poses were initially clustered for ligand heavy atom root mean-squared deviation (RMSD) with a 2.0 Å cutoff. The pose with lowest free energy score in each initial cluster was considered a valid one if it belonged to a cluster with at least 20 members

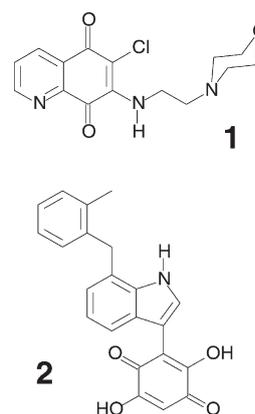


Figure 2

Structure of the quinolinequinone **1** and the indolyldihydroxyquinone **2**.

and if its AutoDock 4.0 free energy score was at most 2.0 kcal/mol higher than the lowest free energy score found within dockings for the same receptor structure. This is the definition for valid pose or valid complex used throughout this work. Then, the set of valid poses for all receptor structures in each ensemble was clustered for ligand heavy-atom RMSD tolerance of 4.0 Å after least-square fitting of C_αs for the stable residues in the receptor (residues 377–530). This procedure gave the best compromise between the size of clustering problem, the amount of clusters finally obtained, and the expected accuracy of the docking function.⁵⁵ The eight clusters with the largest population of valid poses for each ligand and ensemble were analyzed in detail below.

Several criteria in the above procedure were relaxed to test for cluster robustness. For instance, decreasing the number of members (20) or increasing the cutoff energy (2 kcal/mol) in 50% for the first clustering step, enlarged the set of valid poses, but yielded qualitatively the same eight most populated clusters, that is, the same binding sites and cluster centroids. Throughout this work, the centroid geometry represents the clusters obtained in the second clustering step. A cluster centroid is defined as the valid complex with the smallest average distance to the other valid complexes within the same cluster. Again, the same binding sites were obtained if the RMSD tolerance was changed to 3.5 or 4.5 Å in the second clustering step (the population of each cluster was smaller or larger, respectively). Several clustering algorithms (single-linkage, Daura *et al.*,⁵⁶ and self-organizing maps⁵⁷) were tested yielding similar results.

Linear interaction energy

The stability of the binding modes found in the docking and clustering procedure was estimated using the fol-

lowing linear interaction energy (LIE) approximation to the binding free energy⁵⁸:

$$\Delta G_{\text{bind}} \simeq \alpha(\langle V_{\text{comp}}^{\text{vdW}} \rangle - \langle V_{\text{sol}}^{\text{vdW}} \rangle) + \beta(\langle V_{\text{comp}}^{\text{elec}} \rangle - \langle V_{\text{sol}}^{\text{elec}} \rangle) \quad (3)$$

where $\langle \dots \rangle$ represents an ensemble average (obtained here from a MD trajectory) of nonbonded van der Waals (vdw) and electrostatic (elec) interaction energies (V) between the ligand and its surrounding environment. Direct calculation of absolute free energies is possible,⁵⁹ but extremely demanding for computational resources and, hence, unfeasible for such large set of complexes as found here. The LIE is computationally more efficient because it only requires two end states MD simulations, one for the free ligand in solvent (sol) and another for the ligand–protein complex (comp). The parameters adopted were $\alpha = 0.18$, $\beta = 0.37$ for **1**, and $\beta = 0.33$ for **2**.⁶⁰ MD simulations starting from the ligand–protein complex poses obtained with AutoDock and for the free ligand in water were carried out following the same model building, heating, and pre-equilibration procedures adopted for the apo Cdc25B (Molecular dynamics and system set-up section). The equilibration and data acquisition MD trajectories were run for 0.5 and 1.0 ns, respectively, and both simulations were obtained at 300 K and 1 atm in a canonical ensemble using the Nose-Hoover and the Parrinello–Rahman coupling schemes. Long-range electrostatics were treated with a switch function with $r_{\text{on}} = 8 \text{ \AA}$ and $r_{\text{off}} = 12 \text{ \AA}$. The force field for both ligands was derived from the OPLS-AA force field with appropriate partial charges derived from the charges assigned to similar OPLS groups.

RESULTS AND DISCUSSION

Cdc25B is partially unfolded in solution

To explore the intrinsic flexibility of apo Cdc25B in aqueous solvent, a long molecular dynamics simulation was conducted at normal conditions. The observed trajectory begins with a rapid relaxation of the protein structure from the initial crystallographic coordinates in response to the solvated environment. The protein core (residues 374–531) has a constant deviation from the initial coordinates measured by a C_{α} RMSD $\sim 2 \text{ \AA}$ up to the end of the simulation as shown in Figure 3(A). However, at about 40 ns, a steep increase in the C_{α} RMSD from 4 \AA to more than 8 \AA is observed for the last 19 residues in the Cdc25B C-terminus [Fig. 3(A)]. The protein radius of gyration, which is stable at $\sim 16.7 \text{ \AA}$ from 5 to 40 ns also increases to $\sim 17.0 \text{ \AA}$ from 40 ns to the end of the simulation.

Figure 3(B) shows fluctuations of the protein C_{α} s about their mean position obtained from the Cdc25B MD run and from the temperature factors of Cdc25A and Cdc25B crystallographic structures. The Cdc25B

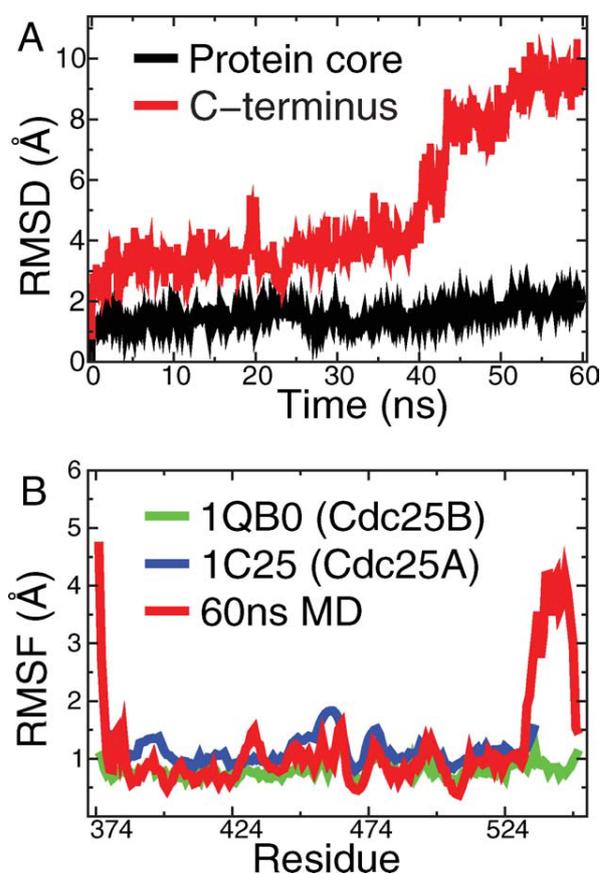


Figure 3

Molecular dynamics of apo Cdc25B in solution. Top panel A shows the C_{α} root mean-squared deviation from the crystallographic structure for the Cdc25B protein core (residues 374–531) and the 19 residues in the C-terminus (residues 532–550). Lower panel B gives the C_{α} root mean-squared fluctuation obtained from the 60ns MD and calculated from temperature factors for the Cdc25A (PDB code 1c25) and Cdc25B (PDB code 1qb0) structures. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

RMSF obtained from the X-ray structure is rather flat and featureless, which maybe a consequence of the refinement procedure on the diffraction data.³¹ There is some fair agreement between the RMSF obtained from the Cdc25A structure and from the Cdc25B simulation. The catalytic domains of the two isoforms share more than 70% identity in their primary sequences and have almost identical folds (Fig. 1), so they should also have equivalent backbone fluctuations. For instance, both RMSFs are larger than 1 \AA from residues 443–466, and drop steadily in residues 469–471. This is the active site region, which is prearranged to bind phosphate esters.^{9,11,13} The main differences between fluctuations calculated from the experimental and simulated data are observed for the protein terminals. The simulated RMSF is $>3 \text{ \AA}$ for the first three residues (374–376) in the N-

terminus and for residues 531–548 in the C-terminus [Fig. 3(B)]. The large and localized flexibility in the N-terminus should be expected as the protein is truncated in this position. In fact, the Cdc25B construct expressed and used for X-ray crystallography has in addition 18 residues (356–373) along the N-terminus and 16 residues (551–566) along the C-terminus, which are not observed in the electron density maps.¹³ The simulated protein contains only residues 374–550, as given in the PDB file, including an α -helix present along residues 533–546.¹³

The large RMSD and RMSF observed for the C-terminus during the MD trajectory correspond to a local unfolding and detaching of the terminal α -helix from the protein main-body. It may be possible that the last 16 residues (551–566) along the C-terminus, which were present in the Cdc25B construct used for crystallization, but absent from the MD simulation, are necessary to stabilize the terminal α -helix. However, no terminal helix is observed in the crystal structure for isoform Cdc25A (Fig. 1) even though the complete C-terminus, residues 336–523 in the Cdc25A numbering, was present in the crystallized construct. In fact, electron densities for the last 28 residues (496–523) in Cdc25A were not observed and this region was considered disordered in the crystal form.¹²

These observations suggest that the terminal Cdc25 regions, especially the last 30–40 residues in the C-terminus, are intrinsically disordered or at least partially unfolded in solution. The terminal α -helix present in the Cdc25B structure may be a crystal packing artifact or may be stabilized because favorable interactions between symmetrically related molecules (the terminal Trp550 side chains are in contact in the 1qb0 model^{13,43}).

To support the partial unfolding hypothesis, the primary sequences of Cdc25A and Cdc25B were submitted to bioinformatic analysis for prediction of disordered protein sequences.⁵² Table I shows the results obtained with four different methods. The whole N-terminal domain is predicted as disordered, with up to 320 unfolded residues for Cdc25A and 370 residues for Cdc25B. To our knowledge, the structure of the N-terminal region of Cdc25s has never been solved, which might be explained by the difficulty in obtaining crystals and a diffraction pattern from partially disordered materials. The ordered segment found by all bioinformatic methods roughly coincides with the catalytic domains of Cdc25A and B observed by X-ray crystallography^{12,13} and with the core region of Cdc25B stable during the MD simulations described above. The disordered segments found for the C-terminus are all continuous, but with variable size. They range from the last 75 residues (FoldIndex) down to the last 40 (VL3H) or 30–20 residues (Disopred2 and VLXT) for both Cdc25A and B. But, clearly, the bioinformatic search confirms the suggestion from MD simulations that the Cdc25 terminals are disordered or partially unfolded in solution.

Table I

Disordered Regions on the Cdc25A and B Complete Primary Sequences Predicted by Various Bioinformatic Methods^a

Method	Cdc25A		Cdc25B	
	Segment	Score	Segment	Score
Disopred2 ⁵⁰	1–323	0.2–1.0	1–368	0.2–1.0
	504–523	0.5	546–566	~0.5
FoldIndex ⁴⁹	61–325	-0.20 ± 0.09	160–375	-0.22 ± 0.08
	450–523	-0.19 ± 0.08	494–566	-0.14 ± 0.06
VL3H ⁵¹	1–323	0.8	1–368	0.8
	480–523	0.8	528–566	0.7
VLXT ⁵²	1–329	0.7	1–365	0.8
	493–523	0.8	541–562	0.9

^aResidue numbering covers the entire protein sequence: 1–523 for Cdc25A and 1–566 for Cdc25B.¹³

Modeling Cdc25B with conformational ensembles

The effects of protein flexibility upon binding and molecular recognition may be accounted for by modeling the receptor geometry with a conformational ensemble or a collection of structures.^{22,29} Statistically significant sampling of such conformations following a Boltzmann distribution⁴⁵ is not feasible with the computer power available nowadays if proteins the size of Cdc25 catalytic domain (~200 residues) are modelled by detailed all-atom force fields, such as the one used here for MD. Different methods exist to generate conformational ensembles.^{38–42,61,62} Here two computational models were adopted: An ensemble selected from the present MD trajectory and an ensemble generated for the Cdc25B catalytic domain by Zuckerman and co-workers.⁴³ See the Methods section for details about ensemble construction. The Zuckerman ensemble contained 250 structures with C_{α} RMSD from the X-ray structure varying from 3 to 10 Å. The core of the Cdc25B catalytic domain (residues 377–530) remains stable and folded. But, large fluctuations are again observed in the C-terminus. In fact, the C-terminal helix is unfolded in all configurations in the Zuckerman ensemble. This is another independent computational evidence of the intrinsic disorder suggested above for the Cdc25 C-terminal region. Figures shown below for ligand–protein complexes with the Zuckerman ensemble are illustrative examples of configurations composing this ensemble. The average radius of gyration for protein conformations in the ensemble is 17.8 ± 0.7 Å, ~1 Å larger than the radius of gyration observed in the MD simulation. Even though the C-terminus unfolds, the main distance from residues to the protein center-of-mass does not change significantly, because the protein core remains stable and folded. In agreement with experimental evaluation of the Cdc25B crystal structure, part of the Zuckerman ensemble had a disulfide bridge formed between C426 and C473 (corresponding to the PDB model 1cwr).¹³ Only the 176 structures without such bridge were used here.

Figure 4(A) shows the projections of the Zuckerman ensemble and of 1200 snapshots from the 60 ns MD trajectory spaced by 50 ps (including the MD ensemble) on the first and second eigenvalues of the C_α covariance matrix of the join set of configurations. These two principal components (PC) describe collective motions mainly by the last 20 residues (531–550) modeled in the Cdc25B C-terminus. The configurational space spanned by the Zuckerman ensemble (~ 14 nm for PC1 and PC2) is considerably larger than the space visited by the MD trajectory (~ 3 nm for PC1 and PC2). The MD trajectory overlaps the same four to eight ensemble structures in the first six PCs calculated [only PC1 and PC2 are shown in Fig. 4(A)]. The cosine content of PC1 is lower than 20% and it is lower than 1% for the other initial PCs.⁶³ These results suggest that the Zuckerman ensemble correctly sampled relevant solution conformations (the MD structures) but also sampled a much larger configurational space corresponding to large scale fluctuations of the disordered C-terminus. Figure 4(B) shows a contact map representation obtained from averaged distances for both ensembles. The tertiary structure is very similar between residues 374 and 530 for both ensembles. The largest differences are observed for contacts involving the C-terminus. The terminal helix is partially formed in the MD ensemble and in contact with the protein main body, particularly to residues 428–431, 480–483, and 445–446. In the Zuckerman ensemble, the C-terminus makes little to no contacts with the protein main body.

Deviations were calculated for both ensembles in comparison with the Cdc25B crystallographic structure. For the backbone atoms in the stable region (residues 377–530), the average RMSD is 1.1 ± 0.1 Å and 2.9 ± 0.1 Å for the MD and Zuckerman, respectively. The difference between the two ensembles is not large and similar to the deviations often observed between different crystallographic structures and solution NMR ensembles of the same protein.³⁰ For C_α of residues 531–550 in the C-terminus, two RMSDs distributions are clearly observed for the MD ensemble, one for configurations visited up to 40 ns, with average RMSD = 3.5 ± 0.7 Å, and another after 40 ns, with average RMSD = 8.2 ± 1.3 Å. The C-terminus average RMSD is 18.3 ± 6.0 Å for the Zuckerman ensemble. Such large RMSDs are a consequence of C-terminus local unfolding as detailed above. For the side chains of 29 residues in the surface of the stable protein region that may interact with small-molecule ligands (see next section), the average RMSD is 3.0 ± 0.4 Å and 3.7 ± 0.1 Å for the MD and Zuckerman, respectively. Together with Figure 4, these RMSD numbers indicate the degree of conformational variability between the two ensembles and the crystal structure.

Given the lack of structural information for Cdc25 in solution or bound to small-molecule inhibitors (see Introduction), which could be used to compare with the present ensembles and with the protein–ligand complexes

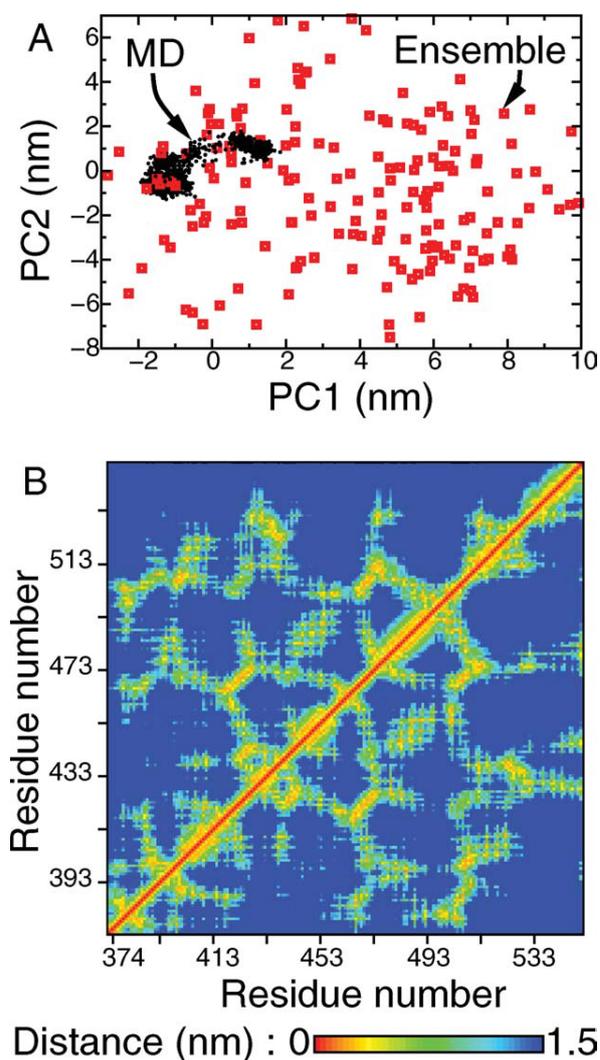


Figure 4

Comparison of the conformational ensembles used here. Top panel A shows the projections along the first and second eigenvalues obtained from principal component analysis of the MD trajectory (black) and the Zuckerman ensemble (red). Lower panel B shows a contact map representation of the MD ensemble (lower right half) and of the Zuckerman ensemble (upper left half) with smaller mean distances between residues color coded. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

generated by modelling, it is difficult to assess the quality of each Cdc25 structural representation. Thus, three models with variable levels of flexibility are identified and used in the next section for analysis of protein–ligand complexes: The rigid model corresponds to the crystal structure (1qb0) and contains a frozen and folded C-terminus; A semiflexible model corresponds to the MD ensemble and represents a distribution between a folded and a partially unfolded C-terminus; A fully flexible model with the C-terminus totally unfolded corresponds to the Zuckerman ensemble.

Each model has distinct structural characteristics and may suffer from different artifacts. The RMSF profiles obtained from experiment and from the MD simulation [Fig. 3(B)] indicate that the local fluctuations Cdc25B undergoes in the crystal are significantly different from fluctuations in solution suggesting that the crystallographic structure might not be a good model for the protein geometry in native conditions. Crystal packing and near-neighbor interactions in the crystal lattice may significantly alter the conformational distribution.^{30,31} In extreme cases, the PDB model, which represents an average of this altered distribution, may not be an appropriate model for a flexible protein in solution. In addition, the rigid model will not be able to account for any induced fit caused by ligand complexation.

Although the MD ensemble was obtained from an MD trajectory that follows the equilibrium distribution, the ad hoc selection of a subset of this trajectory, either as structures equally spaced in time (as used here and often in ensembles derived from MD³²) or in RMSD distance²² may artificially bias the distribution. The MD ensemble will also suffer from undersampling for large-scale fluctuations in the Cdc25 C-terminus.²³ The semi-flexible model will, however, partially account for induced fit effects. Configurations induced by ligand binding should also be present in the apo state ensemble, although possibly in smaller populations.^{64–66} Actually, by observing whether the population of ligand-Cdc25 complexes is enriched by structures obtained before or after 40 ns in the MD trajectory, it will be possible to tell if inhibitor complexation shifts the population distribution towards a state with folded or partially unfolded C-terminus, respectively.

In the Zuckerman ensemble, the large C-terminal fluctuations might be an artifact of the incomplete backbone sequence used in the ensemble generation or caused by an inappropriate choice of simulation parameters, in special the reduced temperature. The distribution of side chain dihedrals might also be incorrect, particularly for side chains exposed to solvent (and in contact with putative ligands).^{67,68} The protocol used to build the side chain orientation for each backbone configuration is based on average rotameric information collected from apo crystal structures.⁵⁴ Nevertheless, assuming that the Cdc25 C-terminus is unfolded in solution as suggested in the previous section, the fully flexible representation will be the best model to analyze the effect of flexibility on small molecule complexation because it samples fluctuations of the C-terminus in a statistically rigorous fashion for the protein backbone.

Binding modes for complexation to Cdc25 models

Ligands 1 and 2 (Fig. 2), two of the most potent Cdc25B inhibitors known to date,^{6,69,70} were docked to

Table II

Summary of the Valid Poses and Clusters Found for Docking to Each Cdc25B Structural Model

Model	1qb0	MD	Zuckerman
Ligand 1			
Valid Poses	3	361	426
Poses per structure	3	2.4	2.4
Number of clusters		48	52
Poses in clusters 1–4		46%	45%
Poses in clusters 1–8		63%	65%
Ligand 2			
Valid Poses	1	389	405
Poses per structure	1	2.6	2.3
Number of clusters		38	56
Poses in clusters 1–4		52%	48%
Poses in clusters 1–8		74%	67%

the Cdc25B PDB structure (code 1qb0) as well as to the MD and Zuckerman ensembles using the AutoDock empirical energy function.

A large number of docking poses was found for each ensemble. Table II shows results of the clustering procedure used to aggregate and classify valid poses (defined in Small molecule docking and clustering section). The average number per ensemble structure indicates that two to three valid poses were obtained for each receptor geometry. Throughout the remainder of this article, each cluster of valid poses will be identified as a possible Cdc25B-ligand binding mode. Up to 50 different clusters were found for each ligand-ensemble pair, but most of these contain very few (1–5) valid poses. Half of the valid poses are aggregated in the four most populated clusters and two-thirds are in the eight most populated clusters.

Because of the lack of structural information about Cdc25-inhibitor complexes, a pragmatic criteria combining high population and favorable energy score was used to choose clusters for further analysis: The eight most populated clusters for each ligand-ensemble pair were validated and rescored based on an estimate of the binding free energy. Several studies suggest that the most populated clusters are better predictors of the native binding mode than the docking pose with most favorable score.^{28,71} This may be viewed as an averaging procedure that compensates noisy scoring functions and is based on the assumption that the global free energy minimum contains the largest amount of structural neighbors, increasing the available configurational entropy.⁷²

As the energy functions used for docking are rather crude,^{23,55,73} a more accurate estimate of the binding free energy was necessary to rescore the clusters found. Absolute binding free energies could be calculated exactly for a given force field by statistical perturbation or thermodynamic integration.⁵⁹ However, the slow convergence and high computational cost preclude using these methods for the large set of complexes studied here. Instead, the linear interaction energy (LIE) estimate of

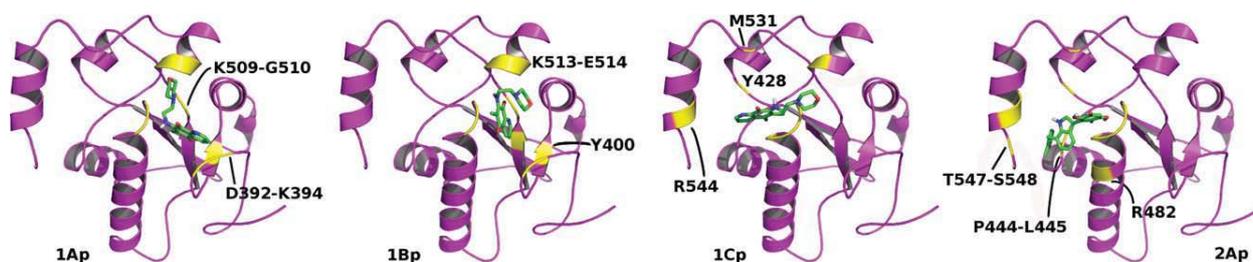


Figure 5

Complexes between the Cdc25B crystal structure and **1** (three structures on the left) and **2** (structure on the right). Residues in contact with the ligand (distance <4.5 Å) are shown in yellow and printed accordingly. P-loop residue numbering was not printed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the binding free energy was used because it is a more economic but still reliable free energy model.^{58,74} LIE approximations have their roots on physical theories of solvation linear response and, in principle, can yield absolute free energies.⁵⁸ However, the formulations used in practice, such as eq. 3, resemble linear structure-activity relationships which are empirically trained by adjusting coefficients [α and β in Eq. (3)]. Thus, parameterized LIE relationships are often used to rank a series of related ligands and its accuracy therefore depends on the similarity between ligands. Because **1** and **2** are structurally different and the LIE relationship used here was not trained for these specific ligands, we do not expect to discriminate the most potent between the two inhibitors, neither to obtain LIE estimates in close agreement with experimental binding free energies. However, the LIE values are well-suited to rank the various clusters or binding modes found within each ligand-structural model pair because errors will largely cancel between complexes with the same ligand.⁵⁸ Because the LIE includes both receptor and ligand local flexibility, a detailed all-atom force field description for the energy function and explicit solvent contributions, its ranking accuracy is expected to be superior to the AutoDock energy function.⁵⁸

The number of valid poses used in the LIE evaluation for each cluster was composed by the maximum between 15% of the cluster size and 4 poses. The cluster centroid was included in the set, and the remaining poses were randomly chosen within each cluster.⁷³ For the rigid model, LIE scores represent averages \pm standard deviations obtained for four geometries in the same valid pose. Complexes were considered relevant binding modes in the discussion below only if their average LIE score less one standard deviation was smaller (or more stable) than the lowest average LIE plus one standard deviation within the same ligand-structural model pair. Only binding modes with the lowest intrinsic binding free energies (or highest partial association constant) will contribute significantly to the measured inhibition constants.^{64,75}

Figure 5 shows the three valid poses found for **1**, and one valid pose found for **2** by docking to the rigid crys-

tallographic model. LIE scores for complexes 1Ap, 1Bp, and 1Cp are -12.4 ± 1.0 , -6.3 ± 1.8 and -12.6 ± 2.2 kcal/mol, respectively. Complex 1Bp will not be relevant because its intrinsic binding free energy is 6.0 kcal/mol lower. In complex 1Ap, the ligand is placed in a shallow pocket formed beside the P-loop, with the quinolinedione moiety in contact with residues D392-K394, Y400 and K509, and the morpholine group in contact to K513-E514. In 1Cp, the inhibitor is docked to the P-loop, with the quinolinedione also in contact to residues Y428, M531, and R544. Complex 2Ap LIE score equals to -7.9 ± 1.4 kcal/mol and shows the dihydroxyquinone group complexed with the P-loop and M531, the indole ring in contact with Y428 and R482, and the 2-methylbenzyl group coordinated to the C-terminus.

AutoDock and LIE results for binding free energies differ qualitatively. For instance, AutoDock gives similar ranking for complexes 1Ap, 1Bp, and 1Cp, but the LIE value for 1Bp is much higher than for the other two complexes. Similar qualitative differences are observed between AutoDock and LIE scores for the ensemble complexes. Tables III and IV give average LIE and standard deviation, as well as the number of valid poses found for the eight most populated clusters in each ligand-ensemble pair. Standard deviations from the average LIE energy are appreciable. RMSDs for the ligand atom positions between initial and final geometries out of the LIE MD trajectories vary from 0.5 to 1.7 Å, without any correlation to LIE deviations. The LIE deviations may be due to a lack of convergence in the average potential energies used in Eq. (3) and obtained from the relatively short (1.5 ns) MD trajectories.^{58,60} For complexes with the ensemble, LIE deviations are also caused by inherent variability in initial geometries among valid poses within each cluster.

Normalized histograms depicting contacts between ligand and protein were built for complexes with both ensembles (four figures in the Supporting Information). A contact was defined if at least one residue atom lays at a distance ≤ 4.5 Å from any ligand atom. A list of contacts was generated for each valid pose and the fraction of each contact found within a cluster was calculated.

Table III
Properties of the Eight Most Populated Clusters Found for **1**
Complexed to Cdc25B Ensembles

MD ensemble								
Complex	1Am	1Bm	1Cm	1Dm	1Em	1Fm	1Gm	1Hm
Valid poses	91	31	24	19	17	15	15	14
Poses used for LIE	14	5	4	4	4	4	4	4
Average LIE ^a	-5.4	-2.8	-0.6	-9.3	-7.7	-4.8	-11.0	-5.0
LIE Std. Deviation ^a	2.7	2.6	1.2	1.9	0.6	2.5	1.8	1.2
Zuckerman ensemble								
Complex	1Az	1Bz	1Cz	1Dz	1Ez	1Fz	1Gz	1Hz
Valid poses	73	51	39	27	26	22	20	19
Poses used for LIE	11	8	6	4	4	4	4	4
Average LIE ^a	-6.7	-10.8	-5.6	-8.8	-5.8	-10.5	-7.6	-6.1
LIE Std. Deviation ^a	2.4	1.7	2.1	2.1	1.2	2.2	1.0	2.5

^aAverage binding free-energy and standard deviation from LIE calculation given in kcal/mol.

The histograms identify the main contacts formed and their dispersion within each cluster.

From the eight most populated clusters found for binding between **1** and the MD ensemble, only 1Dm and 1Gm are identified as relevant modes by LIE scoring. Geometries for the cluster centroids representing these complexes are shown in Figure 6. Although its LIE score is rather low, 1Am is also shown because it corresponds to the most populated cluster found (Tables III and IV). In 1Am, the ligand is complexed to the P-loop and with the C-terminus. In 1Dm, the quinolinedione is buried in the same shallow pocket as in 1Ap. In 1Gm, the quinolinedione is complexed in front of the P-loop and the morpholine is buried inside the P-loop. Notably, all complexes, but 1Dm and 1Em, show contacts between **1** and W550 (See Supporting Information).

Figure 7 shows the centroid geometry for 4 out of the 5 clusters identified as relevant binding modes for complexes between **1** and the Zuckerman ensemble. The 8 most populated clusters found can be roughly divided in two groups. In 1Az and 1Dz, the ligand assumes different orientations but binds to the shallow pocket formed beside the P-loop and also observed in complexes 1Ap and 1Dm. In fact, 1Az and 1Ap correspond to almost identical binding modes with equivalent ligand-residue contacts. The other group is formed by 1Bz, 1Fz, and 1Gz and shows the ligand bound into the P-loop. In 1Gz (geometry not shown), the ligand occupies the same site as in 1Bz and 1Fz, but with morpholine coordinated to the P-loop and the quinolinedione in contact with residues M531-H533.

Thus, the same two binding sites are observed for **1** complexed with Cdc25B in the three structural models analyzed. In complexes 1Ap, 1Dm, 1Az, and 1Dz, the shallow pocket formed beside the P-loop together with residues D392-H395, Y400, K509-G510, and K513-E514 is occupied. It may accommodate two orientations of the

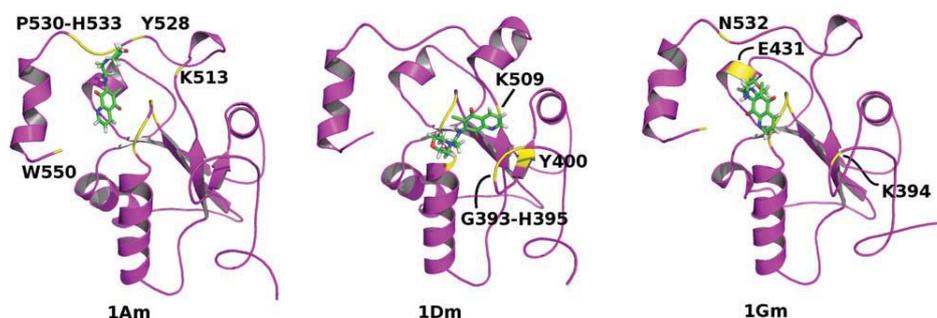
quinolinedione group: complexed with the P-loop (complex 1Dz) or with residues D392-H395 (1Ap, 1Dm, and 1Az). The morpholine group may be exposed to solvent (1Dm) or in contact with K513-E514 (1Ap and 1Az). Backbone flexibility in the C-terminus and the protein main-body has little influence on binding to the shallow pocket since all structural models can account for binding. This is an obvious conclusion given the site distance from the C-terminus and the stability of the protein core (see Cdc25B is partially unfolded in solution section). However, the precise orientation and the contacts formed by **1** depend on the flexibility of P-loop and shallow pocket side chains.

The second binding site observed in 1Cp, 1Am, 1Gm, 1Bz, and 1Fz shows the quinolinedione coordinated to the P-loop in several orientations. The contacts formed and the ligand orientation are largely influenced by the C-terminus backbone flexibility. In the rigid model, dense packing by the C-terminus complexed with the protein main-body and side chains of Y428, M531, and R544 leaves little room for **1** so that the morpholine group points to the shallow pocket beside the P-loop and coordinates K513 (complex 1Cp). The increased receptor flexibility modeled by the MD ensemble corresponds to a partial detachment of the C-terminus from the protein main-body and allows ligand reorientation. Therefore, the morpholine group may coordinate deeper in the direction of residues E431 (1Gm) or Y528 (1Am). Although the C-terminus fluctuates widely in the fully-flexible model, the region composed of residues M531-N532 moves on average to configurations closer to the P-loop and makes favorable contacts with **1** (1Bz and 1Fz). Such fluctuations also allow placing the morpholine group in the space occupied by the C-terminus in the more rigid models. Backbone flexibility thus yields ligand complexation with regions of the stable protein core that were unavailable and unexposed in the rigid model.

Table IV
Properties of the Eight Most Populated Clusters Found for **2**
Complexed to Cdc25B Ensembles

MD ensemble								
Complex	2Am	2Bm	2Cm	2Dm	2Em	2Fm	2Gm	2Hm
Valid poses	71	58	41	32	23	23	21	18
Poses used for LIE	11	9	6	5	4	4	4	4
Average LIE ^a	-8.9	-6.2	-8.7	-9.4	-11.5	-6.7	-6.8	-7.2
LIE Std. Deviation ^a	1.4	0.8	1.6	2.2	1.4	1.6	0.9	2.6
Zuckerman ensemble								
Complex	2Az	2Bz	2Cz	2Dz	2Ez	2Fz	2Gz	2Hz
Valid poses	67	60	37	32	28	18	15	13
Poses used for LIE	10	9	6	5	4	4	4	4
Average LIE ^a	-7.3	-8.9	-8.8	-10.0	-8.4	-7.7	-8.0	-9.8
LIE Std. Deviation ^a	1.3	0.8	1.6	2.5	2.3	1.9	1.4	1.3

^aAverage binding free-energy and standard deviation from LIE calculation given in kcal/mol.

**Figure 6**

Relevant complexes found between **1** and the MD Cdc25B ensemble. Residues in contact with the ligand (distance <4.5 Å) and with a normalized count higher than 5% (see Fig. 11 and explanation on text) are shown in yellow and printed accordingly. P-loop residue numbering were not printed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Noteworthy, the LIE scores suggest that 1Cp, 1Gm, and 1Bz/1Fz are more stable than 1Ap, 1Dm, and 1Az/1Dz, respectively in each structural model. Thus, binding **1** to the P-loop should be more stable than binding to the shallow pocket.

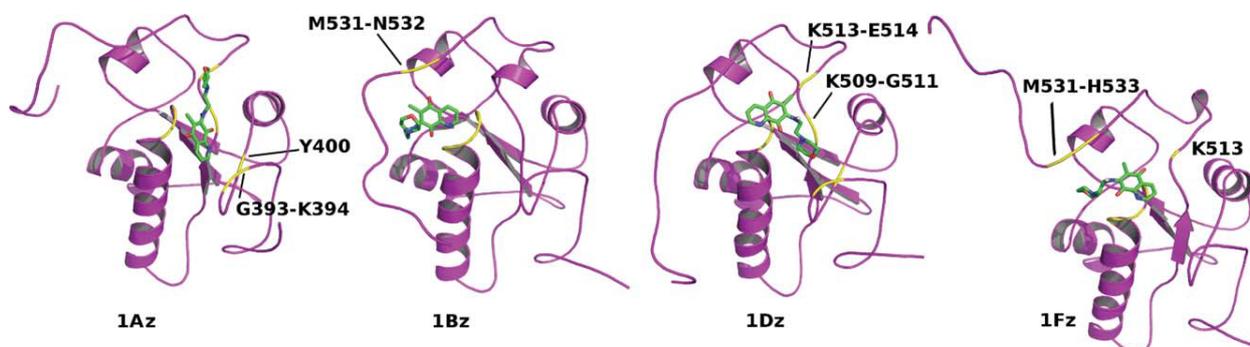
Complexes between **2** and the MD ensemble that represent relevant binding modes are shown in Figure 8. In 2Am, the dihydroxyquinone is coordinated above the P-loop with the 2-methyl-benzyl moiety in contact with the C-terminus. In both 2Cm and 2Em, the ligand is complexed outside the P-loop, with the 2-methyl-benzyl group coordinated to the shallow pocket. Complex 2Em is the most stable and the only binding mode that does not coordinate to W550. In 2Dm, the dihydroxyquinone group coordinates both the P-loop and N532.

Complexes between **2** and the Zuckerman ensemble are shown in Figure 9. All of them have LIE scores within one standard deviation from the cluster with lowest average LIE. Therefore, the eight complexes may represent relevant binding modes. In both 2Az and 2Bz, **2** is complexed to the shallow pocket formed beside the P-loop. Complex 2Gz shows **2** coordinated with the C-terminus and only R479 in the P-loop. In the remaining complexes, the ligand binds to the P-loop. Similar contacts, excluding those with the C-terminus, are observed in 2Fz and 2Ap.

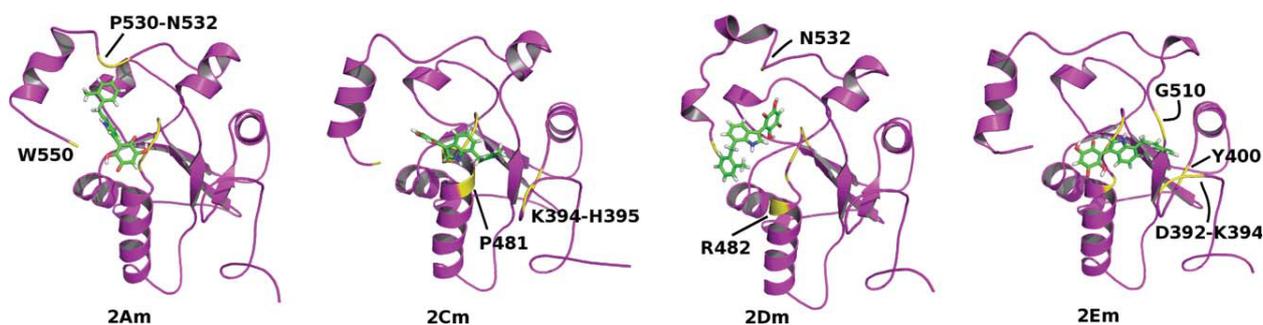
Because **2** is a bulkier ligand, it does not fit entirely in the shallow pocket identified above for **1**. Docking to the rigid crystal model results in only one binding mode where **2** is complexed to the P-loop and the C-terminus (2Ap). Increasing receptor flexibility leads to complexes in which (part of) **2** coordinates the shallow pocket (2Cm, 2Em, 2Az, and 2Bz). In fact, **2** fits completely in the shallow pocket only in the fully flexible model (2Az and 2Bz). Thus, backbone flexibility in the C-terminus grants access of bulkier groups to cavities transiently formed on the surface of the stable folded protein core.⁶⁶ Complexes coordinating the shallow pocket are less stable than complexes bound to the P-loop, as observed for **1**. Increased flexibility also results in a larger diversity of modes for **2** complexed with the P-loop.

For the semiflexible model, the fraction of receptor structures in each cluster obtained after 40 ns in the MD trajectory suggests whether complexation shifts the popu-

minus and only R479 in the P-loop. In the remaining complexes, the ligand binds to the P-loop. Similar contacts, excluding those with the C-terminus, are observed in 2Fz and 2Ap.

**Figure 7**

Relevant complexes found between **1** and the Zuckerman Cdc25B ensemble. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**Figure 8**

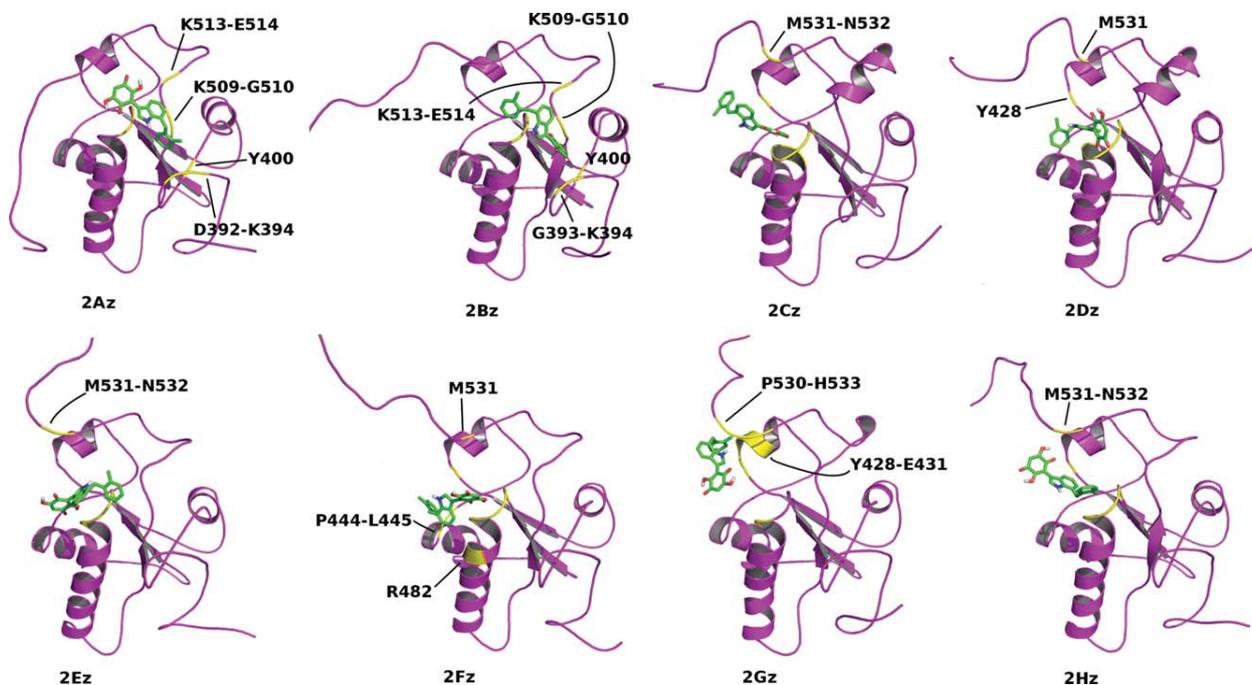
Relevant complexes found between **2** and the MD Cdc25B ensemble. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

lation distribution toward a state with a partially unfolded C-terminus.^{65,66} Because one-third of the MD ensemble corresponds to structures obtained after 40 ns, a similar fraction would be expected in a cluster without a shifted receptor population. Modes 1Dm and 1Gm have fractions of 0.26 and 0.67, respectively. As expected from the complex geometries (Fig. 6), 1Dm does not have a clear dependence on the C-terminal flexibility. But, 1Gm shows significant enrichment for receptor structures corresponding to larger fluctuations of the C-terminus. Ligand coordination in 1Gm requires a receptor with less dense packing between the C-terminus and

the protein main-body. For **2**, 2Dm is the only relevant complex with a fraction (0.59) significantly different from one-third, also in agreement with observation that the ligand requires more orientation freedom than available in the rigid model.

Comparison with previously available data

A series of quinolinediones were tested experimentally for Cdc25 inhibition.⁶⁹ The most potent inhibitor found was **1** ($IC_{50} = 0.21 \mu M$), 15-fold more potent than the unsubstituted quinolinedione. Several substitutions of the

**Figure 9**

Relevant complexes found between **2** and the Zuckerman Cdc25B ensemble. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

morpholine-ethylamino group for other similar sized amino derivatives, removal or substitution of chloride by methoxide as well as different quinones (isoquinoline, quinazoline, etc.) resulted in 2- to 10-fold decrease in inhibition potency. The regioisomer of **1** with chloride and morpholine-ethylamino group positions exchanged was four-fold less potent. Thus, the quinolinedione is the main function responsible for inhibitory activity and the specific contacts of substituents have relatively less importance. Kinetic data for inhibition by **1** fit to a partial mixed competitive model⁶⁹ and displays a time-dependent inhibition.⁷⁰ As it was shown that **1** irreversibly oxidizes the catalytic C473 of Cdc25B,⁷⁶ it is plausible that **1** occupies two different binding sites in Cdc25: one for reversible, competitive binding and another for irreversible oxidation. Average edge-to-edge distances between the catalytic C473 and **1** are 3.3 ± 0.6 Å and 7 ± 1 Å for the binding mode on top of the P-loop (complexes 1Cp, 1Gm, 1Bz, and 1Fz) and in the shallow pocket beside the active site (1Ap, 1Dm, 1Az, and 1Dz), respectively. Thus, both sites may allow fast electron transfer for irreversible oxidation of C473.⁷⁷ On the other hand, only the complexes with **1** on top of the P-loop would display kinetics consistent with competitive inhibition.

Several indolyldihydroxyquinones related to **2** were screened for Cdc25 inhibition.⁷⁰ Modifications in the 2- and 4-position of the indole ring (position 1 corresponds to the nitrogen atom and 3 corresponds to the bond with quinone) lead to loss of inhibitory potency by hindering the torsion around the quinone bond. Modifications at 5- and 6-position have little effect, indicating that these positions do not make specific contacts with the receptor. Substitutions at 7-position with hydrophobic groups bulkier than propyl increase potency suggesting that the 2-methyl-benzyl group in **2** occupies a hydrophobic pocket. Two related inhibitors were screened against various Cdc25B point mutants: Molecule **2A** (in the nomenclature used in the original article⁷⁰) lacks the 2-methyl-benzyl group at 7-position and this position is substituted by the less bulky prenyl group in molecule **6B**. In comparison with wild-type inhibition, mutants E474Q, F475A, and R482L have 5- to 10-fold less inhibition potency by both **2A** and **6B** indicating that these residues make specific contacts with the indole and quinone groups. A Cdc25B construct lacking the last 18 C-terminal residues (549–566) is five-fold less inhibited by **6B**, but has no significant change when inhibited by **2A**, suggesting that specific contacts are not established with the C-terminus by the indole and quinone groups. Mutants E478Q, Y528F, M531A, N532A, and R544L do not show significant changes when inhibited by **6B**, suggesting that these residues do not make specific contacts with ligands similar to **2**. Kinetic data for indolyldihydroxyquinones indicate that these are reversible and competitive ligands with time-independent inhibition.⁷⁰

The above discussion may help to narrow down the possible binding modes found for **2**. Complex 2Ap is not a good model because specific contacts are established between the indole ring and the protein C-terminal. Complex 2Am seems unlikely because it makes no contact with R482 (or the neighbor P481), and coordinates residues which deletion or mutation does not affect binding considerably (W550, N532, and M531). Complex 2Dm is also unlikely because it shows specific contacts with N532 and the 2-methyl-benzyl group is exposed to solvent. Similar reasoning suggests that 2Cz, 2Dz, 2Gz, and 2Hz are also unlikely.

The rigid Cdc25B crystal structure was used as a receptor model in another docking study.⁷⁸ For **1**, the binding mode found resembles complex 1Cp with the ligand slightly dislocated and forming a bidentate hydrogen bond with R482. Two modes were found for **2** corresponding to the same binding site as 2Ap, but with either 2-methyl-benzyl or dihydroxyquinone buried in the P-loop. The main difference in comparison with the results presented here were contacts observed between R544 and both **1** and **2**.⁷⁸ Given the same receptor structure was used, this variance is due to the different docking algorithms used.

CONCLUSIONS

Results from molecular dynamics simulations and bioinformatic searches presented here and of computer-generated conformational ensembles⁴³ indicate that the C-terminus is either unfolded or in equilibrium between a partially folded and a disordered state in Cdc25B under normal conditions in aqueous solution. Because of the growing interest in targeting Cdc25 phosphatases for development of anti-cancer therapeutics,^{1–7} these results should warn those using the available Cdc25 crystal structures in structure-based drug design that such receptor configurations may be flawed or incomplete.

Complexes were obtained between three Cdc25 structural models and two quinone-based small molecules, which are potent Cdc25 inhibitors. The structural models had variable levels of flexibility so that the effect of Cdc25 C-terminal unfolding on small-molecule complexation could be evaluated. The same two binding sites were observed for **1** complexed with Cdc25B in the three structural models analyzed in agreement with experimentally measured mixed kinetics.⁶⁹ Binding to the shallow pocket formed beside the P-loop may account for irreversible oxidation of the catalytic C473.⁷⁶ Backbone flexibility has little influence on this binding site, but the contacts formed by **1** depend on the flexibility of P-loop and shallow pocket side-chains. Binding to the P-loop may occur in several orientations, which are clearly influenced by C-terminus backbone flexibility, and by the appearance of cavities unexposed in the rigid model. Complexation to the P-loop has a more favorable free

energy and may account for both irreversible oxidation and competitive inhibition.

In contrast to **1**, there should be only one mechanism of inhibition by **2** binding competitively to the active site.⁷⁰ This study and previously available data do not allow an unambiguous identification of this binding mode. The most likely possibilities are either described by complexes 2Cm and 2Em or 2Az and 2Bz or 2Fz. However, it is unlike that binding of **2** is a multimodal process.^{64,75} Direct interactions with the C-terminus should be negligible for all complexes, but 2Fz. Nevertheless, inclusion of flexibility is essential to uncover the hydrophobic regions for binding the bulky 2-methyl-benzyl group and to obtain complexes in agreement with experimental data. It is expected that site-directed mutagenesis of Cdc25B residues D392-H395, Y400, Y428, P444-L445, K509-G510, and K513-E514 should help to resolve the occupation of the shallow pocket beside the P-loop and similar hydrophobic cavities.

As noted, backbone fluctuations alter the available binding sites in Cdc25B. The MD ensemble only hints at the Cdc25B transient cavities because of insufficient backbone sampling. A major advantage of using an ensemble following an equilibrium distribution, such as the Zuckerman ensemble as a structural model for receptor docking is the appropriate sampling of large scale fluctuations,⁴³ which help to disclose the transient binding sites. The distribution of side chain rotamers, specially for side chain exposed to the solvent and inside binding pockets is also fundamental for ligand-receptor interactions.^{23,26} But, because the side chain distribution in the Zuckerman ensemble does not follow an equilibrium distribution and given the lack of experimental structural information for comparison, we have refrained from discussing detailed interactions. It may be anticipated that ligands bound to Cdc25 will shift the population of side chain rotamers, as observed in the Cdc25A active site between apo¹² and substrate-bound structures.⁷⁹

Contacts observed between ligands and putative disordered regions (here the Cdc25B C-terminal) in rigid receptor models will not be maintained when backbone flexibility is introduced because such disordered regions fluctuate wildly and do not assume a unique or stable set of configurations. New or cryptic binding sites involving the disordered region are not observed for the same reason. The main effect of modeling backbone flexibility is the formation of transient cavities or compact hydrophobic units⁶⁶ on the surface of the stable, folded protein core that are unexposed or unavailable for ligand binding in rigid and densely packed structures. These cavities are particularly important for binding to bulkier ligands with hydrophobic moieties. Repositioning of flexible but still ordered regions (e.g., residues M531-N532 in Cdc25B) also creates additional sites for contacts with ligands in comparison to a rigid model. Thus, the increased flexibility results in a larger diversity of simulated binding modes, in particular

to bulkier ligands. Further, an approximate and economic receptor model for Cdc25B and other partially disordered proteins could be built by removing or truncating the partially unfolded or disordered regions (residues 531–550 for Cdc25B) from the available crystallographic or solution structure. Corroboration of the present conclusions and simulation results by structural and binding studies for different Cdc25B constructs is now underway in collaboration with experimental groups.

ACKNOWLEDGMENTS

Comments on the manuscript by Shaker Chuck Farah (Universidade de Sao Paulo) are acknowledged.

REFERENCES

- Boutros R, Lobjois V, Ducommun B. Cdc25 phosphatases in cancer cells: key players? Good targets? *Nat Rev Cancer* 2007;7:495–507.
- Brezak MC, Kasprzyk P, Galcera MO, Lavergne O, Prévost GP. Cdc25 inhibitors as anticancer agents are moving forward. *Anti-Cancer Agents in Med Chem* 2008;8:857–862.
- Lazo JS, Wipf P. Is Cdc25 a druggable target? *Anti-Cancer Agents in Med Chem* 2008;8:837–842.
- Bakan A, Lazo JS, Wipf P, Brummond KM, Bahar I. Toward a molecular understanding of the interaction of dual specificity phosphatases with substrates: insights from structure-based modeling and high throughput screening. *Curr Med Chem* 2008;19:2536–2544.
- Vintonyak VV, Antonchick AP, Rauh D, Waldmann H. The therapeutic potential of phosphatase inhibitors. *Curr Opin Chem Biol* 2009;13:272–283.
- Lavecchia A, Giovanni CD, Novellino E. Cdc25A and B dual-specificity phosphatase inhibitors: potential agents for cancer therapy. *Curr Med Chem* 2009;16:1831–1849.
- Park H, Bahn YJ, Ryu SE. Structure-based de novo design and biochemical evaluation of novel Cdc25 phosphatase inhibitors. *Bioorg Med Chem Lett* 2009;19:4330–4334.
- Rudolph J. Cdc25 phosphatases: structure, specificity, and mechanism. *Biochemistry* 2007;46:3595–3604.
- Arantes GM. The catalytic acid in the dephosphorylation of the Cdk2-pTpY/CycA protein complex by Cdc25B phosphatase. *J Phys Chem B* 2008;112:15244–15247.
- McCain DF, Catrina IE, Hengge AC, Zhang ZY. The catalytic mechanism of Cdc25A phosphatase. *J Biol Chem* 2002;277:11190–11200.
- Arantes GM. Free energy profiles for catalysis by dual-specificity phosphatases. *Biochem J* 2006;399:343–350.
- Fauman EB, Cogswell JP, Lovejoy B, Rocque WJ, Holmes W, Montana VG, Piwnicka-Worms H, Rink MJ, Saper MA. Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. *Cell* 1998;93:617–625.
- Reynolds RA, Yem AW, Wolfe CL, Deibel MR, Chidester CG, Watenpaugh KD. Crystal structure of the catalytic subunit of Cdc25b required for G(2)/M phase transition of the cell cycle. *J Mol Biol* 1999;293:559–568.
- Sohn J, Parks JM, Buhrman G, Brown P, Kristjansdottir K, Safi A, Edelsbrunner H, Yang W, Rudolph J. Experimental validation of the docking orientation of Cdc25 with its Cdk2-CycA protein substrate. *Biochemistry* 2005;44:16563–16573.
- Cavanagh J, Fairbrother WJ, Palmer AG, III, Skelton NJ. *Protein NMR Spectroscopy: principles and Practice*. 2nd edn. New York: Academic Press; 2007.
- Wilborn M, Free S, Ban A, Rudolph J. The c-terminal tail of the dual-specificity Cdc25B phosphatase mediates modular substrate recognition. *Biochemistry* 2001;40:14200–14206.

17. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. The roles of intrinsic disorder in protein interaction networks. *FEBS Lett* 2005;272:5129–5148.
18. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21^{Waf1/Cip1/Sdi1} in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* 1996;93:11504–11509.
19. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol* 2009;19:31–38.
20. Jorgensen WL. The many roles of computation in drug discovery. *Science* 2004;303:1813–1818.
21. Sotriffer CA, Flader W, Winger RH, Rode BM, Leidl KR, Vargas JM. Automated docking of ligands to antibodies: methods and applications. *Methods* 2000;20:280–291.
22. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* 2008;18:178–184.
23. Cozzini P, Kellogg GE, Spyrafis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* 2008;51:6237–6255.
24. Kallblad P, Dean PM. Efficient conformational sampling of local side-chain flexibility. *J Mol Biol* 2003;326:1651–1665.
25. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins: Struct Funct Genet* 2002;46:34–40.
26. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J Med Chem* 2004;47:45–55.
27. Knegt RMA, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. *J Mol Biol* 1997;266:424–440.
28. Prasad JC, Goldstone JV, Camacho CJ, Vajda S, Stegeman JJ. Ensemble modeling of substrate binding to cytochromes p450: analysis of catalytic differences between cyp1a orthologs. *Biochemistry* 2007;46:2640–2654.
29. Burra PV, Zhang Y, Godzik A, Stec B. Global distribution of conformational states derived from redundant models in the pdb points to non-uniqueness of the protein structure. *Proc Natl Acad Sci USA* 2009;106:10505–10510.
30. Damm KL, Carlson HA. Exploring experimental sources of multiple protein conformations in structure-based drug design. *J Am Chem Soc* 2007;129:8225–8235.
31. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed* 2003;42:2718–2736.
32. Wong CF, Kua J, Zhang Y, Straatsma T, McCammon JA. Molecular docking of balanol to dynamics snapshots of protein kinase A. *Proteins: Struct Funct Genet* 2005;61:850–858.
33. Frembgen-Kesner T, Elcock AH. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 map kinase. *J Mol Biol* 2006;359:202–214.
34. Ensing B, Vivo MD, Liu Z, Moore P, Klein ML. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc Chem Res* 2006;39:73–81.
35. van Aalten D, de Groot B, Berendsen H, Findlay J, Amadei A. A comparison of techniques for calculating protein essential dynamics. *J Comput Chem* 1997;18:169–181.
36. Zuckerman DM, Lyman E. A second look at canonical sampling of biomolecules using replica exchange simulation. *J Chem Theory Comput* 2006;3:1200–1202.
37. Nagasima T, Sugita Y, Mitsutake A, Okamoto Y. Generalized-ensemble simulations of spin systems and protein systems. *Comp Phys Comm* 2002;146:69–76.
38. de Groot B, van Aalten D, Scheek R, Amadei A, Vriend G, Berendsen H. Prediction of protein conformational freedom from distance constraints. *Proteins: Struct Funct Genet* 1997;29:240–251.
39. Lange OF, Lakomek NA, Fars C, Schrder GF, Walter KFA, Becker S, Meiler J, Grubmiller H, Griesinger C, de Groot BL. Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. *Science* 2008;320:1471–1475.
40. Shehu A, Kaviraki LE, Clementi C. Multiscale characterization of protein conformational ensembles. *Proteins* 2009;76:837–851.
41. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci USA* 2005;102:17002–17007.
42. Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem* 2008;77:363–382.
43. Mamonov AB, Bhatt D, Cashman DJ, Ding Y, Zuckerman DM. General library-based monte carlo technique enables equilibrium sampling of semi-atomistic protein models. *J Phys Chem B* 2009;113:10891–10904.
44. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from nmr residual dipolar couplings. *J Am Chem Soc* 2009;131:17908–17918.
45. McQuarrie DA. *Statistical Mechanics*. 1st edn. New York: Harper and Row; 1976.
46. Hess B, Kutzner C, van der Spoel D, Lindahl E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 2008;4:435–447.
47. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc* 1996;118:11225–11236.
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
49. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann JS, Silman I, Sussman JL. A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 2005;21:3435–3438.
50. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004;337:635–645.
51. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing intrinsic disorder predictors with protein evolutionary information. *J Bioinf Comput Biol* 2005;3:35–60.
52. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. *Biophys J* 2007;92:1439–1456.
53. Liu JS. *Monte Carlo strategies in scientific computing*. 1st ed. New York: Springer; 2001.
54. Canutescu AA, Shelenkov AA, Dunbrack RL. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci* 2001;12:2001–2014.
55. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 2007;28:1145–1152.
56. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angew Chem Int Ed* 1999;38:236–240.
57. Haykin SS. *Neural networks — A comprehensive foundation*. 2nd ed. Upper Saddle River, NJ: Prentice-Hall; 1999.
58. Åqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations. *Acc Chem Res* 2002;35:358–365.
59. Pearlman DA, Rao G. Free energy calculations: methods and applications. In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer IHF, Schreiner PR, editors *Encyclopedia of Computational Chemistry*, Vol. 2, 1st ed. Chichester: Wiley; 1998. pp. 1036–1061.
60. Hansson T, Marelus J, Åqvist J. Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aid Mol Des* 1998;12:27–35.

61. Benedix A, Becker CM, de Groot BL, Caffisch A, Bckmann RA. Predicting free energy changes using structural ensembles. *Nat Methods* 2009;6:3–4.
62. Krzeminski M, Fuentes G, Boelens R, Bonvin AM. Minoes: a new approach to select a representative ensemble of structures in nmr studies of (partially) unfolded states. Application to δ 25-PYP. *Proteins* 2009;74:895–904.
63. Lange OF, Grubmuller H. Generalized correlation for biomolecular dynamics. *Proteins: Struct Funct Genet* 2006;62:1053–1061.
64. Eftink MR, Anusiem AC, Biltonen RL. Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A. *Biochemistry* 1983;22:3884–3896.
65. Bakan A, Bahar I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc Natl Acad Sci USA* 2009;105:14349–14354.
66. Teague SJ. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2003;2:527–541.
67. Heringa J, Argos P. Strain in protein structures as viewed through nonrotameric side chains: II. Effects upon ligand binding. *Proteins: Struct Funct Genet* 1999;37:44–55.
68. Gunasekaran K, Nussinov R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J Mol Biol* 2007;365:257–273.
69. Lazo JS, Aslan DC, Southwick EC, Cooley KA, Ducruet AP, Joo B, Vogt A, Wipf P. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. *J Med Chem* 2001;44:4042–4049.
70. Sohn J, Kiburz B, Li Z, Deng L, Safi A, Pirrung MC, Rudolph J. Inhibition of Cdc25 phosphatases by indolyldihydroxyquinones. *J Med Chem* 2003;46:2580–2588.
71. Kallblad P, Mancera RL, Todorov NP. Assessment of multiple binding modes in ligand-protein docking. *J Med Chem* 2004;47:3334–3337.
72. Shortle D, Simons KT, Baker D. Clustering of low-energy conformations near the native structures of small proteins. *Proc Natl Acad Sci USA* 1998;95:11158–11162.
73. Krol M, Tournier AL, Bates PA. Flexible relaxation of rigid-body docking solutions. *Proteins: Struct Funct Genet* 2007;68:159169.
74. Rizzo RC, Tirado-Rives J, Jorgensen WL. Estimation of binding affinities for hept and nevirapine analogues with HIV-1 reverse transcriptase via Monte Carlo simulations. *J Med Chem* 2001;44:145–154.
75. Khandelwal A, Lukacova V, Kroll DM, Raha S, Comez D, Balaz S. Processing multimode binding situations in simulation-based prediction of ligand-macromolecule affinities. *J Phys Chem A* 2005;109:6387–6391.
76. Brisson M, Nguyen T, Wipf P, Joo B, Day BW, Skoko JS, Schreiber EM, Foster C, Bansal P, Lazo JS. Redox regulation of Cdc25B by cell-active quinolinediones. *Mol Pharmacol* 2005;68:1810–1820.
77. Moser CC, Keske JM, Warncke K, Farid RS, Dutton PL. Nature of biological electron transfer. *Nature* 1992;355:796–802.
78. Lavecchia A, Cosconati S, Limongelli V, Novellino E. Modeling of Cdc25B dual specificity protein phosphatase inhibitors: docking of ligands and enzymatic inhibition mechanism. *Chem Med Chem* 2006;1:540–550.
79. Kolmodin K, Åqvist J. Prediction of a ligand-induced conformational change in the catalytic core of Cdc25A. *FEBS Lett* 2000;465:8–11.

ANEXO H – Conformational Flexibility of the Complete Catalytic Domain of Cdc25B Phosphatases

Conformational Flexibility of the Complete Catalytic Domain of Cdc25B Phosphatases

Raphael S. R. Sayegh, Fabio K. Tamaki, Sandro R. Marana, Roberto K. Salinas, and
Guilherme M. Arantes*

Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof.
Lineu Prestes 748, São Paulo, Brazil

March 28, 2016

Abstract

Cdc25B phosphatases are involved in cell cycle checkpoints and have become possible targets for the development of new anticancer drugs. A more rational design of Cdc25B ligands would benefit from detailed knowledge of its tertiary structure. The conformational flexibility of the C-terminal region of the Cdc25B catalytic domain has been debated recently and suggested to play an important role in the complexation of competitive inhibitors. Here, a combination of experimental NMR measurements and molecular dynamics simulations is presented for the *complete* catalytic domain of the Cdc25B phosphatase. The stability of the C-terminal α -helix is confirmed, but the last 20 residues in the complete catalytic domain are very flexible, establish different contacts with the protein core and occlude the active site. The combination of experimental and computational methods may provide detailed atomic resolution for flexible or disordered protein regions in solution.

1 Introduction

Progression in the cell cycle is controlled by molecular checkpoints where the deactivation of cyclin-dependent kinases (Cdk) complexes by Cdc25 phosphatases is an essential step[1]. Three Cdc25 homologous forms (A, B and C) are encoded in the human genome. For instance, Cdc25A has been linked to the G1/S transition of the cell cycle, while Cdc25B acts in the G2/M transition[2]. Not surprisingly, the three forms have been found to be over-expressed in several types of human tu-

mors [3], and Cdc25 phosphatases have become active targets for the development of new anti-cancer drugs [4, 5]

Cdc25 is considered a member of the protein tyrosine phosphatase (PTP) family[6], but it shares low sequence and structural similarity with other PTPs besides the conserved active site motif Cys-(Xxx)₅-Arg, known as the phosphate binding loop (P-loop). This low similarity may be useful for the design of selective inhibitors. Cdc25s also show substrate promiscuity as they dephosphorylate both pThr and pTyr residues found in Cdk via a cysteine side-chain nucleophile[1, 7, 8].

The similarity between the amino-acid sequences of the three human Cdc25 homologous forms is higher than 70% for their complete catalytic domains but drops to less than 30% in the C-terminus. Crystal structures are available for Cdc25A and Cdc25B catalytic domains[9, 10]. The only notable structural difference between the two forms is a C-terminal α -helix found only for Cdc25B, whereas a coiled coil is found in the same region for Cdc25A. Additional segments of 22 (Cdc25A) and 16 (Cdc25B) amino-acids were present in the C-terminus of both protein constructions used for crystallography, but structures for these regions were not provided in the final model probably due to an undefined electronic density map.

Previous computer simulations suggested that the C-terminal α -helix observed in Cdc25B could be (partially) unfolded in solution[11, 12]. But, recent nuclear magnetic resonance (NMR) experiments have refuted the simulation results and shown that the C-terminal helix has little flexibility and is well folded in the ps–ms time scale[13]. Unfortunately, the protein construction used for this NMR study lacked the last 16 residues present

*Corresponding author; Email: garantes@iq.usp.br

in the C-terminus of wild-type Cdc25B.

Notably, the C-terminus is involved in the recognition and discrimination between Cdk/Cyc complexes[14]. For instance, residues 556R and 562R found in the Cdc25B C-terminus have been found to modulate up to 10× each the stability of the complex with the bis-phosphorylated Cdk2-pYpT/Cyclin-A natural substrate[14]. Given the spatial proximity of the C-terminus to the P-loop in the Cdc25B structure[10], the terminal region might also play important roles in complexation processes such as: release and egress of the product inorganic phosphate[7, 8], binding and stability of bulky artificial phosphatase substrates such as O-methyl fluorescein phosphate[5], and complexation of competitive inhibitors[11].

Here, the conformational flexibility of the *complete* catalytic domain of Cdc25B was probed with experimental NMR and molecular dynamics simulations. The last 20 residues in the Cdc25B C-terminus were found to be highly flexible, partially occlude the active site and make frequent contacts with the protein core. This full-scale representation of the structure will be valuable to rationally design competitive inhibitors of Cdc25 phosphatases.

2 Methods

2.1 Cloning, expression and purification of Cdc25B

The complete Cdc25B catalytic domain (residues 373S–566Q) was amplified from a plasmid purchased from Origene (code: NM_021873) with the cDNA sequence of human Cdc25B and cloned into pET-28a vector using NdeI/BamHI restriction enzyme cloning sites.

Expression of Cdc25B fused to a His₆-tag followed by thrombin cleavage site at the N-terminal end was done in *Escherichia coli* BL21 (DE3-Gold). Cells were grown in M9 medium labeled with ¹⁵NH₄Cl and ¹³C-glucose at 37 °C until they reach log phase of growth (OD_{600nm} = 0.6-0.8). Then, the temperature was set to 20 °C and the protein expression was induced with 0.8 mM IPTG for 18 h before collection of cells by centrifugation. NMR samples used for backbone assignment were expressed in M9 medium prepared with ²H₂O to allow aleatory and partial deuteration of the protein.

Cells were lysed by sonication in buffer containing 20 mM Tris pH 7.4, 500 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol and 1 mM PMSF. After centrifugation, the supernatant was incubated with Ni²⁺ resin and washed with phosphate buffer (20 mM NaPi pH 6.7, 500 mM NaCl, 5 mM β-mercaptoethanol and 20 mM imidazole) before elution of the protein with phosphate buffer containing 500 mM imidazole.

The His₆-tag was cleaved with incubation of the protein in thrombin-agarose resin (Sigma) under gentle shaking for 2 h at room temperature before submitting to size exclusion chromatography with a Hiload 26/600 Superdex 75 (GE Healthcare) column equilibrated with buffer containing 20 mM NaPi (pH 6.7), 50 mM NaCl, 2 mM β-mercaptoethanol and 2 mM DTT. Fractions were eluted by SDS polyacrylamide gel electrophoresis, pooled and concentrated for NMR spectroscopy experiments.

2.2 NMR experiments and assignment of the complete catalytic domain

Samples of ~ 0.2 mM ¹³C/¹⁵N-labeled and partially deuterated Cdc25B in buffer containing 20 mM NaPi (pH 6.7), 50 mM NaCl, 5% ²H₂O, 2 mM β-mercaptoethanol and 2 mM DTT were used in NMR experiments.

Triple-resonance experiments (HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HNCO, HN(CO)CA) using traditional pulse sequences [15] and ¹H–¹⁵N HSQC experiments [16] were used. All spectra were acquired at 298 K on a 800 MHz Bruker Avance™ III spectrometer equipped with a cryogenic probe and processed with NMRPipe package [17]. Spectra visualization and assignment were done in CCPNMR Analysis 2.4 software [18].

2.3 Relaxation rates, heteronuclear NOE and model-free analysis

Spin-lattice relaxation (R_1) and rotating frame relaxation ($R_{1\rho}$) rates of backbone amide resonances from uniformly labeled ¹⁵N Cdc25B were measured by inverse detected 2D NMR experiments [19]. R_1 and $R_{1\rho}$ relaxation rates were calculated using 12 different relaxation delays in the acquisition of 2D spectra. Delay times used for R_1 were: 0.1, 0.2, 0.35, 0.5, 0.6, 0.75, 0.9, 1.2, 1.4, 1.6,

1.8 and 2.0 s. Delay times used for $R_{1\rho}$ were: 0.01, 0.024, 0.04, 0.054 (2 \times), 0.08, 0.096, 0.112, 0.124, 0.14, 0.16 and 0.18 s. Exponential decay of peak intensities were fitted in MATLAB[®]R2009b software. Spectra were acquired with 16 scans with a recycle delay of 3 s.

Spin-spin relaxation rates (R_2) were calculated from $R_{1\rho}$ rates with pulse sequences optimized for weak spin-lock fields (up to 1 kHz) [20]:

$$R_{1\rho} = R_2 \cos^2 \theta + R_1 \sin^2 \theta$$

where $\theta = \tan^{-1}(2\pi\Delta\nu/\gamma_N B_1)$, $\Delta\nu$ is the resonance offset and $\gamma_N B_1$ is the strength of the spin-lock field.

Heteronuclear NOE of ^{15}N spins was obtained by the relation between peak intensity from 2D ^1H - ^{15}N spectra in the presence or in the absence of proton saturation.

Order parameters of ^{15}N spins were calculated with ModelFree [21] and FAST-Modelfree [22] programs. The central structure of the most populated cluster from a MD simulation of the Cdc25B complete catalytic domain was used to obtain the initial parameters of an axially symmetric diffusion tensor using PDBinertia [23] and R2R1_diffusion program [24]: $\tau_c = 12.6$ ns, $D_{\parallel/\perp} = 1.26$, $\theta = 0^\circ$ and $\phi = 0^\circ$.

Parameters were adjusted with 600 Monte Carlo simulations and ^{15}N chemical shift anisotropy and NH bond length values of -160 ppm and 1.02 Å, respectively. Global parameters converged after 5 iterations.

2.4 Residual dipolar couplings

Partial alignment of proteins was achieved in liquid crystalline media prepared with PEG/*n*-hexanol as described previously [25]. Scalar couplings of backbone amides ($^1J_{\text{NH}}$) were measured in isotropic media using anti-phase in-phase experiments [26]. Residual dipolar couplings of backbone amides (D_{NH}) were calculated by subtracting $^1J_{\text{NH}}$ from NH splittings measured in anisotropic media.

A comparison between calculated and measured D_{NH} was performed with the DC program included in NMRPipe package [17]. The crystal structure of the Cdc25B catalytic domain (PDB: 1QB0)[10] was used to calculate D_{NH} . Hydrogens were added to the PDB structure with

pdb2gmx tool, available in GROMACS suite of programs [27], followed by an energy minimization procedure with heavy atoms restrained to their initial coordinates.

2.5 Molecular dynamics simulation

A simulation of Cdc25B catalytic domain was started with coordinates from the crystal structure deposited in PDB (code 1QB0). Additional N-terminal residues (GSHMEFQ) resulting from the adapter sequence used for cloning and the C-terminal residues (552G-566Q) absent in the crystal structure were manually added in an extended conformation. Crystallographic waters, ions and β -mercaptoethanol were removed and hydrogens were added to the structure. A MD simulation in implicit solvent with coordinate restrains for all atoms present in the crystal structure was initially carried out for 12 ns. The final structure was solvated in a dodecahedron water box and NaCl ions were added to neutralize the system up to a 100 mM salt concentration.

The Amber99SB-ILDN force field [28] was used for protein and solvent ions and the TIP3P model [29] was used for water molecules. A 2 μs -long MD simulation was performed at 300 K and 1 bar pressure with a 2 fs integration steps in the NPT ensemble. The PME method [30] with 0.9 nm real space cutoff was used to treat long-range electrostatics, the Berendsen barostat [31] was used with period 0.5 ps and the Bussi thermostat was used with period 0.5 ps [32]. Construction of the models, MD simulations and analysis of results were performed with the GROMACS 4.6.1 suite of programs[27].

3 Results and discussion

3.1 Assignment of the Cdc25B complete catalytic domain

Backbone atoms of the complete catalytic domain of wild-type (WT) Cdc25B (segment 373S-566Q) were assigned by conventional NMR triple resonance experiments. The ^1H - ^{15}N HSQC spectrum (Figure 1A) showed relatively good peak dispersion, except for the region around $\delta^{1\text{H}} = 8.2$ ppm and $\delta^{15\text{N}} = 120$ ppm which suggests the presence of conformational disorder in the polypeptide chain [33]. Out of 187 expected backbone amide

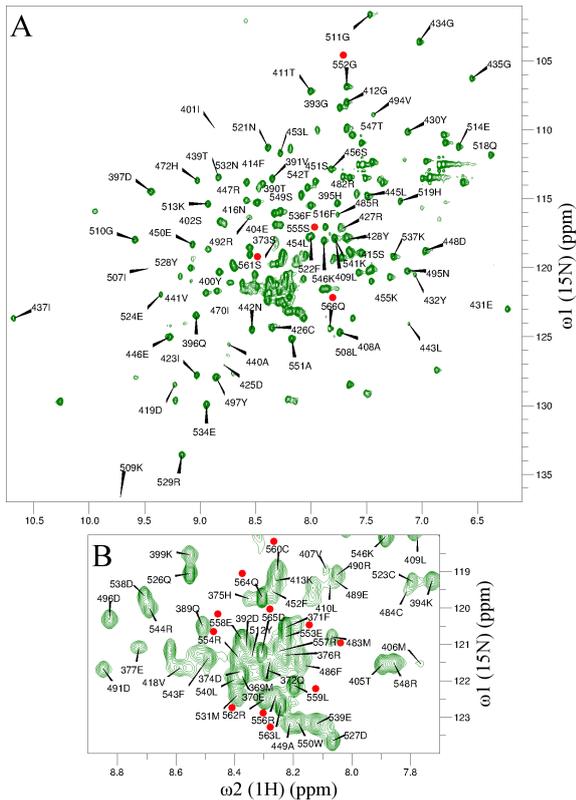


Figure 1: Assignment of backbone amide resonances in the ^1H - ^{15}N HSQC spectrum of the Cdc25B complete catalytic domain. Residues of the C-terminal segment (552G–566Q) are marked with red dots. Panel B is a magnification of the central region in the spectrum.

resonances in the HSQC spectrum, 173 were observed and 143 were assigned ($\sim 82\%$). All backbone amide resonances of the C-terminal segment 552G–566Q absent in the crystal structure and in the Cdc25B construction used by Lund and Cierpicki [13] were located in regions of the HSQC spectrum expected for residues in random coil conformation (red dots in Figure 1)[33].

Assignments were in close agreement with the chemical shifts provided by Lund and Cierpicki, except for $\delta^{15}\text{N}$ of 434G, 511G, 509K and 529R. These differences are probably due to a shorter spectral window used by Lund and Cierpicki for the acquisition of NMR spectra in the ^{15}N dimension [13]. Our results should present the correct $\delta^{15}\text{N}$ for these residues.

Most of the unassigned or unobserved resonances correspond to segments of the protein located in exposed loop regions (378L–388L, 458I–467V and 499S–502Y) where amide protons are probably in fast exchange with water. The active site P-loop (474E–480G) was not assigned due to signal

broadening probably caused by intermediate rate exchange of HOPO_3^{2-} ions with the buffer, as also observed in a low molecular weight protein tyrosine phosphatase [34]. Part of this P-loop (475F–480G) was assigned by Lund and Cierpicki [13], but using a protein construction truncated and with a point mutation at the catalytic cysteine (C473S) that could alter the binding and dynamic properties of the active site.

Residues 373S–377E, 394K, 396Q, 416N, 439T, 446E, 495N–497Y, 510G and 547T–550W were assigned here but not by Lund and Cierpicki. On the other hand, they assigned 387L, 388L, 420K–422V, 433E, 438K, 459A, 468I, 488R, 505M, 506Y and 535A, which was not done here. These differences may be attributed to the partial deuteration of the protein conducted here and to different buffers used in protein preparations for acquisition of NMR spectra.

3.2 Fast dynamics in the Cdc25B terminals

Values of R_1 and R_2 ^{15}N relaxation rates and heteronuclear $\{^1\text{H}\}$ - ^{15}N NOE of the C-terminal 547T–566Q and N-terminal 373S–377E segments shown in Figure 2 are clearly different from the rest of the protein. Higher R_1 and lower R_2 and NOE values found in the Cdc25B terminals in comparison to the rest of the protein chain are related to fast tumbling (lower τ_c)[35], indicating the presence of considerable motion in the ps–ns timescale in these terminal regions.

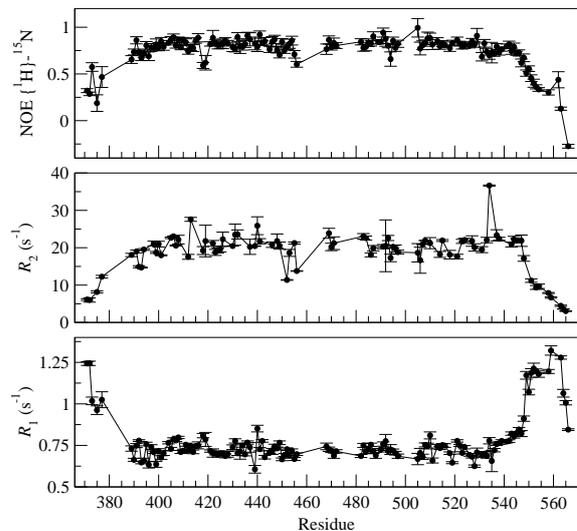


Figure 2: R_1 and R_2 relaxation rates and $\{^1\text{H}\}$ - ^{15}N NOE of backbone ^{15}N spins measured for the Cdc25B complete catalytic domain.

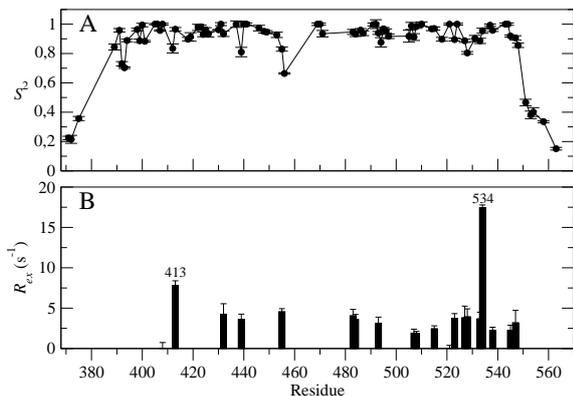


Figure 3: Order parameters (S^2 , panel **A**) and rates of conformational exchange (R_{ex} , panel **B**) obtained after the adjustment of relaxation measurements with Lipari-Szabo formalism.

The amplitude of motions in this timescale was quantified with order parameters (S^2) adjusted from the ^{15}N relaxation and NOE measurements. Lower S^2 values also pointed to increased conformational flexibility of the terminal regions and S^2 values close to unity indicate a relative rigidity of the protein core (Figure 3A). A significant contribution of conformational exchange (R_{ex}) for residues 413K and 534E was determined here, indicating motion in the μs – ms timescale for these two residues (Figure 3B). Lund and Cierpicki have not observed a large R_{ex} contribution for 534E, but their R_2 value for the same residue in the WT Cdc25B is considerably higher than in the C473S mutant, suggesting that this point mutation has altered the conformational flexibility around the active site.

3.3 The last 20 residues in Cdc25B C-terminus are disordered

Experimental residual dipolar couplings (D_{NH}) of the Cdc25B complete catalytic domain measured in neutral PEG/hexanol alignment media agree well with D_{NH} calculated for the residues present in the crystal structure with a $Q=0.26$ quality factor (Figure 4A). Thus, the protein core and the C-terminal α -helix are well folded and populate conformations in solution similar to the crystal structure.

The observed $D_{\text{NH}}^{\text{obs}}$ for residues in the C-terminus (549S–566Q) decay to values closer to 0 Hz when compared to the rest of polypeptide chain (Figure 4B), suggesting again conformational disorder in this region. The same is also observed for the

N-terminus. Given that the Cdc25B C-terminus is enriched with charged residues, D_{NH} measurements in a neutral alignment media should perturb relatively less and more consistently report the conformational distribution of the C-terminus in solution.

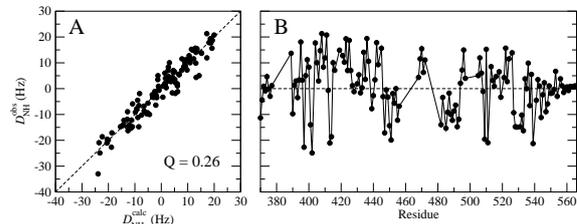


Figure 4: Residual dipolar couplings of backbone amides of the Cdc25B complete catalytic domain. Panel **A** shows a comparison between D_{NH} calculated from crystal structure (PDB: 1QB0) and measured in neutral PEG/*n*-hexanol alignment media. Panel **B** shows the measured D_{NH} for each residue.

The global flexibility of the Cdc25B has been debated in the literature[13]. Previous simulations employing either molecular dynamics[11] with the OPLS-AA force field[36] or a library of rotamers sampled by Monte Carlo[12] suggested that the Cdc25B C-terminal α -helix may (partially) unfold in solution. NMR experiments presented previously for a truncated construction of the Cdc25B catalytic domain[13] and obtained here for the complete catalytic domain show this C-terminal α -helix is folded and has little flexibility. Thus, the force field description used in the previous simulations[11, 12] artificially destabilized helical conformations. This supports the notion that force fields should be carefully tested and further developed, and simulations should be compared and validated with experimental data when available.

In order to probe the conformations visited by the C-terminus and its possible contacts with the protein core, a long MD simulation was carried out with the Amber99SB-ILDN force field which has a balanced helix-coil description and has been successfully used to simulate disordered regions[37, 38].

Figure 5 (insert) shows that the protein core (residues 373–546) and the C-terminal helix (534–546) remained well folded during the whole simulation trajectory, in agreement with the NMR measurements. The root mean-square fluctuation (RMSF) profile shown for the Cdc25B $C\alpha$ is very similar to the fluctuations derived from random

coil index (RCI) analysis of backbone chemical shifts [39]. This is particularly true for the C-terminus and suggests that the simulation correctly samples the amplitude of movements for this protein region.

The lower panel in Figure 5 shows structures obtained from the MD trajectory. The C-terminus fluctuates widely and occupies a large volume close to the active site. Thus, it may directly interact with possible competitive inhibitors that bind to Cdc25B[11]. The C-terminus also makes frequent contacts with the solvent-exposed surface of segments 443–449 and 482–487 in helical regions on the protein core. In particular, the indole group of residue 550W forms a cation- π stacking interaction with the guanidinium group of 482R.

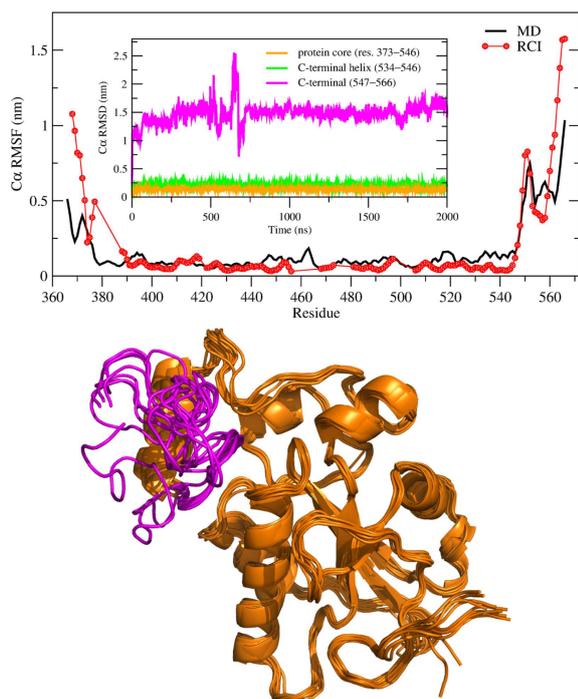


Figure 5: Conformational flexibility of the Cdc25B complete catalytic domain probed by a 2 μ s molecular dynamics simulation. The upper panel shows Cdc25B C α root mean-square fluctuation (RMSF) profiles from MD simulation (black) and derived from RCI analysis (red). The inset shows deviations (RMSD) of Cdc25B C α coordinates of the protein core (orange), C-terminal helix (green) and C-terminal segment (magenta) during the MD trajectory. The lower panel shows a superposition of 10 structures obtained from the MD simulation.

4 Conclusions

The conformational flexibility of the Cdc25B complete catalytic domain was evaluated with experimental NMR measurements and MD simulations. The protein construction used here contains all residues at the C-terminus that were not present in studies previously published[13, 11, 12].

Our results show that the last 20 C-terminal residues (547T–566G) show large fluctuations when compared to the rest of the chain. Conformational disorder for the C-terminus is supported by a range of observations such as assignment of the ^1H – ^{15}N HSQC spectra compatible with random coil conformation, ^{15}N relaxation rates in accord with fast tumbling and derived order parameters, residual dipolar couplings close to zero, and large mean fluctuations found in the long time MD simulations.

The C-terminal helix is well folded during the whole MD trajectory presented here (Figure 5), in disagreement with previous molecular simulations[11, 12]. This is due to the use of the OPLS-AA force field in the previous MD simulation which has been shown to give an unbalanced description of helix-coil conformations[37, 38]. Also, the inclusion of the last 15 residues in the C-terminus might also stabilize the terminal helix in the simulations.

The large fluctuations in position observed here in the MD simulations are validated by the RCI indexes derived from the NMR chemical-shifts. The volume occupied by the terminal residues occludes the entrance of the shallow Cdc25B active site. Frequent contacts between the C-terminus and the protein core were also observed. Thus, the combination of experimental NMR spectroscopy and computer simulation is a powerful method to study with atomic resolution the conformational flexibility of proteins in solution.

References

- [1] Boutros, R.; Dozier, C.; Ducommun, B. *Curr. Opin. Cell Biol.* **2006**, *18*, 185–191.
- [2] Rudolph, J. *Biochemistry* **2007**, *46*, 3595–3604.
- [3] Boutros, R.; Lobjois, V.; Ducommun, B. *Nat. Rev. Cancer* **2007**, *7*, 495–507.

- [4] Lavecchia, A.; Di Giovanni, C.; Novellino, E. *Curr. Med. Chem.* **2009**, *16*, 1831–1849.
- [5] Lazo, J. S.; Wipf, P. *Anti-Cancer Agents in Med. Chem.* **2008**, *8*, 837–842.
- [6] Zhang, Z. Y. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 1–52.
- [7] Arantes, G. M. *Biochem. J.* **2006**, *399*, 343–350.
- [8] Arantes, G. M. **2008**, *112*, 15244–15247.
- [9] Fauman, E. B.; Cogswell, J. P.; Lovejoy, B.; Rocque, W. J.; Holmes, W.; Montana, V. G.; Piwnica-Worms, H.; Rink, M. J.; Saper, M. A. *Cell* **1998**, *93*, 617–625.
- [10] Reynolds, R. A.; Yem, A. W.; Wolfe, C. L.; Deibel, M. R.; Chidester, C. G.; Watenpugh, K. D. *J. Mol. Biol.* **1999**, *293*, 559–68.
- [11] Arantes, G. M. *Proteins* **2010**, *78*, 3017–3032.
- [12] Mamonov, A. B.; Bhatt, D.; Cashman, D. J.; Ding, Y.; Zuckerman, D. M. *J. Phys. Chem. B* **2009**, *113*, 10891–10904.
- [13] Lund, G.; Cierpicki, T. *Proteins* **2014**, *82*, 2889–95.
- [14] Wilborn, M.; Free, S.; Ban, A.; Rudolph, J. *Biochemistry* **2001**, *40*, 14200–14206.
- [15] Sattler, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
- [16] Mori, S.; Abeygunawardana, C.; Johnson, M.; Vanzijl, P. *J. Magn. Reson. Ser. B* **1995**, *108*, 94–98.
- [17] Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
- [18] Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins* **2005**, *59*, 687–96.
- [19] Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972–8979.
- [20] Massi, F.; Johnson, E.; Wang, C.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2004**, *126*, 2247–56.
- [21] Palmer, A. G. ModelFree 4.2. <http://www.palmer.hs.columbia.edu/software/modelfree.html>, 2014.
- [22] Cole, R.; Loria, J. P. *J. Biomol. NMR* **2003**, *26*, 203–213.
- [23] Palmer, A. G. PDBinertia 1.11. <http://www.palmer.hs.columbia.edu/software/pdbinertia.html>, 2014.
- [24] Palmer, A. G. R2R1_diffusion 1.11. http://www.palmer.hs.columbia.edu/software/r2r1_diffusion.html, 2014.
- [25] Rückert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797.
- [26] Cordier, F.; Dingley, A. J.; Grzesiek, S. *J. Biomol. NMR* **1999**, *13*, 175–180.
- [27] Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- [28] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–8.
- [29] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- [30] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- [31] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, a.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- [32] Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*.
- [33] Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J. Biomol. NMR* **1995**, *5*, 67–81.
- [34] Stehle, T.; Sreeramulu, S.; Löhr, F.; Richter, C.; Saxena, K.; Jonker, H. R. A.; Schwalbe, H. *J. Biol. Chem.* **2012**, *287*, 34569–34582.
- [35] Hore, P. J. *Nuclear Magnetic Resonance*; Oxford University Press: New York, NY, EUA, 1995; p 87.
- [36] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [37] Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. **2015**, *11*, 5513–5524.

- [38] Palazzesi, F.; Prakash, M. K.; Bonomi, M.; Barducci, A. **2015**, *11*, 2–7.
- [39] Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970–14971.

ANEXO I – Ligand-receptor affinities
computed by an adapted linear interaction
model for continuum electrostatics and by
protein conformational averaging

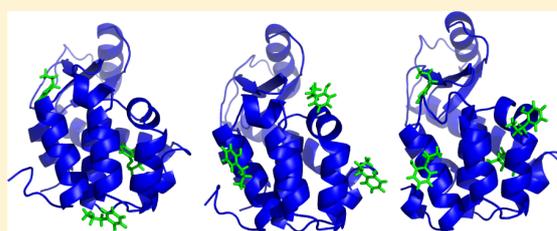
Ligand–Receptor Affinities Computed by an Adapted Linear Interaction Model for Continuum Electrostatics and by Protein Conformational Averaging

Ariane Nunes-Alves and Guilherme Menegon Arantes*

Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900, São Paulo, SP, Brazil

S Supporting Information

ABSTRACT: Accurate calculations of free energies involved in small-molecule binding to a receptor are challenging. Interactions between ligand, receptor, and solvent molecules have to be described precisely, and a large number of conformational microstates has to be sampled, particularly for ligand binding to a flexible protein. Linear interaction energy models are computationally efficient methods that have found considerable success in the prediction of binding free energies. Here, we parametrize a linear interaction model for implicit solvation with coefficients adapted by ligand and binding site relative polarities in order to predict ligand binding free energies. Results obtained for a diverse series of ligands suggest that the model has good predictive power and transferability. We also apply implicit ligand theory and propose approximations to average contributions of multiple ligand–receptor poses built from a protein conformational ensemble and find that exponential averages require proper energy discrimination between plausible binding poses and false-positives (*i.e.*, decoys). The linear interaction model and the averaging procedures presented can be applied independently of each other and of the method used to obtain the receptor structural representation.



1. INTRODUCTION

Prediction of binding affinities between small-molecule ligands and protein receptors has both fundamental and applied importance.¹ In practice, this is a very challenging task² because the ligand functional or bound configurations have a small energy difference from the huge amount of alternative ligand unbound configurations.³ The number and strength of contributions in the ligand bound and unbound states are similar. Consequently, intermolecular interactions have to be evaluated with accuracies much better than 1 kcal mol^{−1} to discriminate the small energy gap between the two states.^{3,4} In addition, a huge number of configurations has to be generated and their energy calculated to sample the important conformational microstates of the molecular system.^{3,5,6} The number of configurations to be sampled will increase if the protein or the ligand has a more flexible structure and if their binding pose is unknown or not unique.^{2,7}

Despite the challenges, there has been enormous progress in the prediction of binding free energies, and several methods have been proposed to tackle the problem.^{1,8,9} In one hand, the application of detailed all-atom force fields, molecular dynamics (MD) simulations (or related approaches), and rigorous free energy estimators^{10–13} have found impressive agreement with experimental affinities;^{14–17} but, given the high computational costs associated, these methods have been successfully applied mainly to less flexible proteins and ligands for which binding sites are known or easy to determine.¹⁸ The high computational costs still prohibit these rigorous methods from being applied

in screenings of large ligand sets. On the other hand, molecular docking^{19–21} employs approximate descriptions of intermolecular interactions usually parametrized against empirical data and efficient conformational search methods to generate binding poses,^{22,23} rank or enrich ligand sets,^{24,25} and determine ligand affinities.^{2,26} However, docking has many documented failures^{27,23,28} which may be due to severe approximations in the calculation of interactions and lack of transferability for ligands or receptors not included in the method parametrization as well as to insufficient conformational sampling.

Another family of methods shows accuracy and computational ease in between the two approaches just mentioned. They are called linear interaction energy (LIE) models^{29–32} because a linear response of the intermolecular interactions³³ is assumed in the estimation of binding free energies by the equation

$$\Delta G_{LIE} = \alpha \Delta \langle V_{vdW}^{l-e} \rangle + \beta \Delta \langle V_{elet}^{l-e} \rangle + \gamma \quad (1)$$

where a force field description of intermolecular van der Waals (*vdW*) and electrostatic (*elet*) interactions between ligand and its environment (V^{l-e}) is employed. The difference (Δ) of ensemble averaged ($\langle \dots \rangle$) interactions between the ligand free state (when environment is the solvent only) and bound state (when environment is the solvated protein complex) is

Received: May 19, 2014

Published: July 30, 2014

multiplied by coefficients derived from the linear response assumption (β) or fit to empirical data (α and γ).^{32,34}

LIE models have been applied successfully to predict affinities for a range of ligand–receptor complexes.^{32,35–38} However, in many of these applications, the LIE models were specifically parametrized to the system studied. In order to increase the model transferability, Hansson et al. proposed the adaptation of coefficients to ligand properties (e.g., the number of possible hydrogen bonds).³⁹ Recently, Linder et al. suggested an adaptive LIE model where coefficients in eq 1 are adjusted by the relative polarities of the ligand and of the binding cavity achieving accuracy and model transferability.⁴⁰

To increase computational efficiency and to avoid the sometimes slow convergence of explicit solvent contributions^{41,42} in eq 1, continuum electrostatics descriptions of solvation^{43–46} have been used in LIE models.^{36,41,47–49} Here, we propose and describe the necessary parametrization of LIE models that combine an implicit solvent description with adaptive coefficients⁴⁰ to predict binding affinities. Local configurational sampling of ligand–receptor complexes usually done by molecular dynamics simulations is substituted by more economic molecular docking and geometry optimizations.^{21,36,47}

The methods mentioned so far rely their predictions on one initial receptor structure, typically obtained from X-ray crystallography. During conformational search in molecular docking, the receptor structure is maintained rigid, maybe allowing for side-chain rotations or smoothed interactions.^{50–52} In methods applying ensemble averages, protein configurations near the initial structure are visited in relatively short MD simulations; but, for flexible receptors, sufficient sampling of protein motions will be difficult to achieve in both approaches. A possible solution in those cases is to start the search or averaging from a conformational ensemble, i.e., from multiple representations of the receptor structure.^{6,7,53,54}

Several approaches, mostly related to docking, are now used to predict binding poses and affinities from receptor conformational ensembles.^{22,55–59} Usually a dominant pose and dominant state approximation is applied.^{57–59} This means that the binding free energy or the related docking score for a given ligand–receptor pair is estimated from the most favorable pose (only one) found after evaluating several complexes obtained from the different receptor structures in the ensemble. This approximation should be appropriate for the level of accuracy expected in docking, but it dismisses important contributions such as multiple binding poses, receptor reorganization energy and thermal fluctuations, and the related entropic effects. Thus, it may be useful to average contributions obtained from an ensemble of ligand–receptor poses.^{60,61}

Based on the implicit ligand theory recently developed,⁶¹ two ensemble averages can be defined for the calculation of binding free energy between a ligand and a receptor represented by a conformational ensemble embedded in an implicit solvent. The first average concerns with the ligand configurational distribution that may be obtained for interaction with one given *rigid* receptor structure. It has been called the binding potential of mean-force, B , and may be estimated by an exponential mean

$$B = -k_B T \ln \frac{1}{P} \sum_{i=1}^P \exp\left(-\frac{\Psi_i}{k_B T}\right) \quad (2)$$

where k_B is the Boltzmann constant, T is the temperature, P is the number of ligand configurations sampled, and Ψ_i is the *implicit solvent-mediated* interaction energy for the i th ligand pose.⁶¹

The second ensemble average accounts for the receptor configurational distribution. Similarly, it may be estimated by an exponential mean, leading to an expression for the binding free energy

$$\Delta G_{ave} = -k_B T \ln \frac{1}{N} \sum_{n=1}^N \exp\left(-\frac{B_n}{k_B T}\right) + \Delta G_\xi \quad (3)$$

where B_n is the binding potential of mean-force (eq 2) for the n th receptor configuration out of a total of N configurations sampled.⁶¹ ΔG_ξ represents a correction to standard concentration due to restriction of the volume sampled by the ligand.

Here, we propose approximations to these two exponential averages in order to account for multiple binding poses and for protein conformation flexibility. In the next section, we provide details of the training and test sets and the methodology used to calibrate adapted LIE models for continuum solvation. System setup, generation of receptor–ligand poses, and definition of the force fields employed are also described. Both Results and Discussion are divided in two parts. First, we report the construction and performance of the adapted LIE models and the procedures for averaging contributions from an ensemble of ligand–receptor poses. Then, we analyze the accuracy and shortcomings of the proposed LIE model as well as the applicability of the ligand–receptor conformational averaging.

2. COMPUTATIONAL METHODS

Calibration and tests of the proposed approximations were conducted using bacteriophage T4 lysozyme mutants L99A⁶² and L99A/M102Q,^{17,63} HIV-1 reverse transcriptase (HIVRT), and human FK506 binding protein 12 (FKBP) as model systems (Table 1). These proteins were chosen based on the

Table 1. Proteins Included in the Training and Test Sets and Ranges of Binding Affinities of Associated Ligands (in kcal mol⁻¹)

protein ^a	PDB code	ligands	$-\Delta G_{exp}$
L99A	3DMV	benzene derivatives	4.5–6.7
M102Q	1LL3	benzene derivatives	4.3–5.8
HIVRT	1RT1	HIV1–HIV6	4.9–11.8
FKBP	1FKG	FKB1–FKB5	7.8–11.2

^aT4 lysozyme mutants L99A and M102Q, HIV-1 reverse transcriptase (HIVRT), and human FK506 binding protein 12 (FKBP).

availability of experimental structures and binding affinities. The ligand set varies in size from fragment-like small molecules which bind T4 lysozyme to lead-like molecules which bind FKBP. Only neutral ligands were considered (Table S3 and Figure S1 in the Supporting Information, SI).

2.1. Parametrization of the Model. Following Linder et al.,⁴⁰ system-derived descriptors are used here to scale coefficients in the linear interaction models. Ligand (π) and cavity or binding site (η) relative polarities were given by the ratio PSA/SA , where SA represents the ligand or cavity total surface area and PSA represents the area of its subset of polar atoms (Table S3 and Table S4). Ligand surface area was obtained from the “3V” server,⁶⁴ and ligand polar surface area

was calculated using the approach of Ertl et al.⁶⁵ Cavity area was obtained from the SA of residues in contact with (or less than 4 Å from) the ligand. Thus, each binding pose has a characteristic η . Protein carbonyl C, O, N, and H bound to O and N were assigned as polar atoms.⁶⁶

Coefficients in the LIE equations were obtained by an optimization procedure that minimized deviations between calculated and experimental affinities for a training set of 10 T4 lysozyme ligands, 3 HIVRT ligands, and 10 T4 lysozyme false-positive poses, as indicated in Table S5. The test set used to check the performance of the parametrized equations was composed of a different set of 15 T4 lysozyme binders, 10 T4 lysozyme false-positive poses, 9 T4 lysozyme nonbinders, 3 HIVRT ligands, and 5 FKBP ligands, as indicated in Table 2. Optimization was carried out with a combination of genetic (GA)⁶⁷ and simplex⁶⁸ algorithms as previously described.⁶⁹ A population of 10 individuals with each coefficient represented by 12 bits was used in the GA. Coefficients in eq 6 could vary between [0,10] for k_s , [-20,20] for k_o , and [-10,10] for the others k_i (eqs 6 and S1–S3). Populations evolved for 10⁶ generations. Simplex optimization was carried from the best GA individuals until the difference in deviations between successive generations was smaller than 10⁻⁶ kcal mol⁻¹.

2.2. Construction of Receptor–Ligand Structures.

Protein structures retrieved from the Protein Databank (PDB) were used after removal of water and other crystallization molecules. Incomplete side chains were built with the WHATIF server.⁷⁰ Hydrogens were constructed using the GROMACS PDB parser⁷¹ for proteins and Babel 2.2⁷² for ligands. Ligand geometry was optimized using Gaussian⁷³ with the AM1⁷⁴ potential if holo crystal structures were unavailable.

To train and test the LIE models, holo structures were taken from the PDB when available. Otherwise, the most favorable binding pose obtained from docking the ligand to an apo structure was used (Table S3). These poses were compared to known crystal structures of congeneric ligands bound to the same protein to confirm the docked ligand was complexed in a plausible binding mode.

Unstable or artificial poses of known binders may be generated in docking due to inaccuracies in the scoring functions.^{27,28,75} Such artificial poses, here called false-positives, were used as a decoy set to assist in the parametrization of the LIE equations. Assuming the ligand will occupy a site different from the known crystallographic site, false-positive poses were obtained by docking ligands to apo crystal structures using a grid excluding the known binding site. Selected poses were submitted to energy minimization, careful heating up ramps, and 10–20 ns explicit solvent molecular dynamics simulations as described below. False-positive poses were retained only if the ligand spent more than 20% of the trajectory dissociated from the protein. Ligand-protein dissociation was monitored by the ligand solvent accessible surface area (SASA).

Tentative configurational ensembles were generated for apo T4 lysozyme L99A and M102Q mutants, HIVRT bound to ligand HIV1 and FKBP bound to ligand FKBP1 (Table 1). Ensembles generated from apo HIVRT and FKBP could not be used for docking due to large conformational changes which occluded the binding sites (see Discussion for further details). Receptor structures were submitted to energy minimization, and implicit solvent molecular dynamics simulations were run for 160–235 ns. For each protein, an ensemble was constructed by 50 configurations (excluding the ligand in the case of HIVRT and FKBP) collected along trajectories at regular time

Table 2. Binding Free Energies (in kcal mol⁻¹) Experimentally Measured and Estimated by Eq 6 for the Ligand Test Set

ligand	ΔG_{exp}^a	ΔG_{ALICE}
L99A		
<i>n</i> -butylbenzene ^b	-6.7	-6.3
propylbenzene	-6.5	-5.7
ethylbenzene ^b	-5.7	-5.1
toluene	-5.5	-3.7
benzene ^b	-5.2	-3.0
3-ethyltoluene	-5.1	-5.5
meta-xylene	-4.7	-4.5
2-ethyltoluene	-4.5	-4.8
<u>propylbenzene</u>	>-2.0	-1.4
<u>ortho-xylene (A)</u>	>-2.0	0.7
<u>toluene</u>	>-2.0	-1.7
<u>4-ethyltoluene</u>	>-2.0	-2.1
<u>benzene</u>	>-2.0	-0.8
3-methylpyrrole	>-2.0	-2.6
phenol	>-2.0	-3.1
1,3,5-trimethylbenzene	>-2.0	-4.4
cyclohexane	>-2.0	-2.6
2-fluoroaniline	>-2.0	-2.8
M102Q		
(phenylamino)acetonitrile ^b	-5.8	-4.8
toluene	-5.2	-4.1
3-methylpyrrole	-5.2	-2.9
thieno[3,2-b]thiophene ^b	-4.9	-3.5
2-ethylphenol ^b	-4.8	-5.0
catechol ^b	-4.4	-2.5
2-ethoxyphenol ^b	-4.3	-4.9
<u>thieno[3,2-b]thiophene</u>	>-2.0	-0.3
<u>(phenylamino)acetonitrile</u>	>-2.0	-0.4
<u>catechol</u>	>-2.0	-2.4
<u>2-propylphenol (A)</u>	>-2.0	-1.7
<u>2-ethoxyphenol</u>	>-2.0	-0.7
phenylhydrazine	>-2.0	-3.0
2-methoxyphenol ^b	>-2.0	-4.5
4-vinylpyridine	>-2.0	-4.2
<i>N</i> -(<i>o</i> -tolyl)cyanofornamide	>-2.0	-5.0
HIVRT		
HIV3	-8.1	-10.5
HIV4	-10.6	-9.5
HIV5	-6.4	-7.4
FKBP		
FKBP1 ^b	-11.0	-10.9
FKBP2 ^b	-11.2	-11.4
FKBP3	-7.8	-7.2
FKBP4	-8.5	-7.8
FKBP5	-9.6	-9.9

^aRepeated from Table S3. ^bHolo structure taken from the PDB. False-positive poses of binder molecules are underlined. These poses and nonbinders were assumed to have $\Delta G_{exp} > -2.0$ kcal mol⁻¹. The label (A) represents different false-positive poses of the same ligand.

intervals (3–4 ns) after stabilization of C_α root mean-squared deviation (RMSD). For each configuration in an ensemble, 20 docking poses were generated resulting in a total of 1000 ligand-bound structures for each protein–ligand pair.

Dockings to crystal structures were performed with AutoDock 4.0⁵⁰ with its genetic algorithm search run with 150 individuals for 27,000 generations maximum. Dockings to

the configurational ensemble were done with AutoDock Vina²¹ setting the exhaustiveness level to 8. Conformational search options were chosen in order to thoroughly search for the possible docking poses in a given protein structure. Grids with 0.375 Å spacing and 60 to 80 points were centered in the known binding sites. Protein structures were kept frozen, but bond torsions were allowed in ligands. Typically, T4 lysozyme ligands had 0–4 torsions activated, HIVRT ligands had 3–12 torsions, and FKBP ligands had 6–13 torsions (Table S2). The correction of the restricted volume sampled by the ligand to the standard concentration (1 M)⁶⁰ in eq 3 was calculated from the average volume of the grid used for docking, $2.7 \times 10^4 \text{ \AA}^3$.

2.3. Protein Force Field and Simulation Details. Energy contributions for the linear interaction models (eqs 6 and S1–S3) were obtained after geometry optimizations of protein–ligand complexes in implicit solvent using the conjugate gradient approach (T4 lysozyme) or the BFGS algorithm (HIVRT and FKBP).⁶⁸ Free protein and ligand contributions were obtained without the ligand or protein, respectively, but using the same geometry of the complex. The GBR⁶ method⁴³ was used to calculate the solvent polarization free energies, G_{GB} in eq 6.

GROMACS 4.5⁷¹ was used for all protein geometry optimizations and MDs. Dynamics were carried out at 300 K with a 2 fs time-step, and covalent bonds were constrained with LINCS.⁷⁶ Proteins were represented by the OPLS-AA force field.⁷⁷

In explicit solvent simulations, structures were solvated in a dodecahedral box with edges at least 8 Å far from the protein. The SPC/E potential⁷⁸ was employed for water, and chloride ions were added to neutralize the charge of the systems. Periodic boundary conditions were activated. The velocity rescale method⁷⁹ was used to control the temperature at 300 K, and pressure control at 1 bar was applied with the Parrinello–Rahman method.⁸⁰ PME⁸¹ was used to treat long-range electrostatics, and a switched potential (cutoffs 0.8, 1.2 nm) was used to treat van der Waals interactions. Before production MD, systems were heated in cycles of short 20 ps simulations with gradual temperature increase (10 K, 50 K, 100 K, 200 K, and 300 K) and reduction of position restraints over heavy atoms (240 kcal nm⁻², 120 kcal nm⁻², 24 kcal nm⁻², 2 kcal nm⁻², and 0).

In implicit solvent simulations, the generalized Born (GB) approximation was used.⁴⁴ The OBC model was used to estimate Born radii,⁴⁵ and the nonpolar contribution was calculated as in Schaefer et al.⁴⁶ with a surface tension of 5.4 cal mol⁻¹ Å⁻² for all atoms. MDs were run with a leapfrog stochastic dynamics integrator, with a friction coefficient $\tau = 10 \text{ ps}^{-1}$.

2.4. Ligand Force Field. Topologies for ligands were built manually based on the OPLS-AA force field. Bonding, Lennard-Jones, and implicit solvation parameters unavailable for certain atom types in OPLS-AA were approximated from similar chemical functions. Parameters for dihedral angles of the thymine ring in HIVRT ligands were taken from the AMBER99 force field^{82,83} OPLS-AA partial charges were used for nonpolar ligands or for ligands with one polar group. For ligands with more than one polar group, partial charges were recalculated with AM1-CM2.^{74,84} For small ligands, partial charges for the whole molecule were recalculated. For the bulky FKBP and HIVRT ligands, the molecule was divided in fragments, and those with more than one polar group had their partial charges

recalculated. For HIVRT ligands containing sulfur, the partial charges were recalculated with HF/6-31G*.

For all ligands, the partial charges used here resulted in total and component dipole moments in good agreement with a quantum mechanical (QM) reference (HF/6-31G*, Table S1). For instance, OPLS-AA partial charges were used for 4-vinylpyridine resulting in a total dipole moment $\mu = 2.7 \text{ D}$ which is in good agreement with the QM reference $\mu = 2.6 \text{ D}$. Another example is 2-fluoroaniline which has two polar groups. Its partial charges were recalculated as described above and resulted in $\mu = 2.0 \text{ D}$ which is in good agreement with the QM reference, $\mu = 1.9 \text{ D}$.

All ligand topologies are available online⁸⁵ or from the authors upon request.

2.5. Approximations to Implicit Ligand Theory. Given that only configurations with favorable interaction energies will contribute significantly to the exponential average in eq 2, here we use ligand docking to quickly generate ligand–receptor poses with favorable interactions for a rigid receptor conformation and approximate $\Psi \approx \Delta G_{int}$. Thus, eq 2 leads to

$$B_E = -k_B T \ln \frac{1}{P} \sum_{i=1}^P \exp\left(-\frac{\Delta G_{int,i}}{k_B T}\right) \quad (4)$$

where ΔG_{int} is an intrinsic binding free energy used to estimate the stability of a given ligand binding pose (see Results, section 3.1). As docking does not generate an equilibrium distribution of ligand–receptor configurations, application of eq 4 is approximate. Substitution of interaction energies (Ψ in eq 2) for an intrinsic binding free energy parametrized against experimental data may partially correct inaccuracies in the docking energy function and introduce entropic contributions.

In the limit that one individual sample dominates the exponential average in eq 4, the dominant pose approximation may be used

$$B_D = \min_i(\Delta G_{int,i}) \quad (5)$$

where only a single intrinsic free energy of binding contribution is used for each rigid receptor structure.

For the receptor ensemble average, eq 3 is used with B_i calculated by either eqs 4 or 5. A dominant state approximation may also be invoked where only a single receptor configuration [$\min_i(B_i)$] is used.⁶¹

Values of $N = 50$ and $\Delta G_\varepsilon = 0.9 \text{ kcal mol}^{-1}$ were employed here (see section 2.2). A maximum of $P = 20$ complex configurations were drawn from docking a ligand to each rigid receptor configuration. Thus, a maximum of 1000 binding poses were used in eq 3. For the calculation of B_E in eq 4, poses with intrinsic free energies less favorable by 2.0 kcal mol⁻¹ than the most stable pose were discarded, effectively leading to $1 \leq P \leq 20$ (see Table S7).

3. RESULTS

3.1. Parametrization and Performance of LIE Models.

The first goal here was to obtain an accurate yet computationally efficient free energy function to estimate the stability of a given binding pose obtained for a small-molecule and a given receptor configuration. This function was called an intrinsic binding free energy (ΔG_{int}).

Several equations based on LIE models previously proposed for implicit^{41,47,49} and explicit^{29,32,40,86} solvents were tested. A combination of implicit solvent and geometry optimization of

ligand–receptor complexes proved reasonably accurate and computationally fast. As indicated below by comparisons of errors observed between LIE models parametrized here and previously available, the following adapted linear interaction model for continuum electrostatics (ALICE) gave the best results

$$\Delta G_{\text{int}} \approx \Delta G_{\text{ALICE}} = k_1(2 - \eta - \pi)V_{\text{vdW}}^c + k_2(\eta + \pi)V_{\text{elet}}^c + k_3\eta(G_{\text{GB}}^c - G_{\text{GB}}^p) + k_4\pi G_{\text{GB}}^l + k_5\Delta\text{SASA}^l + k_6 \quad (6)$$

where solvent polarization free energy (G_{GB}), van der Waals (V_{vdW}), and electrostatic (V_{elet}) potentials were calculated for optimized geometries of the complex (c), protein (p), and ligand (l) species. ΔSASA^l is the difference in SASA between bound and free ligand. The processes of ligand insertion in solution and in the receptor cavity were assumed to be fully decoupled so that the 6 LIE coefficients were independent.

Energy contributions and relative polarity descriptors for ligand (η) and receptor cavity (π) are given in the Supporting Information for all ligands (Table S3 and Table S4). Parameters obtained after optimization of eq 6 against the training set described above were $k_1 = 0.09$, $k_2 = 0.31$, $k_3 = 1.16$, $k_4 = -2.85$, $k_5 = 0.017 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and $k_6 = 3.36 \text{ kcal mol}^{-1}$ (Table S6).

Binding free energies estimated with eq 6 are shown in Figure 1, Table 2, and Table S5. For the 42 small-molecule

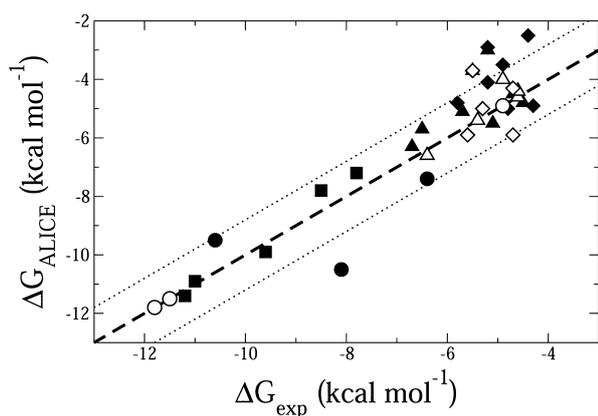


Figure 1. Binding free energies estimated by eq 6. T4 lysozyme L99A ligands are shown as triangles (▲), M102Q ligands are lozenges (◆), HIVRT ligands are circles (●), and FKBP ligands are squares (■). Ligands in the training set are shown as empty symbols, and ligands in the test set are shown as filled symbols. Dashed and dotted lines indicate $y = x \pm 1.2 \text{ kcal mol}^{-1}$.

complexes in the test set, the RMSD between calculated (ΔG_{ALICE}) and experimental affinities is $1.2 \text{ kcal mol}^{-1}$, the coefficient of determination (R^2)⁸⁷ is 0.8, and the maximum error (E_{max}) is $3.0 \text{ kcal mol}^{-1}$ (Table S6), which drops to $2.4 \text{ kcal mol}^{-1}$ if only binder molecules are considered. All false-positive poses were properly recognized, but many nonbinders were not.

In order to compare the results of eq 6 with the Vina docking energy function,²¹ ligands in our test set were docked with Vina to their respective protein crystal structure (native docking) or to a congeneric holo structure if a native one was not available. The binding pose given by the most favorable score in Vina was chosen for comparisons. Error analysis shows that binding free energies calculated with eq 6 resulted in smaller deviations from

experiment than those estimated with Vina (RMSD = 1.1 or 1.7 kcal mol^{-1} , $E_{\text{max}} = 2.2$ or $4.3 \text{ kcal mol}^{-1}$, and $R^2 = 0.7$ or 0.4 , respectively). In particular, the performance of our ALICE model is significantly better than Vina if only FKBP ligands are considered (RMSD = 0.6 or 2.6 kcal mol^{-1} , $E_{\text{max}} = 1.0$ or $4.3 \text{ kcal mol}^{-1}$, and $R^2 = 0.8$ or -2.8 , respectively).

It is instructive to describe some of the other LIE models parametrized and tested here. An implicit solvent LIE equation equivalent to the formulation given by Su et al.⁴⁹ but with the cavity and van der Waals free energies of solvation condensed to one nonpolar contribution⁴⁶ (eq S1 in the SI) resulted in a RMSD of $2.2 \text{ kcal mol}^{-1}$ for the test set ($E_{\text{max}} = 5.6 \text{ kcal mol}^{-1}$). Affinities estimated with an adaptive version of the same model (eq S2) resulted in a RMSD of $1.6 \text{ kcal mol}^{-1}$ ($E_{\text{max}} = 4.6 \text{ kcal mol}^{-1}$). An adapted LIE eq (eq S3) in which k_6 was scaled by $(1-\eta)$ presented a RMSD of $1.4 \text{ kcal mol}^{-1}$ ($E_{\text{max}} = 2.8 \text{ kcal mol}^{-1}$). The number of outliers found for predictions with this last model was, however, 50% larger than found with eq 6 (Table S6).

Adapted LIE models with the same energetic contributions but with different combinations of the polarity descriptors were also tested, but eq 6 is the most accurate model. A similar result was observed by Linder et al.⁴⁰ for adapted LIE models in explicit solvent.

3.2. Averaging Multiple Ligand and Receptor Configurations. The second goal of this study was to test procedures and approximations based on the implicit ligand theory⁶¹ to average the intrinsic free energies calculated for an ensemble of ligand–receptor complexes. Three combinations of ligand pose and receptor configuration averages were tested: In ΔG_{EE} , eq 4 is used to average the ligand poses and to calculate the binding potential of mean-force for each receptor structure, and eq 3 is used to average the receptor configurational distribution. In ΔG_{DE} , eq 5 is used to calculate the binding potential of mean-force, and eq 3 is used to average the receptor distribution. Finally in ΔG_{DD} , eq 5 is again used to calculate the binding potential of mean-force, and a dominant state approximation is used for the receptor distribution (see section 2.5).

Table 3 shows results obtained by the averaging procedures for the full ligand set (previously divided in training and test sets). Error analysis in comparison to experimental affinities is shown in Table 4.

The highest deviations observed in Table 3 are due to the L99A ligands benzene, toluene, and 1,3,5-trimethylbenzene and to the M102Q ligands 2-fluoroaniline, toluene, 3-methylpyrrole, thieno[3,2-b]thiophene, and *N*-(*o*-tolyl)cyanoformamide. All of these also show high ΔG_{ALICE} deviations. In order to isolate contributions of the averaging procedures from inaccuracies in the intrinsic free energy function, all ALICE outliers, i.e., the ligands cited above and catechol, 2-methoxyphenol, 4-vinylpyridine, and HIV3, were removed from the error analysis.

Deviations calculated for this ligand set show slightly smaller RMSDs and determination coefficients closer to one when going from the exponential averages (ΔG_{EE}) to the dominant pose (ΔG_{DE}) and state (ΔG_{DD}) approximations. However, the maximum errors (E_{max}) are higher for ΔG_{DD} due to over-stabilization of HIVRT and FKBP ligands.

It is useful to analyze errors for each receptor separately. For T4 lysozyme mutants, the dominant pose and state approximation results in smaller deviations than the exponential averaging procedures. In fact, ΔG_{DD} shows a RMSD smaller than that observed for ΔG_{ALICE} for L99A ligands (Table 2 and Table S6) suggesting that receptor conformational selection

Table 3. Binding Free Energies (in kcal mol⁻¹) Experimentally Measured and Estimated by the Averaging Procedures Described in the Text for the Full Ligand Set

ligand	ΔG_{exp}^a	ΔG_{EE}	ΔG_{DE}	ΔG_{DD}
L99A				
isobutylbenzene	-6.4	-4.1	-4.4	-5.6
4-ethyltoluene	-5.4	-3.7	-4.1	-4.7
para-xylene	-4.6	-3.1	-3.5	-4.1
indole	-4.9	-2.7	-3.2	-3.8
ortho-xylene	-4.6	-3.1	-3.6	-4.2
n-butylbenzene	-6.7	-4.4	-4.7	-5.7
propylbenzene	-6.5	-3.8	-4.2	-5.1
ethylbenzene	-5.7	-3.1	-3.7	-4.1
toluene	-5.5	-2.5	-3.1	-3.5
benzene	-5.2	-1.7	-2.3	-3.0
3-ethyltoluene	-5.1	-3.7	-4.1	-4.8
meta-xylene	-4.7	-3.1	-3.6	-4.0
2-ethyltoluene	-4.5	-3.7	-4.1	-4.8
3-methylpyrrole	>-2.0	-1.3	-2.0	-2.3
phenol	>-2.0	-1.5	-2.1	-2.6
1,3,5-trimethylbenzene	>-2.0	-3.6	-3.9	-4.9
cyclohexane	>-2.0	-1.9	-2.4	-2.7
2-fluoroaniline	>-2.0	-1.7	-2.3	-2.8
M102Q				
2-fluoroaniline	-5.5	-1.5	-2.1	-2.8
5-chloro-2-methylphenol	-5.3	-2.6	-3.2	-4.0
benzyl acetate	-4.7	-4.4	-4.5	-5.6
ortho-cresol	-4.7	-2.1	-2.6	-3.2
2-propylphenol	-5.6	-3.5	-3.8	-4.7
(phenylamino)acetonitrile	-5.8	-3.2	-3.4	-4.4
toluene	-5.2	-2.3	-2.6	-3.1
3-methylpyrrole	-5.2	-0.9	-1.6	-2.0
thieno[3,2-b]thiophene	-4.9	-1.7	-1.9	-2.7
2-ethylphenol	-4.8	-2.7	-3.2	-3.9
catechol	-4.4	-2.8	-3.0	-4.5
2-ethoxyphenol	-4.3	-2.6	-3.1	-3.9
phenylhydrazine	>-2.0	-1.6	-2.3	-3.2
2-methoxyphenol	>-2.0	-2.1	-2.7	-3.3
4-vinylpyridine	>-2.0	-2.3	-2.7	-3.2
N-(o-tolyl)cyanoformamide	>-2.0	-1.8	-2.4	-3.6
HIVRT				
HIV1	-11.5	-10.4	-11.1	-12.2
HIV2	-4.9	-4.6	-5.2	-6.3
HIV3	-8.1	-8.6	-9.2	-10.0
HIV4	-10.6	-8.7	-9.2	-10.6
HIV5	-6.4	-7.5	-8.2	-9.8
HIV6	-11.8	-10.3	-10.5	-12.0
FKBP				
FKB1	-11.0	-11.9	-12.6	-13.8
FKB2	-11.2	-11.7	-12.2	-13.3
FKB3	-7.8	-7.4	-8.2	-9.3
FKB4	-8.5	-7.6	-8.2	-9.4
FKB5	-9.6	-10.7	-11.2	-12.8

^aRepeated from Table S3. Nonbinders were assumed to have $\Delta G_{exp} > -2.0$ kcal mol⁻¹.

contributes to the calculation of binding free energies even for the small and hydrophobic L99A ligands and for the relatively rigid T4 lysozyme engineered cavity. Results for the M102Q mutant show higher deviations which in part may be due to the higher inaccuracies in the ALICE model for this receptor (Table 2).

Table 4. Error Analysis of the Binding Free Energies Calculated by the Averaging Procedures Proposed for Different Ligand Sets

	ΔG_{EE}	ΔG_{DE}	ΔG_{DD}
Full Ligand Set (Table 3)^a			
RMSD	1.7	1.5	1.3
E_{max}	2.9	2.6	3.4
R^2	0.6	0.7	0.8
L99A Ligands^a			
RMSD	1.7	1.4	0.7
E_{max}	2.7	2.3	1.5
R^2	0.0	0.3	0.8
M102Q Ligands^a			
RMSD	2.2	1.8	1.3
E_{max}	2.9	2.6	2.1
R^2	-3.0	-1.8	-0.4
HIVRT Ligands			
RMSD	1.2	1.2	1.7
E_{max}	1.9	1.8	3.4
R^2	0.8	0.8	0.6
FKBP Ligands			
RMSD	0.8	1.1	2.3
E_{max}	1.1	1.7	3.2
R^2	0.6	0.3	-1.9

^aLigands with a ΔG_{ALICE} deviation from the experimental affinity higher than one RMSD (1.2 kcal mol⁻¹, Table 2) were removed from the error analysis. Deviations (in kcal mol⁻¹, except for R^2) were calculated for each set in comparison to experimental affinities.

For HIVRT and specially for FKBP ligands, the opposite trend is observed. The best predictions were obtained for exponential averaging of both pose and receptor configurations (ΔG_{EE}). The dominant approximations overstabilize the binding free energies because some binding poses are as much as 3.5 kcal mol⁻¹ more stable than the experimental free energy.

The relative rankings of binding affinities among the HIVRT and the FKBP ligands are recovered with all averaging procedures, except for the two ligands with the most favorable affinities in each receptor. However, experimental and calculated differences between these two ligands are smaller than 0.3 kcal mol⁻¹.

If the Vina docking energy function is used to approximate Ψ in eq 2, the free energies predicted show significant disagreement with experiment for all ligand sets. For example, ΔG_{EE}^{Vina} calculated for the FKBP ligands give RMSD = 3.8 kcal mol⁻¹, $E_{max} = 4.6$ kcal mol⁻¹, and $R^2 = -7.0$ suggesting that the docking energy function will give meaningless results if used to approximate the solvent-mediated interaction energy in eq 2. This result can be traced to the Vina inability to discriminate false-positive poses. Almost all poses generated from docking were on average used to calculate B_E^{Vina} . On the other hand, less than half of the generated poses were on average used to calculate B_E^{ALICE} for the same set of ligand-receptor poses (Figure S2 and Table S7).

4. DISCUSSION

4.1. ALICE Model Contributions, Performance, and Limitations.

Several LIE equations with different definitions of the nonpolar solvation contribution and with different combinations of polarity descriptors were described and tested here. The best predictions for a test set composed of 42 small-

molecule complexes of 4 different receptors were obtained with eq 6 with 6 adjustable parameters. None of the ligands in the test set were used in the parametrization training set. Deviations observed with this ALICE model and with previously proposed LIE models are similar. For instance, the implicit solvent LIE model proposed by Su et al. has 4 adjustable parameters and resulted in a RMSD of 1.3 kcal mol⁻¹ with $R^2 = 0.62$ for a set of 57 HIVRT ligands (including the 6 HIVRT ligands used here).⁴⁹ The implicit solvent LIE model with 2 adjustable parameters proposed by Kolb et al. showed a RMSD of 1.6 kcal mol⁻¹ with a correlation coefficient of 0.52 for a set of 128 EGFR kinase ligands.³⁶ The LIE model proposed by Wall et al. has 3 adjustable parameters and showed a RMSD of 1.6 kcal mol⁻¹ and a correlation coefficient of 0.62 for a set of 15 neuraminidase inhibitors.³⁵ Finally, the adapted LIE model for explicit solvent proposed by Linder et al. has 3 adjustable parameters and resulted in a mean absolute deviation (similar to a RMSD) of 1.6 kcal mol⁻¹, $E_{max} = 3.4$ kcal mol⁻¹, and $R^2 = 0.72$ for a diverse set of 38 ligands and their respective 16 receptors.⁴⁰ Deviations reported for these four LIE models were obtained with the same ligands (or with a congeneric set of ligands) used for training the models. Still, the ALICE model proposed has the smallest RMSD and the determination coefficient closest to 1.

In order to analyze the energy contributions included in the ALICE model, it should be noted that implicit solvation is not pairwise decomposable in general. Consequently, the splitting of solute–solvent interactions necessary for LIE calculations is not unique. The solvent polarization energy in the bound ligand state is given by $(G_{GB}^c - G_{GB}^p)$ where the initial state, c , indicates a complex with ligand charges turned off. As discussed by Su et al.,⁴⁹ the initial state is approximated here to the free protein (p). With the GBr⁶ method,⁴³ the electrostatic polarization energy calculated for the training set changes by 0.2–0.7 kcal mol⁻¹ between these two initial state definitions. This polarization response is scaled by the cavity polarity descriptor in eq 6. A combination with the ligand polarity did not result in better predictions.

Nonpolar solvent contributions from the free ligand and the receptor complex were replaced by a simple Δ SASA term without loss of accuracy. Previous work suggested that the constant term in a LIE equation, γ in eq 1 or k_6 in eq 6, may be related to binding site hydrophobicity^{32,34} or nonpolar surface area.⁸⁸ An ALICE model (eq S3 in the SI) in which the constant term is scaled by the nonpolar cavity surface $(1-\eta)$ was tested, but this modification also did not result in better performance.

The comparable accuracy obtained here for different LIE models (eq 6 and eqs S1–S3) suggests the exact form of a LIE equation is less important given a proper parametrization is conducted. Significant departure from theoretical values is observed for the parameters obtained here. For instance, $k_2(\eta+\pi)$, equivalent to the parameter β in previous LIE models (eq 1), ranged from 0.03 to 0.27. This is below the theoretical value of $\beta = 0.5$.^{29,49} It is expected that parameter values will mutually compensate model assumptions and inaccuracies in the solvent model, molecular mechanical potentials, etc. Thus, the parametrized LIE equations presented here may be cast as linear free energy relationships which coefficients are only bounded by the linear response theory.³⁰

Applications of these LIE models depend, however, on their transferability for receptors and ligands not included in the training set used to parametrize the equations. Here coefficients

were scaled by ligand and cavity polarities in order to increase model transferability.⁴⁰ Eq 6 correctly predicts affinities for ligands which receptors were either included (T4 lysozyme and HIVRT) or not (FKBP) in the training set. The sensitivity of ΔG_{ALICE} on the η and π descriptor values is small. Variations of ~ 0.2 kcal mol⁻¹ were observed when descriptor values were scaled by $\pm 20\%$. Predictions for other receptors and ligands should have similar accuracy, but an extensive test of transferability is left for future studies.

The computational efficiency observed for eq 6 suggests it can be used to predict affinities for large ligand sets. For instance, our ALICE model could be used instead of the scoring or energy functions currently employed in molecular docking. To this end, ligand–receptor poses would have to be generated by the conformational search procedures found in docking^{21,50} or by another method such as mining-minima.⁸⁹ Ligand and protein topologies containing connectivity, force field parameters, and polarity descriptors would have to be available or built. Although cumbersome when manually done, this process can be made fairly automatic.^{64,66,90}

In order to improve the ALICE model proposed, it may be useful to analyze the highest deviations found. The following ligands are described as eq 6 outliers since they show deviations larger than one RMSD: benzene, toluene, 1,3,5-trimethylbenzene (L99A), 2-fluoroaniline, catechol, 3-methylpyrrole, thieno-[3,2-*b*]thiophene, 4-vinylpyridine, 2-methoxyphenol, *N*-(*o*-tolyl)cyanofornamide (M102Q), and HIV3. All but 2-fluoroaniline belong to the test set, and all T4 lysozyme outliers have underestimated free energies. Most of these ligands are also outliers for eqs S1–S3 (see Table S6).

The binding affinity increases for L99A ligands upon addition of linear methylene units, as seen for benzene to toluene and up to *n*-butylbenzene. The experimental free energy difference upon methylene addition in this series is 0.2–0.3 kcal mol⁻¹ but between ethylbenzene and propylbenzene, which is 0.8 kcal mol⁻¹. Although eq 6 incorrectly predicts a small stability to benzene and toluene, appropriate affinities are predicted upon increasing the number of methylene units. This observation suggests a slightly unbalanced description of the nonpolar contributions involving aliphatic and aromatic carbons. An atom-type dependent surface tension, k_5 in eq 6, could amend this problem.

The hydrophobic ligand 1,3,5-trimethylbenzene should interact more favorably with the L99A nonpolar engineered binding site than with water, as suggested by the free energy calculated with the ALICE model. Docking suggests that there is enough room to accommodate this relatively bulky ligand in the L99A cavity (Table S7). However, the experimental free energy shows that 1,3,5-trimethylbenzene is a L99A nonbinder. As T4 lysozyme must show some breathing or opening movement to allow the entrance or exit of ligands from the engineered cavity,⁹¹ we speculate that 1,3,5-trimethylbenzene is a kinetic nonbinder and that there may not be a low energy pathway allowing its entrance into the L99A cavity.

The highest number of outliers were found for M102Q ligands. Possibly electrostatic interactions were not described or sampled correctly in the LIE models tested. Yet no correlation was found between outliers and significant flaws in the description of their dipole moments (Table S1) or the availability of experimental ligand–receptor structures. The lack of holo crystal structures for some ligands and the possibility that receptor structures used for the calculation of ΔG_{ALICE} are not representative of complexes observed in

aqueous solution may contribute for inaccuracies in the prediction of intrinsic binding free energies. This appears to be the case for catechol as discussed in the next section.

4.2. Analysis of the Proposed Conformational Averaging Schemes. Ensembles for all target receptors were initially built with apo protein representations. For the T4 lysozyme mutants, 3 structures out of 50 in the L99A ensemble and 11 structures out of 50 for the M102Q ensemble had their binding cavity fully blocked by side-chains rotations. Docking to these receptor structures did not yield complexes with ligand inside the engineered binding site even for small ligands such as toluene. This problem was more pronounced for the HIVRT and FKBP bulky ligands as none of the receptor structures obtained from apo molecular dynamics after equilibration were able to accommodate ligands in its crystallographic binding site. Thus, molecular dynamics obtained from holo structures were used to generate the HIVRT and FKBP ensembles.

Affinities were consistently underestimated for T4 lysozyme mutants by all three combinations of ensemble averages tested. Although ΔG_{DD} shows a small RMSD and a favorable error analysis for L99A ligands, the dominant approximation that counts only the most favorable pose underestimates affinities for almost all T4 lysozyme ligands but catechol (see below). This tendency suggests that the tentative apo M102Q ensemble used here systematically degrades the structural representation in comparison to the crystal structures. As noted, side-chains rotations block the binding cavity even for the relatively rigid engineered site in T4 lysozyme. Apo protein ensembles have to be used carefully and possibly enlarged or modified⁹² to accommodate ligand binding.

The only notable exception is catechol binding to M102Q. Although the ALICE free energy function predicts unstable binding for catechol to the crystal (holo) configuration, favorable binding is predicted by ΔG_{DD} . Thus, receptor configurational sampling is important for the prediction of catechol binding, and the M102Q crystal configuration may have a low contribution to the binding affinity. In fact, it has been shown that catechol has at least two binding modes⁹³ and that enhanced sampling is required to compute its binding free energy correctly.¹⁸

Increasing the number of poses that contribute to exponential averaging results in less favorable binding free energies for all ligands tested. This “dilution” effect has two possible causes. One is artificially related to the discrimination of poses that should contribute to the ensemble averages. It depends on the discriminatory quality of the intrinsic free energy function. The second cause has a physical origin as the macroscopically measured affinity may be an average of several binding poses, some of which will have higher intrinsic free energies of binding.

Free energies calculated for the T4 lysozyme mutants decrease up to 50% on going from ΔG_{DD} to ΔG_{EE} . The dilution effect is more pronounced for T4 lysozyme because the intrinsic free energy differences between ligand poses inside or outside the engineered binding site is small (<2.0 kcal mol⁻¹). Consequently, it is harder to discriminate poses that should contribute to the ensemble averages from decoy poses. In fact, for all T4 lysozyme ligands, the average number of poses included in the calculation of the binding potential of mean-force (eq 4) is higher than the average number of poses inside the receptor binding site (Table S7).

The importance of pose discrimination on ensemble averages is also illustrated when the Vina energy function is used as an

approximation instead of ΔG_{ALICE} (eq 6). The exponentially averaged free energies obtained in this case are significantly less favorable than the corresponding dominant approximation (ΔG_{DD}^{Vina}) for the same set of ligand–receptor poses.

For HIVRT and FKBP ligands, pose discrimination with the ALICE model is rather accurate. Although the absolute values of binding free energies are higher, the predictions with the ALICE intrinsic free energy function show smaller deviations for both receptors (Table 2). The better discrimination is reflected by the average number of poses used to calculate B_E which is closer to the average number of poses found inside the receptor binding site (Table S7).

Thus, for HIVRT and FKBP ligands, the decrease of calculated free energies on going from ΔG_{DD} to ΔG_{EE} does not appear artificial but a physical effect due to averaging a distribution of ligand–receptor conformations. Induced fit is expected to change the receptor conformational distribution upon ligand complexation. Indeed, some of the configurations found for the complexes with the receptor ensemble show more favorable intrinsic free energies of binding. This also explains the overstabilization observed for ΔG_{DD} predictions.

Receptor reorganization free energy will be accounted for if the ensemble is canonically distributed and wide enough to represent the relevant receptor motions. If ligand binding has a large reorganization free energy, the associated receptor motions will have high free energy costs, and the number or fraction of binding-competent configurations in the ensemble will be small. Consequently, a ligand complex formed with such rare receptor configurations will only contribute significantly to the final exponential average if it has a highly favorable intrinsic binding free energy. On the other hand, if the reorganization free energy for ligand binding is small, the fraction of binding-competent configurations in the ensemble will be large, and a reasonable number of ligand complexes with these popular receptor configurations will contribute to the average even if their intrinsic binding free energies are not highly favorable.

We do not investigate here what is the appropriate size of the ensemble to account for the reorganization energy correctly, but 50 configurations appear to be insufficient, specially if an apo structure is used to generate the ensemble.⁹² When a holo structure is used to generate a tentative ensemble, the full receptor reorganization energy may not be accounted for, but the relative contribution upon a series of congeneric ligands can be retrieved.

Finally, which of the three averaging procedures should be applied? The answer depends on the receptor ensemble and the intrinsic free energy function. For receptor ensembles with unknown or biased distributions, with insufficient sampling and for free energy functions that cannot discriminate correct binding poses from decoys, the dominant pose and state approximation (ΔG_{DD}) should be applied. For more flexible receptors and upon increasing the quality of the conformational ensemble, the state exponential average (ΔG_{DE}) might be included. The exponential average for multiple ligand configurations (ΔG_{EE}) should be employed only when a calibrated and discriminatory intrinsic free energy function is available. This is probably the case for the ALICE model proposed (eq 6) with lead-like ligands.

5. CONCLUSIONS

We have parametrized an adapted linear interaction energy model that employs a simple combination of energy

minimization of ligand-protein geometries with implicit solvation. This model is able to retrospectively predict binding affinities for different receptors with accuracy similar to other LIE models which employ more expensive molecular dynamics simulations to sample configurations and more detailed solvent models.^{29–31,40,49} LIE models using geometry optimization and implicit solvation have already been successfully applied by other authors.^{36,47}

Conformational sampling is divided and approximated in several steps here. Solvent degrees of freedom, in particular the important dielectric response, are treated implicitly with a continuum electrostatic model. Ligand internal torsion and relative orientation in receptor complexes are evaluated within the docking algorithm when poses are generated. Finally, receptor conformations are sampled from a configurational ensemble, which is generated once and repeatedly used for preparing ligand-protein complexes for a series of congeneric ligands. It should be noted that the proposed ALICE model can be used to predict binding affinities given either a single receptor structure or a conformational distribution.

For the prediction of binding to flexible protein targets, it is useful to represent the receptor structure by conformational ensembles.^{6,7,53,56,57} Based on the implicit ligand theory,⁶¹ averaging procedures were proposed and tested to estimate affinities for ligand binding to four different receptors with structures represented by tentative conformational ensembles. The scheme proposed is computationally efficient and could be applied to average contributions of $\sim 10^6$ ligand-receptor poses for each ligand tested.

In principle, to be useful in the prediction of ligand binding, a receptor conformational ensemble should follow the Boltzmann distribution and describe all motions or structural rearrangements relevant for small-molecule complexation. Here, tentative ensembles were built without much attention to statistical distribution or to sampling of relevant motions by running simple molecular dynamics in implicit solvent. Consequently, results obtained by averaging contributions from these receptor ensembles do not aim to reproduce experimental affinities and should be analyzed only qualitatively. It should be noted that there is no consensus on how to obtain conformational ensembles that are suitable to predict ligand binding.^{7,53,54,92}

Nevertheless, we find that good discrimination between binding poses and decoys is essential to calculate accurate binding affinities, particularly when contributions from several putative binding poses and different receptor configurations are (exponentially) averaged. Approximations used here such as sampling by docking instead of a statistical distribution or inaccuracies in the intrinsic free energy function may contribute to the difficulty in distinguishing binding poses from decoys. For lead-like ligands, we found that the ALICE model proposed is able to discriminate poses resulting in binding free energy predictions in good agreement with experiment.

■ ASSOCIATED CONTENT

● Supporting Information

Equations S1–S3, figures with HIVRT and FKBP ligand structures and an example of ligand pose distribution for T4 lysozyme, and tables with dipole moments, number of torsions activated in docking, experimental energies, descriptors and LIE contributions for all ligands considered, training set predictions, error analysis, and number of poses used in averaging

procedures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: garantes@iq.usp.br.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding from FAPESP (projects 11/04354-6 and 12/02501-4) and CNPq (project 141950/2013-7) are gratefully acknowledged.

■ REFERENCES

- (1) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303*, 1813–1818.
- (2) Ballester, P. J.; Schreyer, A.; Blundell, T. L. Does a More Precise Chemical Description of Protein–Ligand Complexes Lead to More Accurate Prediction of Binding Affinity? *J. Chem. Inf. Model.* **2014**, *54*, 944–955.
- (3) Fleishman, S.; Baker, D. Role of the Biomolecular Energy Gap in Protein Design, Structure, and Evolution. *Cell* **2012**, *149*, 262–273.
- (4) Faver, J. C.; Yang, W.; Merz, K. M. The Effects of Computational Modeling Errors on the Estimation of Statistical Mechanical Variables. *J. Chem. Theory Comput.* **2012**, *8*, 3769–3776.
- (5) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40*, 41–62.
- (6) Benedix, A.; Becker, C. M.; de Groot, B. L.; Cafilisch, A.; Bockmann, R. A. Predicting Free Energy Changes Using Structural Ensembles. *Nat. Methods* **2009**, *6*, 3–4.
- (7) Arantes, G. M. Flexibility and Inhibitor Binding in Cdc25 Phosphatases. *Proteins* **2010**, *78*, 3017–3032.
- (8) Gilson, M. K.; Zhou, H. X. Calculation of Protein–Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (9) Guvench, O.; MacKerell, A. D., Jr. Computational Evaluation of Protein–Small Molecule Binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 56–61.
- (10) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
- (11) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: a Review. *J. Comput. Chem.* **2010**, *31*, 1569–1582.
- (12) Gallicchio, E.; Levy, R. M. Advances in All Atom Sampling Methods for Modeling Protein–Ligand Binding Affinities. *Curr. Opin. Struct. Biol.* **2011**, *21*, 161–166.
- (13) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
- (14) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV–1 Reverse Transcriptase Via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 145–154.
- (15) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
- (16) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized–Over–Parts Computation of Absolute Binding Free Energy with Docking and Molecular Dynamics. *J. Chem. Phys.* **2006**, *125*, 084901.
- (17) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009**, *394*, 747–763.

- (18) Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.* **2006**, *125*, 084902.
- (19) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule–Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (20) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein–Ligand Docking: Current Status and Future Challenges. *Proteins* **2006**, *65*, 15–26.
- (21) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.
- (22) Bowman, A. L.; Nikolovska-Coleska, Z.; Zhong, H.; Wang, S.; Carlson, H. A. Small Molecule Inhibitors of the MDM2–P53 Interaction Discovered by Ensemble–Based Receptor Models. *J. Am. Chem. Soc.* **2007**, *129*, 12809–12814.
- (23) Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can We Trust Docking Results? Evaluation of Seven Commonly Used Programs on PDBbind Database. *J. Comput. Chem.* **2011**, *32*, 742–755.
- (24) Keenan, S. M.; Geyer, J. A.; Welsh, W. J.; Prigge, S. T.; Waters, N. C. Rational Inhibitor Design and Iterative Screening in the Identification of Selective Plasmodial Cyclin Dependent Kinase Inhibitors. *Comb. Chem. High Throughput Screening* **2005**, *8*, 27–38.
- (25) Bisson, W. H.; Cheltsov, A. V.; Bruey-Sedano, N.; Lin, B.; Chen, J.; Goldberger, N.; May, L. T.; Christopoulos, A.; Dalton, J. T.; Sexton, P. M.; Zhang, X.-K.; Abagyan, R. Discovery of Antiandrogen Activity of Nonsteroidal Scaffolds of Marketed Drugs. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11927–11932.
- (26) Frimurer, T. M.; Peters, G. H.; Iversen, L. F.; Andersen, H. S.; Møller, N. P. H.; Olsen, O. H. Ligand–Induced Conformational Changes: Improved Predictions of Ligand Binding Conformations and Affinities. *Biophys. J.* **2003**, *84*, 2273–2281.
- (27) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
- (28) Kim, R.; Skolnick, J. Assessment of Programs for Ligand Binding Affinity Prediction. *J. Comput. Chem.* **2008**, *29*, 1316–1331.
- (29) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer–Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.
- (30) Carlson, H. A.; Jorgensen, W. L. An Extended Linear Response Method for Determining Free Energies of Hydration. *J. Phys. Chem.* **1995**, *99*, 10667–10673.
- (31) Gallicchio, E.; Kubo, M. M.; Levy, R. M. Enthalpy–Entropy and Cavity Decomposition of Alkane Hydration Free Energies: Numerical Results and Implications for Theories of Hydrophobic Solvation. *J. Phys. Chem. B* **2000**, *104*, 6271–6285.
- (32) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.
- (33) King, G.; Warshel, A. Investigation of the Free Energy Functions for Electron Transfer Reactions. *J. Chem. Phys.* **1990**, *93*, 8682–8692.
- (34) Almlof, M.; Brandsdal, B. O.; Åqvist, J. Binding Affinity Prediction with Different Force Fields: Examination of the Linear Interaction Energy Method. *J. Comput. Chem.* **2004**, *25*, 1242–1254.
- (35) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding Constants of Neuraminidase Inhibitors: an Investigation of the Linear Interaction Energy Method. *J. Med. Chem.* **1999**, *42*, 5142–5152.
- (36) Kolb, P.; Huang, D.; Dey, F.; Caflich, A. Discovery of Kinase Inhibitors by High–Throughput Docking and Scoring Based on a Transferable Linear Interaction Energy Model. *J. Med. Chem.* **2008**, *51*, 1179–1188.
- (37) Stjerschantz, E.; Marelus, J.; Medina, C.; Jacobsson, M.; Vermeulen, N. P. E.; Oostenbrink, C. Are Automated Molecular Dynamics Simulations and Binding Free Energy Calculations Realistic Tools in Lead Optimization? An Evaluation of the Linear Interaction Energy (LIE) Method. *J. Chem. Inf. Model.* **2006**, *46*, 1972–1983.
- (38) de Amorim, H. L. N.; Caceres, R. A.; Netz, P. A. Linear Interaction Energy (LIE) Method in Lead Discovery and Optimization. *Curr. Drug Targets* **2008**, *9*, 1100–1105.
- (39) Hansson, T.; Marelus, J.; Åqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27–35.
- (40) Linder, M.; Ranganathan, A.; Brinck, T. Adapted Linear Interaction Energy²: A Structure–Based LIE Parametrization for Fast Prediction of Protein–Ligand Affinities. *J. Chem. Theory Comput.* **2013**, *9*, 1230–1239.
- (41) Zhou, R. H.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. New Linear Interaction Method for Binding Affinity Calculations Using a Continuum Solvent Model. *J. Phys. Chem. B* **2001**, *105*, 10388–10397.
- (42) Alves, A. F. N. M.Sc. thesis, Instituto de Química, Universidade de São Paulo, São Paulo, 2013. Available at <http://www.teses.usp.br/teses/disponiveis/46/46131/tde-08052013-144801/> (accessed June 20, 2014).
- (43) Tjong, H.; Zhou, H.-X. GBr⁶: A Parameterization–Free, Accurate, Analytical Generalized Born Method. *J. Phys. Chem. B* **2007**, *111*, 3055–3061.
- (44) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (45) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large–Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383–394.
- (46) Schaefer, M.; Bartels, C.; Karplus, M. Solution Conformations and Thermodynamics of Structured Peptides: Molecular Dynamics Simulation with an Implicit Solvation Model. *J. Mol. Biol.* **1998**, *284*, 835–848.
- (47) Huang, D.; Caflich, A. Efficient Evaluation of Binding Free Energy Using Continuum Electrostatic Solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.
- (48) Carlsson, J.; Ander, M.; Nervall, M.; Åqvist, J. Continuum Solvation Models in the Linear Interaction Energy Method. *J. Phys. Chem. B* **2006**, *110*, 12034–12041.
- (49) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. Linear Interaction Energy (LIE) Models for Ligand Binding in Implicit Solvent: Theory and Application to the Binding of NNRTIs to HIV–1 Reverse Transcriptase. *J. Chem. Theory Comput.* **2007**, *3*, 256–277.
- (50) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A Semiempirical Free Energy Force Field with Charge–Based Desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.
- (51) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
- (52) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: the Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 45–55.
- (53) Wong, C. F.; Kua, J.; Zhang, Y.; Straatsma, T.; McCammon, J. A. Molecular Docking of Balanol to Dynamics Snapshots of Protein Kinase A. *Proteins* **2005**, *61*, 850–858.
- (54) Mamonov, A. B.; Bhatt, D.; Cashman, D. J.; Ding, Y.; Zuckerman, D. M. General Library–Based Monte Carlo Technique Enables Equilibrium Sampling of Semi–Atomistic Protein Models. *J. Phys. Chem. B* **2009**, *113*, 10891–10904.
- (55) Kallblad, P.; Mancera, R. L.; Todorov, N. P. Assessment of Multiple Binding Modes in Ligand–Protein Docking. *J. Med. Chem.* **2004**, *47*, 3334–3337.
- (56) Novoa, E. M.; de Pouplana, L. R.; Barril, X.; Orozco, M. Ensemble Docking from Homology Models. *J. Chem. Theory Comput.* **2010**, *6*, 2547–2557.
- (57) Totrov, M.; Abagyan, R. Flexible Ligand Docking to Multiple Receptor Conformations: A Practical Alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.

- (58) B-Rao, C.; Subramanian, J.; Sharma, S. D. Managing Protein Flexibility in Docking and Its Applications. *Drug Discovery Today* **2009**, *14*, 394–400.
- (59) Cozzini, P.; Kellogg, G. E.; Spyraakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. Target Flexibility: an Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, *51*, 6237–6255.
- (60) Gallicchio, E.; Lapelosa, M.; Levy, R. M. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein–Ligand Binding Affinities. *J. Chem. Theory Comput.* **2010**, *6*, 2961–2977.
- (61) Minh, D. D. L. Implicit Ligand Theory: Rigorous Binding Free Energies and Thermodynamic Expectations from Molecular Docking. *J. Chem. Phys.* **2012**, *137*, 104106.
- (62) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A Cavity-Containing Mutant of T4 Lysozyme Is Stabilized by Buried Benzene. *Nature* **1992**, *355*, 371–373.
- (63) Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. A Model Binding Site for Testing Scoring Functions in Molecular Docking. *J. Mol. Biol.* **2002**, *322*, 339–355.
- (64) Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B. The Geometry of the Ribosomal Polypeptide Exit Tunnel. *J. Mol. Biol.* **2006**, *360*, 893–906.
- (65) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area As a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714–3717.
- (66) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: Design and Development of a Database for Comprehensive Analysis of Protein–Ligand Interactions. *J. Mol. Biol.* **2003**, *326*, 607–620.
- (67) Carroll, D. L. *Genetic Algorithm Driver, Version 1.7*; 1998.
- (68) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in FORTRAN 77: the Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, 1992.
- (69) Arantes, G. M.; Loos, M. Specific Parametrisation of a Hybrid Potential to Simulate Reactions in Phosphatases. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347–353.
- (70) China, G.; Padron, G.; Hooft, R. W. W.; Sander, C.; Vriend, G. The Use of Position-Specific Rotamers in Model-Building by Homology. *Proteins* **1995**, *23*, 415–421.
- (71) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (72) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: an Open Chemical Toolbox. *J. Cheminf.* **2011**, DOI: doi:10.1186/1758-2946-3-33.
- (73) Frisch, M. J. et al. *Gaussian 09, Revision A.1*; Gaussian, Inc.: Wallingford, CT, 2009.
- (74) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (75) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the Development of Universal, Fast and Highly Accurate Docking/scoring Methods: a Long Way to Go. *Br. J. Pharmacol.* **2008**, *153*, S7–S26.
- (76) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (77) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (78) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (79) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (80) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (81) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: an N×log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (82) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (83) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (84) Li, J.; Xing, J.; Cramer, C. J.; Truhlar, D. G. Accurate Dipole Moments from Hartree–Fock Calculations by Means of Class IV Charges. *J. Chem. Phys.* **1999**, *111*, 885–892.
- (85) *Ligand Topologies*. http://gaznevada.iq.usp.br/wp-content/uploads/ops_site.tar.gz (accessed May 2014).
- (86) Jones-Hertzog, D. K.; Jorgensen, W. L. Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method. *J. Med. Chem.* **1997**, *40*, 1539–1549.
- (87) Zar, J. H. *Biostatistical Analysis*, 4th ed.; Prentice Hall: NJ, 1999.
- (88) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. Empirical Correlation Between Hydrophobic Free Energy and Aqueous Cavity Surface Area. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 2925–2927.
- (89) Chen, W.; Gilson, M. K.; Webb, S. P.; Potter, M. J. Modeling Protein–Ligand Binding by Mining Minima. *J. Chem. Theory Comput.* **2010**, *6*, 3540–3557.
- (90) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM General Force Field: a Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (91) Mulder, F. A. A.; Hon, B.; Muhandiram, D. R.; Dahlquist, F. W.; Kay, L. E. Flexibility and Ligand Exchange in a Buried Cavity Mutant of T4 Lysozyme Studied by Multinuclear NMR. *Biochemistry* **2000**, *39*, 12614–12622.
- (92) Seeliger, D.; de Groot, B. L. Conformational Transitions upon Ligand Binding: Holo-Structure Prediction from Apo Conformations. *PLoS Comput. Biol.* **2010**, *6*, e1000634.
- (93) Graves, A. P.; Brenk, R.; Shoichet, B. K. Decoys for Docking. *J. Med. Chem.* **2005**, *48*, 3714–3728.

ANEXO J – Partition, orientation and mobility of ubiquinones in a lipid bilayer



Partition, orientation and mobility of ubiquinones in a lipid bilayer



Vanessa Viviana Galassi, Guilherme Menegon Arantes *

Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900, São Paulo, SP, Brazil

ARTICLE INFO

Article history:

Received 19 May 2015

Received in revised form 2 July 2015

Accepted 4 August 2015

Available online 6 August 2015

Keywords:

Membrane localization
Membrane insertion
Computer simulation
Free-energy profile
Molecular dynamics
Force-field parametrization

ABSTRACT

Ubiquinone is the universal mobile charge carrier involved in biological electron transfer processes. Its redox properties and biological function depend on the molecular partition and lateral diffusion over biological membranes. However, ubiquinone localization and dynamics within lipid bilayers are long debated and still uncertain. Here we present molecular dynamics simulations of several ubiquinone homologs with variable isoprenoid tail lengths complexed to phosphatidylcholine bilayers. Initially, a new force-field parametrization for ubiquinone is derived from and compared to high level quantum chemical data. Free energy profiles for ubiquinone insertion in the lipid bilayer are obtained with the new force-field. The profiles allow for the determination of the equilibrium location of ubiquinone in the membrane as well as for the validation of the simulation model by direct comparison with experimental partition coefficients. A detailed analysis of structural properties and interactions shows that the ubiquinone polar head group is localized at the water–bilayer interface at the same depth of the lipid glycerol groups and oriented normal to the membrane plane. Both the localization and orientation of ubiquinone head groups do not change significantly when increasing the number of isoprenoid units. The isoprenoid tail is extended and packed with the lipid acyl chains. For ubiquinones with long tails, the terminal isoprenoid units have high flexibility. Calculated ubiquinone diffusion coefficients are similar to that found for the phosphatidylcholine lipid. These results may have further implications for the mechanisms of ubiquinone transport and binding to respiratory and photosynthetic protein complexes.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Ubiquinone or coenzyme-Q was described over five decades ago as a central molecule in cellular respiration and photosynthesis [1, 2]. It is the ubiquitous mobile charge carrier linking distant redox centers in biological electron transfer chains. Ubiquinone receives electrons from NADH:ubiquinone reductase (respiratory complex I), succinate:ubiquinone reductase (complex II) in bacteria or mitochondria and reaction centers in purple and cyanobacteria. Cytochrome bc_1 (complex III) and the related cytochrome bf_6 then oxidize ubiquinol, the reduced counterpart of ubiquinone, to continue the electron transfer chain. Biological membranes harboring the proteins involved in bio-energetical processes contain a pool of ubiquinone and ubiquinol molecules which is responsible for 95% of the energy transduction in aerobic organisms [3–5].

Ubiquinone is an amphipathic molecule formed by a modified benzoquinone polar head and a lipid tail composed of 6 to 10 isoprenoid units in biologically relevant homologs [6,3]. The quinone head may be reduced by up to two electrons and receive two protons, passing through a semi-quinone intermediate up to the quinol

form. Ubiquinone redox properties and biological function as charge carriers will depend on its partition and diffusion over biological membranes. But, despite much experimental progress obtained during the last decades, the localization and orientation of ubiquinones in biological membranes are still under heavy debate [6–15].

It has been proposed that the polar head of ubiquinones with more than 3 isoprenoid units are fully embedded in the membrane core and oriented parallel to the interfacial plane, lying in the bilayer midplane space [6–10]. Other authors have suggested that the quinone head is located close to the phospholipid polar groups and oriented parallel to its hydrocarbon tails, with the isoprenoid chain spanning the membrane [11–15]. Other proposals include the formation of ubiquinone aggregates or the partition between separate pools inside the lipid bilayer [14,16].

There is also a lack of consensus on the dynamical properties of ubiquinone embedded in lipid bilayers [17–21] as diffusion constants experimentally determined range from 10^{-9} up to 10^{-6} $\text{cm}^2 \text{s}^{-1}$. This high variation is due to the application of various measurement techniques [22,19] and to the usage of probes (fluorescent, spin labels, etc.) attached to ubiquinone which will alter its intrinsic diffusion properties. Nevertheless, it has been suggested that ubiquinone diffusion between respiratory complexes is the rate limiting step in the whole mitochondrial electron transfer chain [17].

* Corresponding author.

E-mail address: garantes@iq.usp.br (G.M. Arantes).

Computer simulations can give detailed information on the partition, orientation and mobility of solutes embedded in lipid bilayers [23–25]. The accuracy and precision of the results obtained in a simulation will depend both on the energy description – the molecular mechanics (MM) force-field – and on appropriate sampling of the relevant degrees of freedom. Although there are a few simulation studies focusing on ubiquinone properties when bound to proteins [26–29], to our knowledge there is only one simulation study of ubiquinone complexed to a lipid bilayer [22,30]. Unfortunately, this study reports a very short simulation time (less than 20 ns in aggregate) which is not enough to reach equilibration and proper sampling.

Here we apply long molecular dynamics (MD) simulations and free energy calculations to investigate the properties of several ubiquinone homologs bound to model lipid bilayers. Although ubiquinone force-fields were previously available [26,28], we have tested their performance both in comparison to quantum chemical reference properties and to experimental data and judged necessary to build a new parametrization for ubiquinone compatible with the CHARMM force-field [31,32]. This energy function has been carefully validated for a series of lipids [33], including 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) which was used here to build a model bilayer. We find that the polarity and flexibility of isoprenoid units should be carefully calibrated in order to obtain an accurate description for ubiquinone partition in membranes. The insertion process is simulated with enhanced sampling of several degrees of freedom that may be relevant to the interaction of ubiquinone with the bilayer. After validation of the simulations by comparison to available experimental measurements, we present a detailed analysis of the simulation results providing interesting points for the ubiquinone localization, interactions and lateral mobility when embedded in a POPC bilayer.

2. Materials & methods

2.1. Reference set and force-field parameterization strategy

A detailed parametrization for ubiquinone was performed here in accordance with the CHARMM force-field [31,32]. Details are given below for the ab initio electronic structure calculations used to build the training and test sets, and for the fitting strategy used to adjust the parameters. Ubiquinone-1 (UQ1 or UQ n , for n isoprenoid units), ubiquinol-1 (UQ1H2) and 3-methyl-3-hexene (isoprene or ISP) were used here as molecular fragments for the force-field parametrization. Please refer to Fig. 1 for atom naming used throughout the text. An equivalent naming scheme is used for ketonic ubiquinone. Isoprene was used to adjust parameters for the isoprenoid tail, whereas UQ1 and UQ1H2 were used for the quinone head.

A reference set of ubiquinone interaction energies, bond torsion potentials and molecular dipoles was built from quantum chemical (QC) calculations and used as targets for the force-field parametrization. QC calculations were conducted with GAUSSIAN 09 (Rev. A01) [34] for molecules and complexes in the gas phase. All geometry optimizations were performed with the B3LYP functional [35] and the 6–31+G** basis set [36]. All single-point energies and quantum mechanical dipole moments used in the reference set were calculated at the MP2/6-311+G(2df,2p) level [37].

Hydrogen-bonded ubiquinone–water complexes were used to model solvent interactions. Three different orientations were used for ubiquinone with oxygens O4, O1 and O3 (equivalent to O2) accepting the water hydrogen and two orientations were used with oxygen in ubiquinol (O4 and O1) donating a hydrogen to water. Hydrophobic contacts were modeled by complexes between UQ1 and methane and a ISP dimer. Interaction energies (E_i in Eq. (1)) were computed with a counterpoise scheme to correct for the basis-set superposition error [38]. The distance and orientation of all complexes were optimized whereas internal coordinates of each molecule were kept frozen at

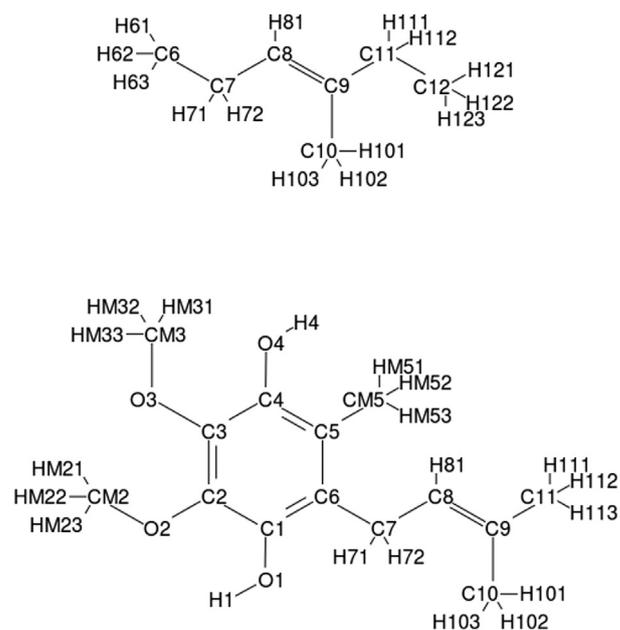


Fig. 1. Structure and atom naming of isoprene (ISP) and ubiquinol-1 (UQ1H2).

their isolated geometry. Water internal coordinates were kept frozen at the TIP3P [39] equilibrium geometry. Please refer to the Results (Fig. 2) and Supporting Information (SI, Fig. S1) for the relative orientation of the dimer complexes.

Torsions around bonds O2–C2, C6–C7 in UQ1, O1–C1, O2–C2, O4–C4 in UQ1H2, and C7–C8 and C9–C11 in ISP were evaluated (Figs. 3 and S2). Torsion potentials (E_t in Eq. (1)) were scanned in 15° or 30° angle increments by partial geometry optimizations with fixed bond lengths but flexible angle and dihedrals to avoid clashes and obtain torsional profiles representative of configurations visited at normal thermal energies. Torsions were further divided into polar (tp) and non-polar (tnp). Torsion rotamers were selected for the calculation of molecular dipoles (D in Eq. (1)).

The following error function was minimized by adjusting the force-field parameters:

$$P = \sum_i N_{int} w_i \sum_j N_{sp}^{int} \left| \frac{E_{i,j}^{MM} - E_{i,j}^{QC}}{E_{i,j}^{QC}} \right| + \sum_t N_{tp} w_{tp} \sum_j N_{sp}^{tp} \left| \frac{E_{t,j}^{MM} - E_{t,j}^{QC}}{E_{t,j}^{QC}} \right| + \sum_t N_{tnp} w_{tnp} \sum_j N_{sp}^{tnp} \left| \frac{E_{t,j}^{MM} - E_{t,j}^{QC}}{E_{t,j}^{QC}} \right| + w_D \sum_j N_{sp}^D \sum_{\alpha=x,y,z} \left| \frac{D_{j,\alpha}^{MM} - D_{j,\alpha}^{QC}}{D_{j,\alpha}^{QC}} \right| \quad (1)$$

where w are the weights assigned to each property, $N_{int} = 5$ is the number of interaction pair complexes (3 hydrogen-bonded and 2 hydrophobic), and $N_{tp} = 3$ and $N_{tnp} = 4$ are the number of polar and non-polar bond torsions included in the training set. N_{sp} geometries were used for each property: $N_{sp}^{int} = 3$, including the minima and two adjacent points along the interaction profile; N_{sp}^{tp} and N_{sp}^{tnp} were either 6 or 12, depending on the angle increments, and $N_{sp}^D = 2, 4$ and 4 conformations respectively for UQ1, UQ1H2 and ISP. Superscripts MM and QC indicate the calculation method.

In order to obtain parameters for ubiquinone compatible with the CHARMM force-field, atom types were assigned by chemical analogy with previously parametrized molecules (Table S1). Bond and valence angle parameters were kept from the standard CHARMM values. Improper dihedrals were not included in aromatic rings or unsaturated bonds as they were not necessary to maintain planarity. Root mean square deviations of Cartesian coordinates are smaller than 0.01 nm for UQ1 geometries optimized with the force-field parametrized here in comparison to geometries optimized with the reference QC method.

Partial charges were adjusted for atoms C2/C3, O2/O3, CM2/CM3, CM5, C6, ketonic groups of UQ1 (C1/C4 and O1/O4) or phenolic groups of UQ1H2 (C1/C4, O1/O4 and H1/H4), and for isoprene C9, C10 and C11 (Fig. 1). Groups with neutral total charge were defined among the following bound atom sets: (C2, O2, CM2, HM2) and the equivalent CM3 methoxide group, (C6, C5, CM5, HM5), (C9, C10, H10, C11, H11), (C1, O1) and (C4, O4) in quinones and (C1, O1, H1) and (C4, O4, H4) in quinols. Charges for atoms in the aromatic ring and its substituents were made symmetric. Charges for isoprenoid tails longer than found in UQ1 were adapted accordingly. Partial charges for all other atom types such as aliphatic hydrogens were kept in their standard CHARMM values.

Bond torsions were fitted by altering the number of expansion terms of order n , their phase δ and force constant K_ϕ (Table S3) in the corresponding Fourier dihedral terms [31]. For ISP, torsions around bonds C7–C8 and C9–C11 were adjusted. For UQ1, polar torsion C2–O2 (and the equivalent C3–O3) and non-polar torsions C6–C7 and C7–C8 and for UQ1H2 polar torsions C1–O1 and C4–O4 were optimized.

Partial charges were optimized first by setting $w_i = 1$, $w_D = 1$, $w_{tp} = 0.04$ and $w_{mp} = 0$ and minimizing Eq. (1). Then, dihedral parameters were optimized by setting $w_i = 1$, $w_D = 0$, $w_{tp} = 0.01$ –1.0 and $w_{mp} = 1$. After these two optimization cycles, torsions around bonds C6–C7 and C2–O2 (and the equivalent C3–O3) were still not described satisfactorily. Then, the R_{min} Lennard-Jones parameter for atoms C8, CM5, O2, O3, CM2 and CM3 were slightly changed by evaluating only the error of these three torsions. A final optimization cycle was performed by including the full training set and varying all partial charges and dihedral parameters at the same time.

All MM energies and properties used in the force-field parametrization were obtained using GROMACS package version 4.6.1 [40]. For the parameter optimization, a combination of genetic algorithm and simplex was used as described before [41]. See Tables S1, S2 and S3 for the final optimized parameter values. GROMACS compatible topologies are available upon request from the authors.

2.2. System setup and molecular dynamics simulations

The force-field obtained was tested by molecular dynamics simulations and free energy calculations. All MD simulations reported here were carried out with GROMACS package version 4.6.7 [40]. Lipid parameters were taken from the CHARMM36 force-field [33] and water was described by TIP3P [39]. The NPT ensemble was used and temperature kept at 310 K with the Bussi thermostat [42] and a coupling constant of 0.1 ps with two separate coupling groups (water and everything else). Pressure was kept at 1.0 bar with the Berendsen barostat [43] with a coupling constant of 1 ps and a compressibility of $1.0 \cdot 10^{-5} \text{ bar}^{-1}$ for the unrestrained simulations and of $0.5 \cdot 10^{-5} \text{ bar}^{-1}$ for the restrained free energy simulations. Semi-isotropic coupling in the direction normal to the bilayer was applied in the lipid simulations. Electrostatic interactions were handled by the particle mesh Ewald method [44] with a real space cutoff of 1.0 nm, grid spacing of 0.13 nm and quartic interpolation. All bonds were constrained using the LINCS algorithm [45] and van der Waals interactions were truncated at 1.0 nm. No dispersion corrections were applied in order to avoid artifacts in mean area per lipid and bilayer thickness [46]. The integration time step was set to 2 fs and the nonbonded list was updated every 20 fs.

Symmetric lipid bilayers containing different numbers of POPC and water molecules were built: 128 lipids and ~7900 water molecules (large system, LS), 72 lipids and ~6100 water molecules (medium system, MS) and 50 lipids and ~3300 water molecules (small system, SS). NaCl was added up to a final concentration of 0.150 M. After energy minimization using the conjugate gradient algorithm, an equilibration MD was run for 30 ns for the SS and up to 200 ns for the LS, resulting in a mean area per molecule of 0.63, 0.61 and 0.62 nm^2 respectively for the small, medium and large systems. This is in reasonable

agreement with the experimental value of 0.68 nm^2 [47]. The calculated bilayer thickness was 3.9 nm in all systems which is near the experimental value of 3.7 nm [47]. Mean area was computed as the ratio between the area of the membrane plane and the number of lipids per leaflet. Membrane thickness was calculated as the distance between the peaks of the lipid headgroup electron density.

Ubiquinone was inserted in the membranes using two different protocols: For the insertion of UQ1 in the LS bilayer, a pulling from the water phase towards the membrane center of mass (COM) was conducted in the membrane normal direction with the pull code in GROMACS and a rate of $4.10^{-5} \text{ nm ps}^{-1}$. For all other insertions, the protocol described by Javanainen was used [48]. In short, ubiquinone was placed at the membrane edge using a visualization tool, the box dimensions were increased accordingly, water and lipids were restrained in the membrane normal direction to avoid escaping to the empty space generated, and a 5 ns MD simulation was run with an anisotropic barostat and a reference pressure of 1000 bar in the direction which ubiquinone was placed. After ubiquinone was immersed into the membrane, the system was relaxed by removing all restrains and the pressure reversed back to semi-isotropic and 1.0 bar in another 30 ns MD run. The mean area and bilayer thickness were monitored to check for equilibration.

Ubiquinone concentration in percentage per mole was 0.8% in LS, 1.4% in MS and 2.0% in SS membranes. This is similar to previous ubiquinone simulations [22] and to the physiological concentration range of 1–3% [8].

Unconstrained MD simulations were performed for the embedded ubiquinones using the last frame from the umbrella sampling window with minimum free energy (see below). Trajectories with 200 ns were collected for UQ1, UQ1H2 and UQ2 and trajectories with 400 ns were collected for UQ6 and UQ10. Two simulations starting from different isoprenoid tail conformations (folded and fully extended) were performed for UQ10 with the quinone head positioned near the minimum free energy configuration found for the other ubiquinones. All properties calculated from these two simulations are equivalent and, hence, independent of the tail starting conformation.

Diffusion coefficients were calculated from the linear phase of the mean squared displacements with dimensionality set to 2, as only lateral diffusion along the membrane plane was processed [49]. The orientation of the ubiquinone head was calculated from the angle formed between the quinone ring plane (defined by atoms C1, C3 and C5) and the bilayer midplane. Flexibility and extension of the isoprenoid tail were evaluated from the distance between the last carbon of the tail (CT) and both the quinone head COM and the bilayer midplane. Contacts between ubiquinone atoms and different groups in POPC were defined with a 0.3 nm cut-off. Order parameters S_{CD} for lipid methylene units were calculated using the program LOOS [50].

2.3. Free energy calculations

Umbrella sampling (US) [51] and bias-exchange metadynamics (BEMD) [52] were used to compute the free energy profile for ubiquinone insertion into the model bilayers. The distance between the ubiquinone head COM and the membrane COM along its normal z -axis was used as the insertion reaction coordinate.

Initial configurations for each US window were extracted from a non-equilibrium trajectory generated with an extra pulling force acting over ubiquinone along the z -axis in the range [0,5] nm for the POPC bilayer (Section 2.2). The pulling spring constant was $300 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ and the pulling rate was $4.10^{-5} \text{ nm ps}^{-1}$. US windows were chosen equally spaced by 0.125 nm in the range $z = [0.00, 2.75] \text{ nm}$ while ubiquinone is inside the membrane or in the interface and by 0.250 nm in the range $z = [3.0, 5.0] \text{ nm}$ while ubiquinone is in water. The umbrella potential was set with $k_{umb} = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ in the membrane range and with $k_{umb} = 500 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ in the water range.

Simulations run for 60 ns in each US window with a total aggregate time of 5.5 μ s for all US simulations. The reaction coordinate was recorded each 200 fs. Potentials of mean-force (PMF) were calculated from the reaction coordinate occurrence using the weighted histogram analysis method [53]. The PMFs shown were obtained from 30 ns of equilibration and 30 ns of accumulation in each window. PMF uncertainties were estimated by bootstrap analysis using the Bayesian method [54]. It should be noted that ubiquinone insertion in the US simulations was carried out over only one bilayer leaflet (z -axis > 0).

The bias-exchange metadynamics method [52] was applied to enhance sampling along degrees of freedom orthogonal to the insertion reaction coordinate. BEMD is an extension of metadynamics [55] for replica exchange in a collective variable (CV) space. The PLUMED plug-in version 2.1 [56] was used for these simulations.

The following CVs were chosen besides the z -axis coordinate: C6–C7 bond torsion which was identified with a high torsional barrier (Fig. 3); the angle formed between ubiquinone atoms C1 and C4 and the COM of the membrane (θ_{UQ}). Following a recent publication that studied the insertion of ethanol in a model bilayer with BEMD simulations [24], coordination numbers N between ubiquinone and membrane lipid tails or water molecules were also included as CVs. These contacts were calculated with the soft function:

$$N_{g_1-g_2} = \sum_{i \in g_1} \sum_{j \in g_2} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \quad (2)$$

where r_{ij} is the distance between atoms i and j and the sums run over two sets of atoms (g_1 and g_2). Values of $r_0 = 0.3$ nm, $m = 12$, and $n = 6$ were adopted. Four contact number CVs ($N_{g_1-g_2}$) were defined by the combination of $g_1 = \{H, I\}$ and $g_2 = \{A, W\}$, where H indicates the set of ubiquinone head atoms, I is the set of isoprenoid tail atoms, A is the set of lipid acyl chain carbon atoms and W is the set of water oxygen atoms.

For UQ1, UQ1H2 and UQ2, BEMD simulations were performed for 120 ns with the 7 CVs defined in the previous paragraph and an extra replica without metadynamics boost. For UQ0, two BEMD simulations were performed with a different combination of CVs. One simulation was performed for 160 ns with 4 CVs: z -axis, θ_{UQ} , N_{H-A} and N_{H-W} , and another was done for 120 ns with 4 CVs: z -axis, z -axis with a slower gaussian function deposition, and torsions around C2–O2 and C3–O3 bonds. All BEMD simulations were performed in SS bilayers (Section 2.2). For UQ1, a BEMD simulation was repeated with the MS bilayer. The gaussian height for all CVs was set to 0.5 kJ mol⁻¹ and gaussian widths were set to 4 nm for the coordination number CVs and to 0.2 nm for all other CVs. The frequency of gaussian deposition was set to 4 ps for the z -axis CV, 10 ps for the coordination number CVs and 20 ps for all other CVs (including the z -axis with slower deposition). Exchanges were attempted randomly among replicas every 20 ps.

PMFs were obtained from the BEMD simulations by averaging N_{BE} profiles generated at different accumulation times from the history dependent potential of the sum of gaussian functions deposited for a given CV [57,58]. Typically, 10 to 12 profiles were generated, the first 2 or 3 with the shortest accumulation times were discarded and $N_{BE} \sim 8$ –10 profiles were averaged. For total simulation times longer than about 80 ns, the system could freely diffuse along the enhanced CV. Uncertainties were estimated from the standard deviation of the mean [57].

Standard binding free energies between ubiquinone and the lipid membrane, ΔG_b^* , were determined by trapezoidal integration of the PMF profiles according to [59]:

$$e^{-\beta \Delta G_b^*} = \frac{\int_{\text{bound}} e^{-\beta |PMF(z)|} dz}{\int_{\text{unbound}} e^{-\beta |PMF(z)|} dz} \quad (3)$$

A separate ΔG_b^* was calculated for each leaflet in the BEMD profiles with bound regions set to $z \in [0.0, 2.5]$ nm or $z \in [-2.5, 0.0]$ and unbound region set to $z \in [2.5, 5.0]$ nm or $z \in [-5.0, -2.5]$ nm, respectively. Results shown are an average of the two separate ΔG_b^* . For the PMFs obtained from US, the bound region was set to $z < 2.6$ nm, 2.9 nm and 4.0 nm for UQ1, UQ2 and UQ6, respectively. These boundaries correspond to the z value where the PMF(z) reached a plateau. The unbound region was set such that the same integration volume was used, resulting in the calculation of a standard free energy. Uncertainties were propagated from the PMFs following standard formulas [60].

3. Results & discussion

3.1. Performance of the new CHARMM compatible parametrization

As described in Section (2.1), partial charges, Lennard-Jones parameters (R_{min}) and dihedral force constants of a CHARMM compatible force-field were adjusted in order to reproduce high-level QC data for bond torsion potentials, molecular dipoles and interaction energies of fragment molecules ubiquinone-1, ubiquinol-1 and isoprene isolated and in complex with water or apolar molecules.

The description of molecular and energetic properties was greatly improved with the force-field parametrized here, both qualitatively for the shape of bond torsion potentials or the description of hydrogen-bonded complexes, as well as quantitatively for the relative stability between rotamers or complexation energies. In particular, it was noted that the molecular dipole of the apolar isoprene molecule (ISP) was incorrectly described by up to 2 Debye using the previously available CHARMM compatible force-fields (Table S4) [26,28]. This was corrected in the current parametrization which leads to significant improvements on the bilayer partition free energy of ubiquinones.

A few examples are noted below to illustrate the performance of the available force-fields in comparison to high-level QC calculations. Please refer to S1 (Figs. S1 and S2) for a complete set of results for interaction and torsion profiles used for training and testing the force-fields. This analysis may help to identify flaws in the potential energy description and suggest internal degrees of freedom which should be sampled carefully in condensed-phase simulations.

For the interaction energy profiles used in the parametrization training set (Figs. 2A and S1), deviations of the ubiquinone force-field obtained here from the QC reference are always smaller than 5 kJ/mol near minima and over longer separation distances. Differences are at most 10 kJ/mol for interaction energy profiles used only for testing such as the one in Fig. 2B. Other force-field parametrizations [28,26] show much higher deviations and even qualitatively wrong descriptions such as in a hydrogen bond between water and the quinol hydrogen (Fig. 2B).

The shapes of the potential energy profile and relative energies between minima are well described by the force-field obtained here for ubiquinone bond torsions either included or not in the parametrization (Figs. 3 and S2). Deviations smaller than 10 kJ/mol are observed near energy barriers. On the other hand, previously available MM parametrizations result in torsions with incorrect shapes (e.g., in Fig. 3A) and in high deviations near energy barriers (Fig. 3B and C).

Torsions in the isoprenoid tail around bond C(4 + 5i)–C(6 + 5i) in the i -th isoprenoid unit were also improved (Fig. 3A) leading to a more accurate description of ubiquinone homologs with long isoprenoid tails. The torsion around the C6–C7 bond was particularly challenging to be described with the available force-fields (Fig. 3B). During the parametrization of the dihedral force constants, it was noted that the barriers along this bond torsion could not be adjusted due to too high repulsion interactions between atoms C8 and C5 and their respective hydrogens. Thus, Lennard-Jones parameters for both carbon atoms were changed and a significant improvement (~ 20 kJ/mol) was obtained. Nevertheless, the force-field parametrized here is still unable to describe the barrier near dihedral angle $\sim 270^\circ$ with an accuracy

better than 15 kJ/mol. In fact, this is the highest deviation observed over all potential energy comparisons done here. Because the C6–C7 bond torsion contributes significantly for the ubiquinone flexibility as it connects the quinone head with the isoprenoid tail, sampling along this torsion was carefully monitored in the insertion free energy calculations.

Methoxide group torsion around bond C2–O2 (or C3–O3) also shows high barriers when described with previous parametrizations (see Fig. 3C for UQ1H2 and Fig. S2 for UQ1). The barrier deviations were also due to too high repulsion interaction between CM2 (or CM3) and O1 (or O4) and were again corrected with changes in the Lennard-Jones R_{min} parameter. The deviations were greatly improved with the force-field parametrization presented here.

3.2. Free energy profiles for partition into the membrane

Potentials of mean-force (PMFs) for the insertion of various ubiquinone homologs into model POPC bilayers were calculated in order to check the accuracy of the parametrized force-fields and to determine the equilibrium ubiquinone localization along the membrane normal. Umbrella sampling and bias-exchange metadynamics were used to enhance sampling over the membrane insertion coordinate (z -axis defined above) and along orthogonal collective coordinates in the case of BEMD. In this section, we describe the PMFs and derived free energy quantities (ΔG_b^i). Details of the ubiquinone localization and contacts within the bilayer are left to Section 3.3 below.

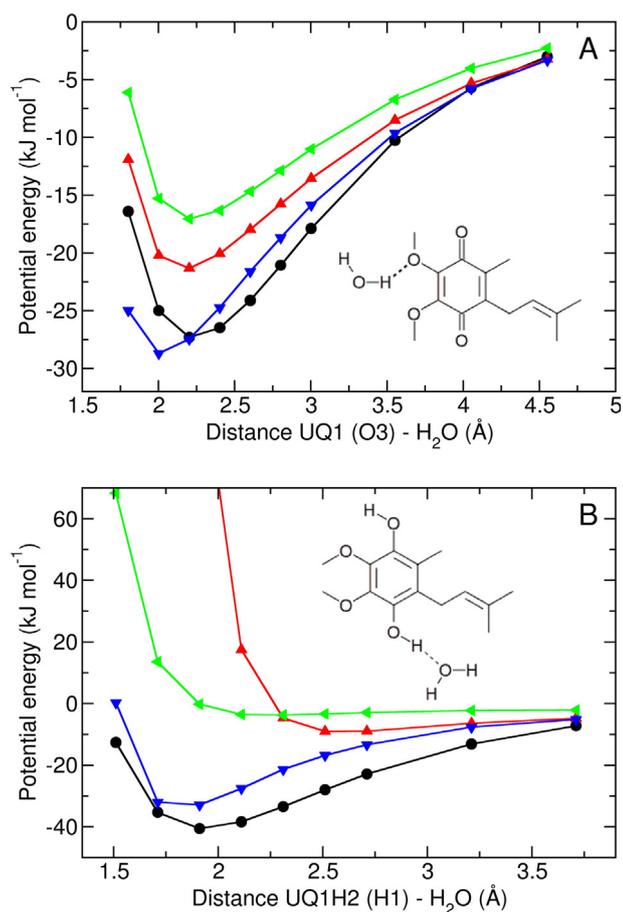


Fig. 2. Interaction energies for ubiquinone–water hydrogen-bonded complexes. QC reference energies were obtained at the MP2/6-311+G(2df,2p) level (black circles). The other curves are different ubiquinone MM parametrizations proposed by Kaszuba et al. [28] (red up triangles), Autenrieth et al. [26] (green left triangles) and here (blue down triangles). Complex orientation is shown as insets. Data in panel A was included in the force-field parametrization while panel B was used only for testing purposes.

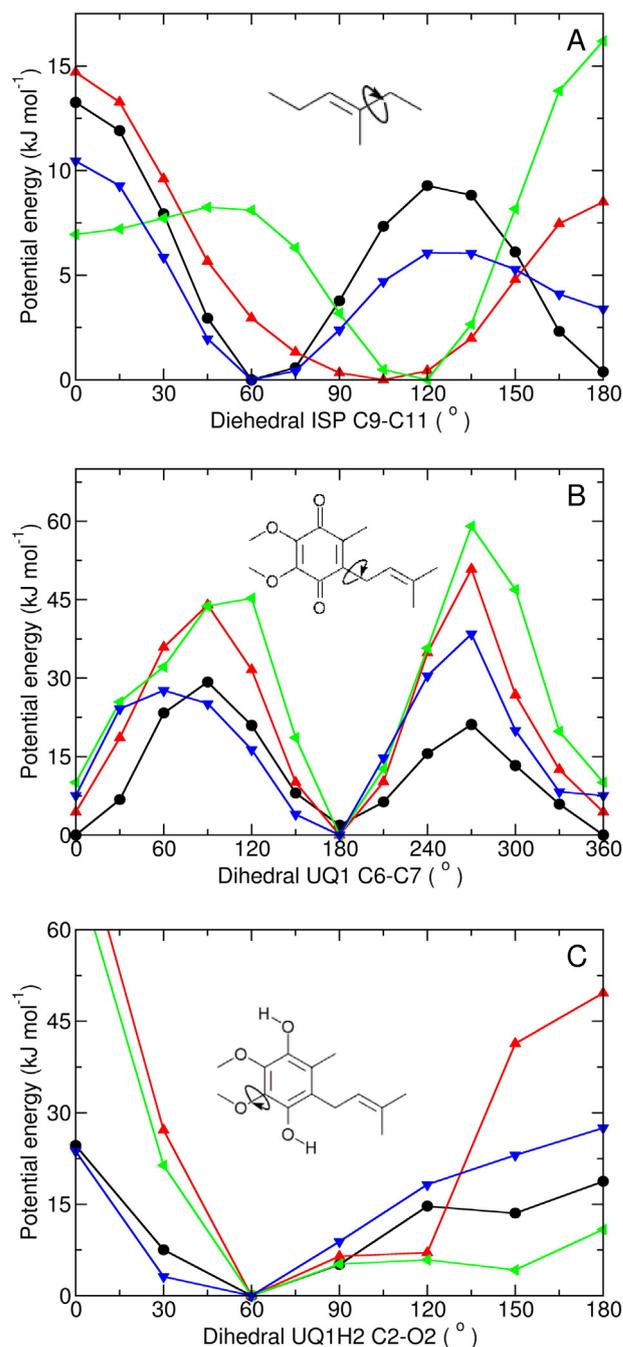


Fig. 3. Bond torsion potentials for ubiquinone obtained at QC and MM levels as indicated in the legend for Fig. 2. Data from panels A and B were included in the force-field parametrization while panel C was used only for testing.

Fig. 4 shows that the general shape and minima of the PMF obtained by US change little for the three ubiquinones tested. One isoprenoid unit is enough to partition ubiquinone UQ1 into the bilayer and, as expected, increasing the number of isoprenoid units enhances hydrophobicity and consequently membrane affinity. But, the z -axis position of the PMF minima is rather constant (1.68 nm for UQ1 to 1.55 nm for UQ6). These minima correspond to the preferential localization of the ubiquinone head in the membrane and have been observed in the same z -axis range for the insertion of other amphiphile molecules [23]. Thus, previous proposals of the ubiquinone head localization in the middle of the bilayer are not plausible [6–10,22].

The translocation of ubiquinone from one bilayer leaflet to the other will depend on the free energy barrier separating the

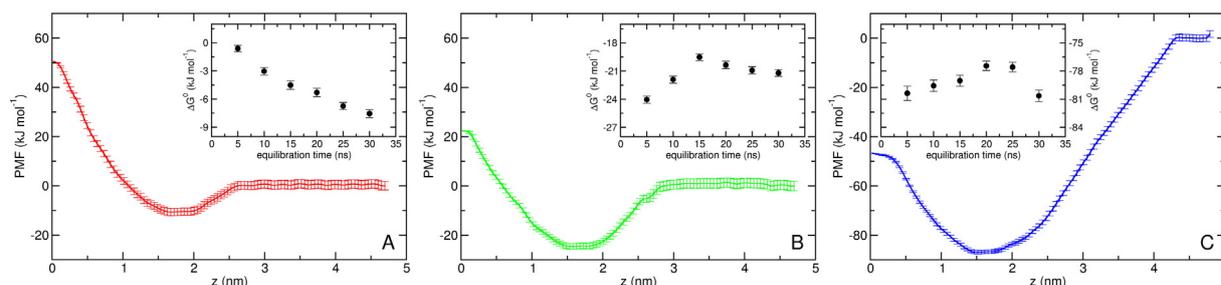


Fig. 4. PMFs for ubiquinone insertion in large (LS) POPC bilayer obtained by umbrella sampling for UQ1 (panel A), UQ2 (B), and UQ6 (C). The insets show the calculated binding free energy ΔG_b as a function of equilibration time for fixed 30 ns of accumulation time in each US window.

membrane center and the minima in the PMF. The barrier decreases from $62 \pm 2 \text{ kJ mol}^{-1}$ for UQ1, to $47 \pm 2 \text{ kJ mol}^{-1}$ for UQ2 and to $40 \pm 2 \text{ kJ mol}^{-1}$ for UQ6. This is again in line with the higher hydrophobicity of ubiquinones with longer tails and suggests that an approximate barrier of $\sim 40 \text{ kJ mol}^{-1}$ should be observed for the “flip-flop” of natural ubiquinones with long isoprenoid tails (UQ6–UQ10). Thus, the translocation of ubiquinone is a rare event and will not be observed faster than the micro- to milli-second time scale. This is in agreement with the flip-flop rate estimated from NMR measurements [11]. There is no significant barrier in the PMFs for the water-membrane insertion process, as also observed for the insertion of small polar and charged molecules [61,62,24] or large amphiphiles [23], suggesting that ubiquinone insertion will be diffusion-controlled. It should be noted that the lack of an insertion barrier may also be due to incomplete sampling for the ubiquinones with longer isoprenoid tails.

Although we expect to have obtained converged PMFs for UQ1 and UQ2 (see discussion below), the same cannot be said for UQ6 which has a longer tail and would require longer simulation times for convergence. It has been recently noted that the calculation of PMFs in the membrane–water interface region for the insertion of amphiphiles with long hydrophobic tails can be problematic [23]. Although errors

in this region of the PMFs result in incorrect calculated binding free energies, they do not effect significantly the z -axis location of PMF minima nor the translocation barrier which should be well described by the PMFs shown in Fig. 4 even for UQ6. Thus, our discussion in the previous paragraph is accurate, but the value estimated below for the free energy of binding of UQ6 (Table 1) should only be considered qualitatively.

It has been noted [63,61,62,23] that the solvation substitution occurring in membrane–water interfaces is often the process with the slowest relaxation to be described during the simulation of solute insertion into lipid bilayers. In accordance, we have noticed a sudden change in the hydration number of the ubiquinone tail for homologs with two or more isoprenoid units in US windows located at the membrane–water interface. For UQ2, it was observed at $z \sim 2.8 \text{ nm}$ (Fig. S3), corresponding to simulation windows where the isoprenoid tail does not transition between water and bilayer and stays in only one of these phases during all accumulation time. This sudden change may also be causing the small flat region observed at similar z coordinate on the calculated PMF for UQ2 (Fig. 4) [23].

To enhance sampling along slow degrees of freedom such as the number of contacts of ubiquinone with water and with the lipid we

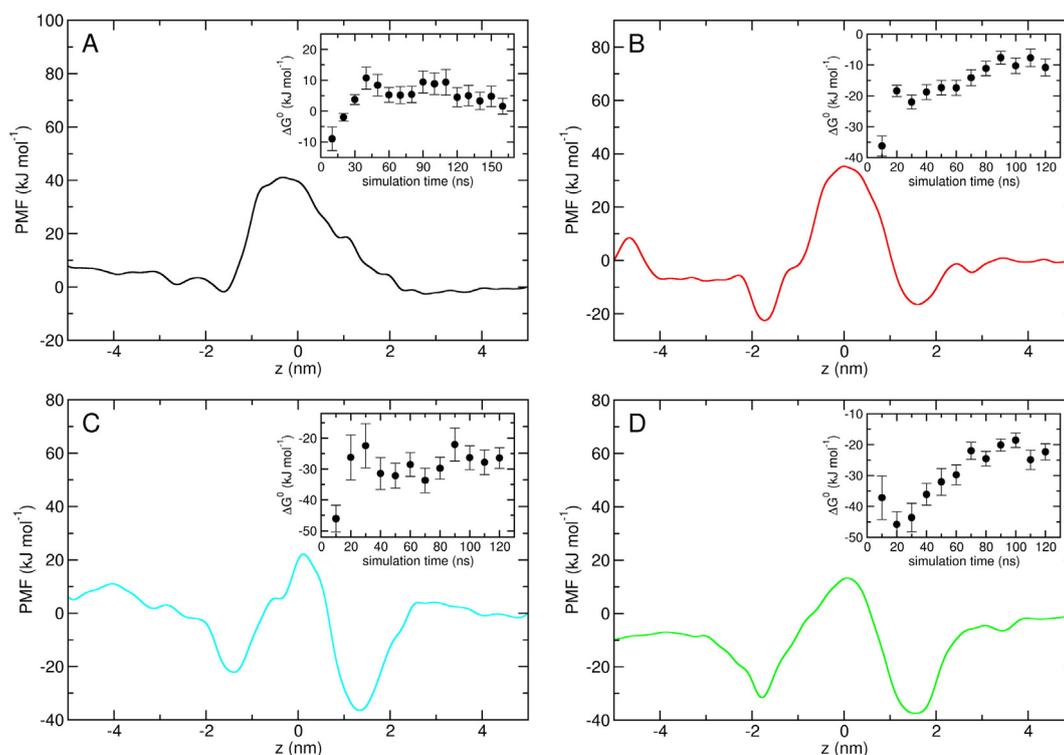


Fig. 5. PMFs for ubiquinone insertion in small (SS) POPC bilayer obtained by bias-exchange metadynamics for UQ0 (panel A), UQ1 (B), UQ1H2 (C), and UQ2 (D). The insets show the calculated binding free energy ΔG_b as a function of total simulation time.

Table 1
Experimental and calculated standard binding free energies (ΔG_b , in kJ mol^{-1}) between ubiquinone homologs and model bilayers.^a

	Experimental	US	BEMD
UQ0	−2[65]		2 ± 2
UQ1	−17[66], −20[18]	−8 ± 1	−11 ± 3
UQ1H2	−18[67]		−26 ± 3
UQ2	−21[18], −24[66]	−21 ± 1	−22 ± 3
UQ6		−81 ± 1	

^a Experimental free energies were obtained from partition coefficients measured for bovine heart submitochondrial particles [66], asolectin vesicles [18,67] and water–octanol [65] mixtures assuming a temperature of 310 K. Values calculated in POPC bilayers were obtained from US and BEMD simulations employing the force-field parametrized here. Uncertainties were propagated from the corresponding PMFs.

have employed BEMD simulations. Internal bond torsions of the inserted molecule can also difficult sampling [25]. Thus, we have enhanced sampling over ubiquinone tail C6–C7 which was identified with large energy barrier (Fig. 3).

PMFs shown in Fig. 5 were obtained from the BEMD simulations for insertion over the two bilayer leaflets. The PMF shape is more rugged than the US PMFs as a result of the gaussian deposition scheme that enhances sampling along the membrane normal z-axis [24]. Nevertheless, for UQ1 and UQ2 the position of the PMF minima (~ 1.65 nm in modulo) and the translocation energy barrier (55 ± 10 kJ mol^{-1} for UQ1 and 47 ± 10 kJ mol^{-1} for UQ2) are equivalent to those observed in the US PMFs. It should be noted that uncertainties reported for free energies obtained by BEMD are higher than those reported for US due to higher standard deviations obtained in the averaging procedure used to calculate the PMFs from BEMD (see Materials & methods). This scheme is formally different and more likely to accumulate statistical errors than the bootstrap analysis used to estimate the uncertainties of PMFs from US.

The minima observed in the PMF for UQ0 insertion were not pronounced in either the simulation shown in Fig. 5 or in a simulation with enhanced sampling for C2–O2 and C3–O3 bond torsions (data not shown) suggesting that our UQ0 model gives a very low affinity for the membrane. Ubiquinol UQ1H2 insertion PMF and localization in the membrane are similar to the other quinone homologs. This is in agreement with their comparable experimental reactivity [64] and free energy of binding (Table 1).

The accuracy of the PMFs simulated here for ubiquinone insertion into a POPC bilayer can be accessed by comparisons of the calculated free energy of binding (ΔG_b , Eq. (2)) with free energies derived from experimental partition coefficients as shown in Table 1. It should be noted that the available experimental data were obtained in conditions and system composition which are different from those simulated here. For instance, the partition coefficient reported for UQ0 was obtained in a water–octanol mixture [65]. For the other ubiquinones, partition coefficients were obtained in bovine heart submitochondrial particles with a mixed lipid composition [66], or in sonicated asolectin vesicles mainly composed by phosphatidylcholine [67]. Thus, part of the divergence between experimental and simulated values may be attributed to differences in the model composition.

The highest differences from experiment in the binding free energy calculations are observed for UQ1 and UQ1H2 in Table 1. For the quinone, the calculated ΔG_b suggests an affinity of 6–12 kJ mol^{-1} lower than the experimental values. For the quinol, an affinity of 8 kJ mol^{-1} higher than experiment is found. These differences may be due to unbalanced interactions of the quinone (or quinol) head with the lipid polar group. Values calculated for UQ0 and UQ2 are almost equivalent to the experimental measurements. No experimental value for UQ6 partition coefficient is available, but the binding free energy for biological ubiquinones was estimated to be more negative than -60 kJ mol^{-1} [66]. As noted above, we do not expect to have a quantitative value for UQ6 binding free energy. Thus, given the statistical uncertainty in the simulations, the composition differences between

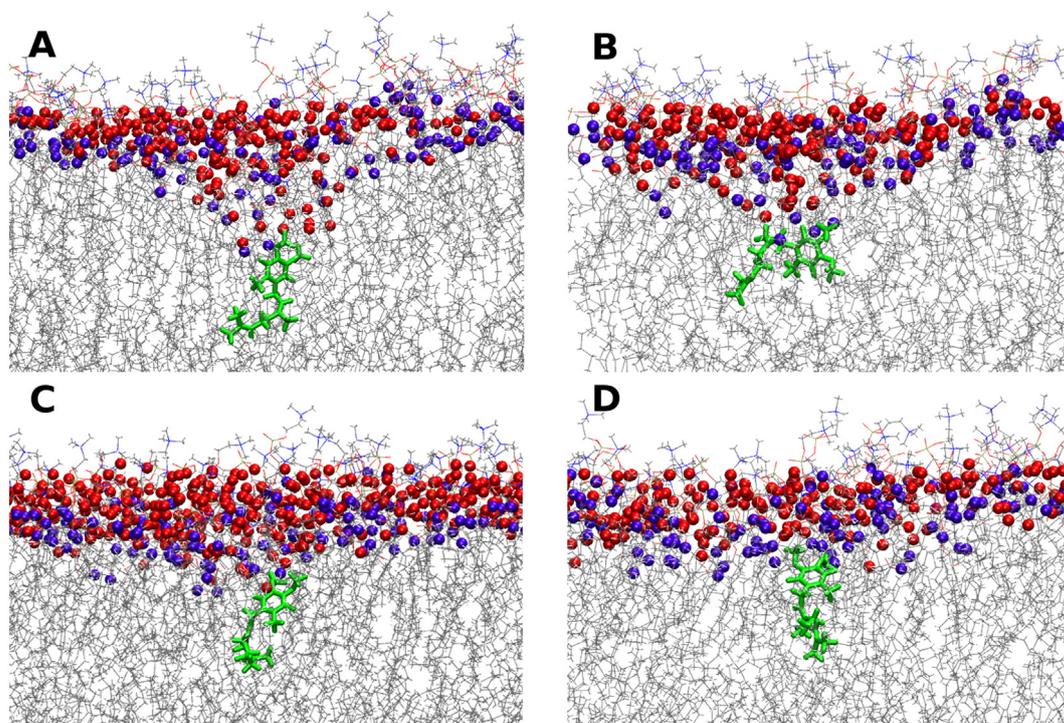


Fig. 6. Membrane protrusion upon ubiquinone insertion. Snapshots were obtained from the last frames of UQ2 simulation in US windows centered at $z = 0.2$ nm (panel A), $z = 0.4$ nm (B), $z = 0.6$ nm (C) and $z = 0.9$ nm (D). Ubiquinone is represented in green sticks. Membrane is represented in lines with lipid glycerol oxygen atoms in violet balls and interfacial water oxygens in red balls.

the simulated models and experimental setups, and the variations of measured values, we conclude that our force-field yields PMFs and derived free energy quantities in good agreement with experimental observations.

The agreement of ΔG_b calculated for UQ1 and UQ2 between US and BEMD methods also suggests a good accuracy for the calculated PMFs. These two methods estimate free energies using different assumptions and the simulations were carried out independently in different bilayer preparations (LS for US and SS for BEMD). The influence of the bilayer size on the binding free energy calculated here from the BEMD simulation should be negligible. For instance, $\Delta G_b = -10 \pm 3 \text{ kJ mol}^{-1}$ for UQ1 insertion in the MS bilayer is equivalent to the value calculated for the SS bilayer (Table 1).

On the other hand, the force-field description changes the calculated binding free energy considerably. For the insertion of UQ1 in the LS membrane using US and the force-field proposed by Kaszuba et al. [28], the calculated $\Delta G_b = 1 \pm 1 \text{ kJ mol}^{-1}$ is 18 kJ mol^{-1} higher than the experimental value. Such high hydrophilicity of the Kaszuba et al. potential can be attributed to the incorrect high polarity of the isoprenoid tail observed with their charge parametrization (Table S4). This artifact will be more pronounced for ubiquinones with longer isoprenoid tails as the total dipole for the longer tails will be a vector sum of the contributions of each isoprenoid unit.

Convergence of the calculated PMFs can be accessed from the derived free energies of binding obtained over different equilibration (US) or total simulation (BEMD) times as shown in the insets of Figs. 4 and 5. For the US simulations, the calculated ΔG_b shows variations smaller than $\sim 3 \text{ kJ mol}^{-1}$ after 15 ns of equilibration time in each window. For the BEMD simulations, ΔG_b shows variations smaller than $\sim 6 \text{ kJ mol}^{-1}$ after 80 ns of total simulation time. Both variations are comparable to the uncertainties estimated from the PMFs. Higher precision would require much longer simulation times [61]. However, we do not judge a higher precision necessary as the variations of reference partition coefficients measured for same ubiquinone in different experimental preparations are similar to the present calculated statistical uncertainties ($\sim 3 \text{ kJ mol}^{-1}$).

Besides essential sampling over the bilayer insertion coordinate (z -axis), we identify that enhanced sampling over the coordination numbers between water or lipid molecules and the ubiquinone head and tail (N_{g1-g2} in Eq. (1)) are the most important ones in order to describe the slow relaxation of the solvation substitution process in the membrane interface. Thus, convergence of calculated PMFs in BEMD simulations can be significantly accelerated for ubiquinones with longer isoprenoid tails by enhanced sampling of the coordination between water and tail. Enhanced sampling of ubiquinone orientation and internal degrees of freedom are less important. For instance, a BEMD simulation with enhanced sampling of methoxide bonds (C2–O2 and C3–O3) resulted in a $\Delta G_b = 1 \pm 1 \text{ kJ mol}^{-1}$ for UQ0 which is equivalent to the binding free energy calculated for UQ0 without such enhancement (Table 1). Rotations through these methoxide bonds as well as in quinol hydroxide bonds (H1–O1 and H4–O4) are observed over the simulation time of the US and BEMD simulations (Fig. S5). It should be noted that rotations over C6–C7 bonds were not observed during the accumulation time of the US windows, but were observed on the longer free MD simulations (see Section 3.3 and the discussion of Fig. 8). However, increasing the sampling of this C6–C7 bond as done in the BEMD simulations does not lead to significant differences in the calculated PMF and derived quantities.

3.3. Ubiquinone-membrane structure and interactions

Ubiquinone insertion induced a structural perturbation in the bilayer when the ubiquinone polar head group is located near the membrane midplane (ubiquinone head COM with $z < 0.6 \text{ nm}$) as shown in Fig. 6. This protrusion has a “funnel”-like shape and corresponds to both lipid head groups and water molecules dragged towards the membrane

center. The number of contacts with water also reports the significant hydration of the ubiquinone head when partitioned inside the membrane (Fig. S3). The opposite effect when ubiquinone desorbs from the bilayer was not observed [61], suggesting that the solvation substitution in the bilayer interface is more favorable than membrane deformation.

Results presented over the remaining of this and the next sections (Figs. 7–12) were obtained from unconstrained MDs started from the last frame of the 60 ns US window with lowest free energy value in the respective PMF. This should correspond to the equilibrium configuration of ubiquinone embedded in the bilayer. Isoprenoid tails should be well equilibrated as restrictions in US were included only in the ubiquinone head. The last 180 ns for UQ1–UQ2 and the last 360 ns for UQ6 and UQ10 of the unconstrained MD trajectories are used for analysis.

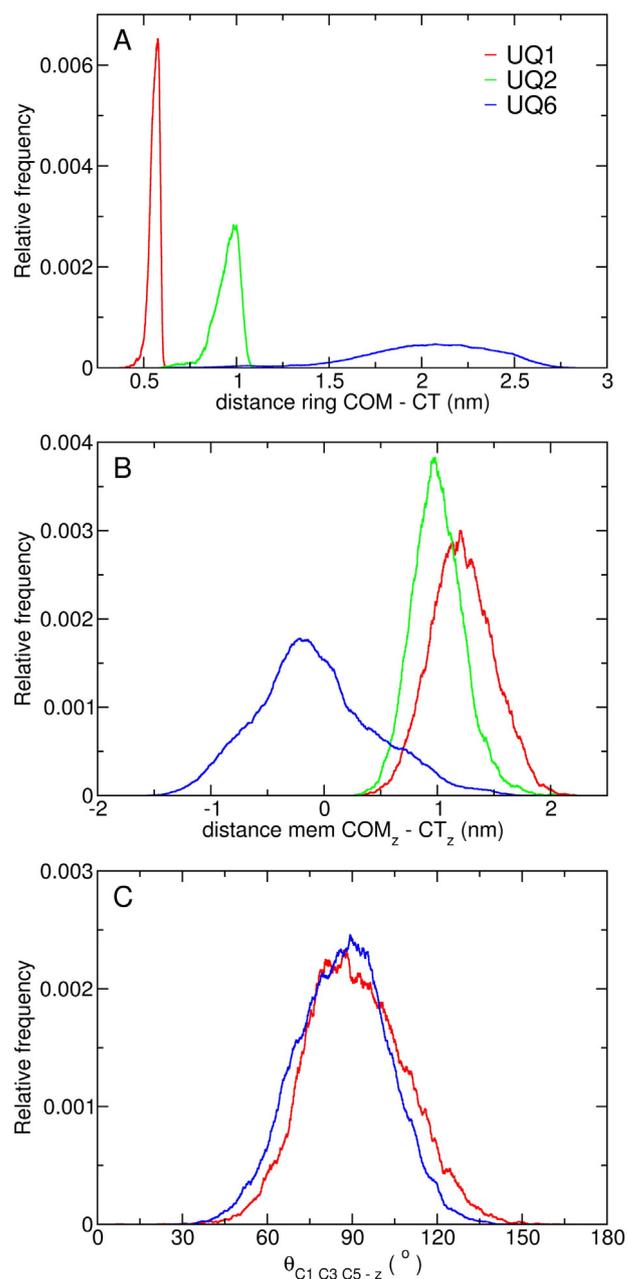


Fig. 7. Ubiquinone localization in the bilayer obtained from unconstrained MD simulations. The distance between ring COM and the last carbon (CT) of ubiquinone isoprenoid tail (panel A), the distance between CT and the membrane COM (B) and the angle between ubiquinone head plane and the membrane midplane (C) for UQ1 (red), UQ2 (green) and UQ6 (blue) are shown. UQ2 curve is not shown in panel C as it is equivalent to UQ1 and would difficult its visualization.

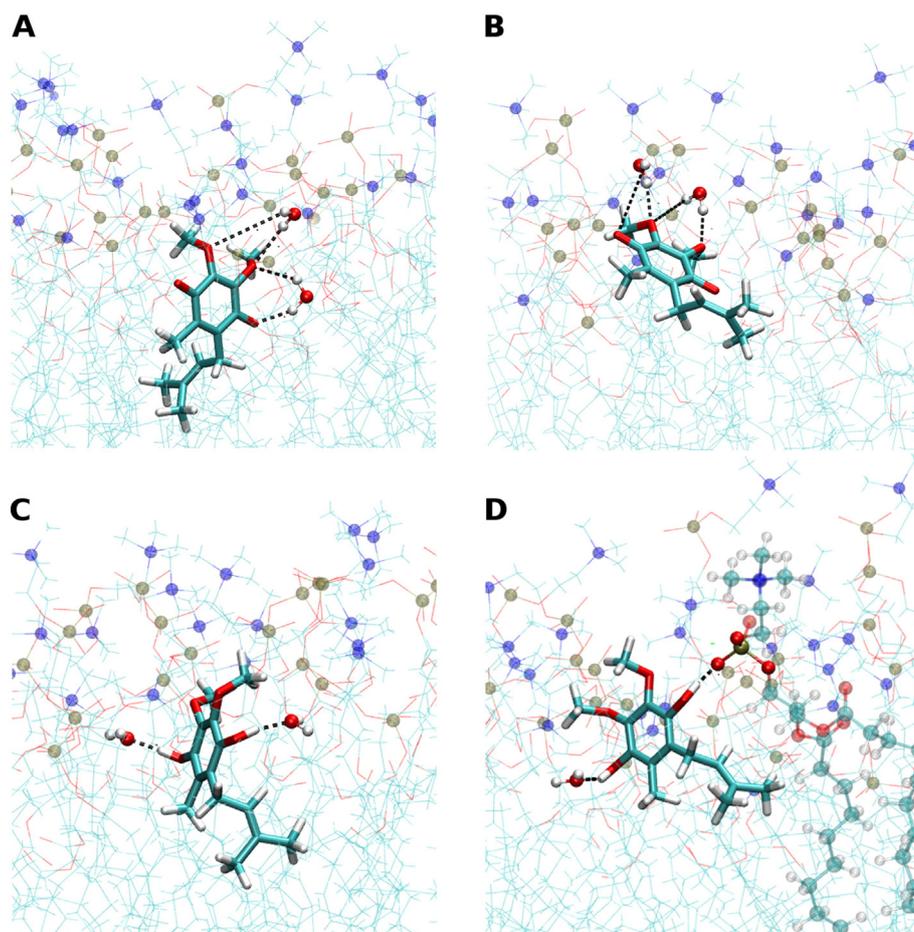


Fig. 8. Selected interactions of ubiquinone with the membrane environment. Ubiquinone is represented in sticks and the lipid bilayer is represented in lines and balls with phosphate P in ochre and choline N in blue. Water molecules and atoms of lipid that interact with UQ1 and UQ1H2 are depicted in balls and sticks, and hydrogen bonds are indicated in black dashed line. Carbon atoms are colored cyan, H in white and O in red. Panels A and B show bidentate water–ubiquinone oxygen hydrogen bonds. Hydrogen bonds of ubiquinol hydrogen with water are show in panel C and with POPC phosphate is shown in panel D.

Considerable effort has been made towards describing ubiquinone localization in lipid bilayers. The two main proposals in the literature suggest that ubiquinone head lies in the bilayer midplane, oriented parallel to membrane plane [6–10] or that the head group is localized near the water–bilayer interface close to the glycerol average position [12–15]. Our results strongly support the second model as the minima observed in all PMFs calculated here corresponds to $z \approx 1.6$ nm. The z insertion coordinate is equivalent to the distance between the membrane midplane and the COM of the ubiquinone head. Thus, contrary to previous suggestions [16], the localization of the ubiquinone head does not change significantly with the length of the isoprenoid tail.

It is remarkable that the average positions of ubiquinone ring in cytochrome bc_1 (complex III) for both the Q_o and Q_i redox sites are observed around the same z -axis values (± 1.6 nm) [68,69]. The entrance of the narrow ubiquinone chamber in NADH:ubiquinone reductase (complex I) is also located near the membrane interface at an approximately similar z -axis value [70]. Thus, the equilibrium location of ubiquinones in a bilayer matches the position of protein binding sites in respiratory complexes, probably facilitating the binding mechanism and increasing the rate of protein binding and unbinding.

The sharp monotonic distribution of the distance between the last isoprenoid carbon (CT) and the COM of the ubiquinone head shown in Fig. 7A for UQ1 and UQ2 suggests a reduced internal flexibility for the tails of ubiquinones with few isoprenoid units. The UQ6 tail, however, has high internal conformational flexibility and a broad distribution centered at 2 nm. A similar broad distribution is observed for the

UQ10 terminal isoprenoid carbon (data not shown). The UQ6 tail distribution is altered when ubiquinone is moved inside the membrane (Fig. S4) suggesting that a tail rearrangement is observed during the ubiquinone flip-flop pathway.

Fig. 7B shows the localization of CT regarding membrane normal. The isoprenoid tail is mostly extended and in contact with lipid acyl chains up to about the sixth isoprenoid unit. Given the polar head localization discussed above, ubiquinones span the membrane similar to a POPC molecule. For UQ6 and longer ubiquinones, the tail length is longer than the POPC acyl chain. CT interdigitates over the two bilayer leaflets and it is preferentially localized in the leaflet opposed to its head. For UQ10, the four terminal isoprenoid units have increasingly higher flexibility. In fact, the terminal CT in UQ10 equally samples the whole apolar region of both membrane leaflets (Fig. S4).

This is contrary to previous suggestions that the isoprenoid tail would fold over itself [22,19] and suggests that no aggregation or clustering of ubiquinones inside the lipid bilayer should be observed in the concentration range studied here ($\sim 2\%$ per mol) [66]. It should be noted that the distribution of z -axis position of the terminal carbon in POPC (both chains, data not shown) is about an order of magnitude less broad than observed here for ubiquinone CT. The higher flexibility in the ubiquinone tail might be related to its higher diffusion rates (Section 3.4) [19] and may also facilitate binding into the narrow NADH:ubiquinone reductase ubiquinone chamber.

The equilibrium ubiquinone head orientation is indicated in Fig. 7C. For all ubiquinones studied (UQ1 up to UQ10), the quinone head is

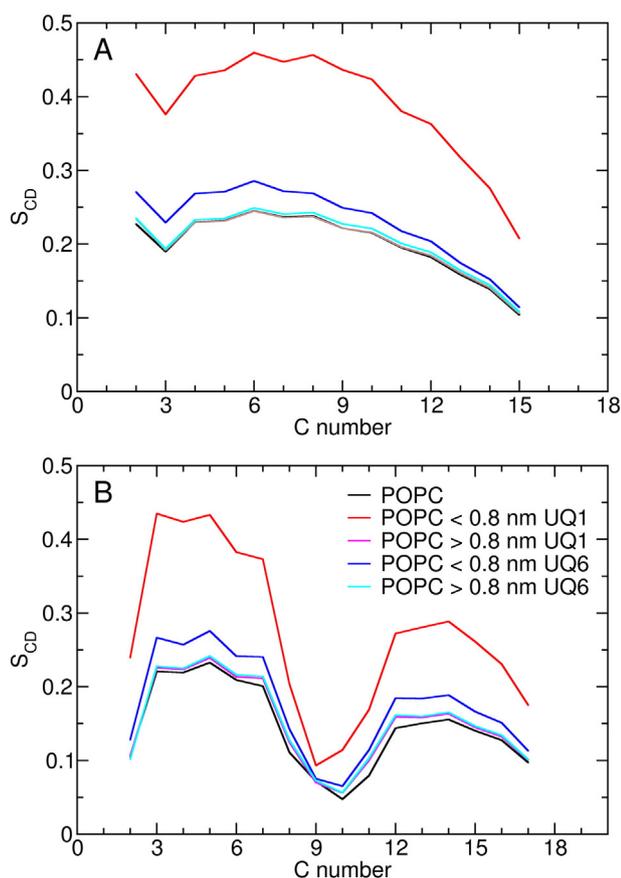


Fig. 9. Calculated carbon-hydrogen bond vector order parameters for POPC sn-1 (panel A) and for sn-2 (B) acyl chains. Order parameters were calculated for a pure POPC bilayer (black lines) and in the presence of ubiquinone separately for lipid molecules within or further away than 0.8 nm of UQ1 or UQ6 as colored in the legend.

oriented normal to the membrane, forming an angle of $\sim 90^\circ$ with the midplane. The angle distribution is rather sharp, with fluctuations smaller than 30° in the time scale of the unconstrained MD simulations. Atoms C5 and C6 of the ubiquinone ring point to the center of the bilayer and atoms C1–C4 point towards the solution phase as expected from the more hydrophilic groups attached to the last centers (Fig. 8). Thus, the ubiquinone orientation when bound to the membrane does not change with increased isoprenoid tail length. When the quinone head is inserted into the low-packing bilayer center ($z \sim 0$ nm) as well as when it is free in solution ($z > 4.0$ nm) during the US simulations the whole orientation space is sampled and ubiquinone head tumbles almost freely (Fig. S4).

Since ubiquinones have been suggested to order lipid membranes [9], we have computed carbon-hydrogen order parameters S_{CD} for sn-1 and sn-2 chains of POPC for simulations of ubiquinone-free bilayers and for bilayers containing ubiquinones of different isoprenoid chain length. While average S_{CD} order parameters show small shifts for ubiquinone containing bilayers, there was a significant change in order parameters of POPC molecules approximately in the first ubiquinone lipid shell (within 0.7–0.8 nm) as shown in Fig. 9. Acyl chain ordering upon ubiquinone addition observed in fluorescence anisotropy measurements [9] was more pronounced for short tail homologs. This is again in agreement with our simulation results.

Ubiquinone and ubiquinol oxygens were highly hydrated when partitioned into the membrane as shown in Fig. 10. This is due to the interfacial localization of the quinone head and to the bilayer protrusion (Fig. 6). Ubiquinone oxygens, both ketonic and methoxyl (Figs. 10A and S6A) have a first solvation layer formed by water and lipid choline groups. Glycerols have also significant contribution to the first

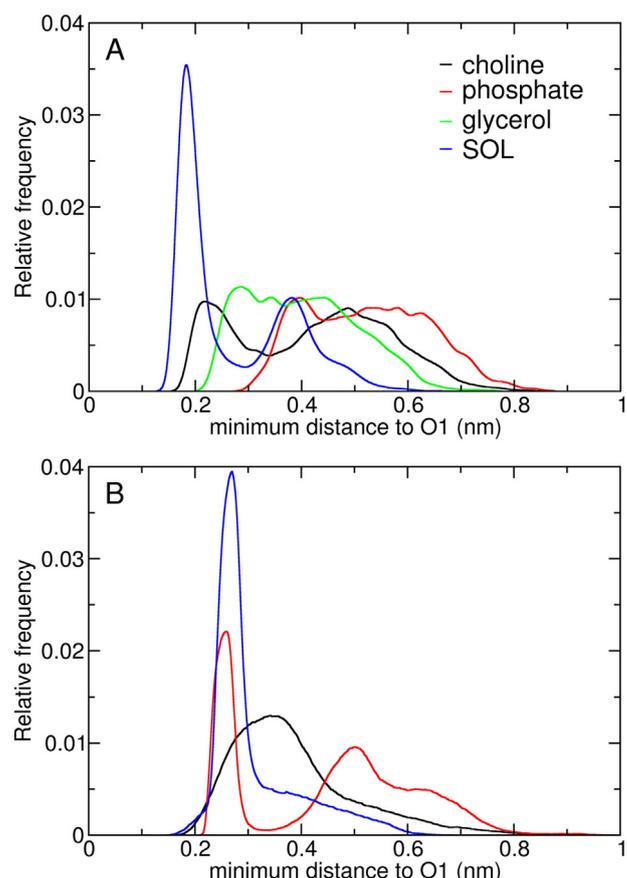


Fig. 10. Minimum distance distribution. Calculations were performed over unconstrained MD simulations between different lipid or solvent groups as shown in the legend and O1 in UQ1 (panel A) and O1 in UQ1H2 (B). SOL is the water solvent.

interaction layer, while phosphate groups were farther away. This is again in agreement with previous suggestions from experimental studies [12–15].

Ubiquinol showed a very similar interaction pattern for methoxyl oxygens (Fig. S6B), while hydroxyl oxygens show a perturbed pattern in comparison to the quinone oxygens (Fig. 10B). These hydroxyl oxygens have sharper interactions with the water solvent and with the lipid phosphate group because of the donation of hydrogen bonds. H1 and H4 establish bonds with lipid phosphate groups (Fig. 8D) and water molecules (Fig. 8C) for half of the simulation time. In approximately the other half of the simulation, intramolecular bonds are formed between H1 (H4) and O2 (O3). When located inside the bilayer, the intramolecular hydrogen bonds are prevailing. However, it should be noted that all these hydrogen bonds break and form quickly with an average life time of 1 ns.

The intramolecular hydrogen bonds in UQ1H2 can also be analyzed from the respective C1–O1 (or C4–O4) bond torsions (Fig. 11). The energy minimum of the force-field calculated in vacuum found at $\sim 0^\circ$ is normally populated in solution as it corresponds to the intramolecular hydrogen bond H1–O2 and H4–O3. The second energy minimum at $\pm 180^\circ$ is not populated in solution or in the membrane, as water or any H-bond acceptor (such as the lipid phosphate, Fig. 8) is hindered by HM5 or H7 and no hydrogen bond can be established for this configuration. The intermolecular acceptors bind to UQ1H2 hydrogens when this torsion is $\pm 60^\circ$, resulting in a broad distribution for C1–O1 (or C4–O4) bond torsions in the $[-90^\circ, 90^\circ]$ range.

Rotamers of the methoxide groups with angles of $\pm 70^\circ$ for torsion of bonds C2–O2 and C3–O3 are more populated in condensed phase than expected from their force-field energy profiles (Fig. 11B). As described

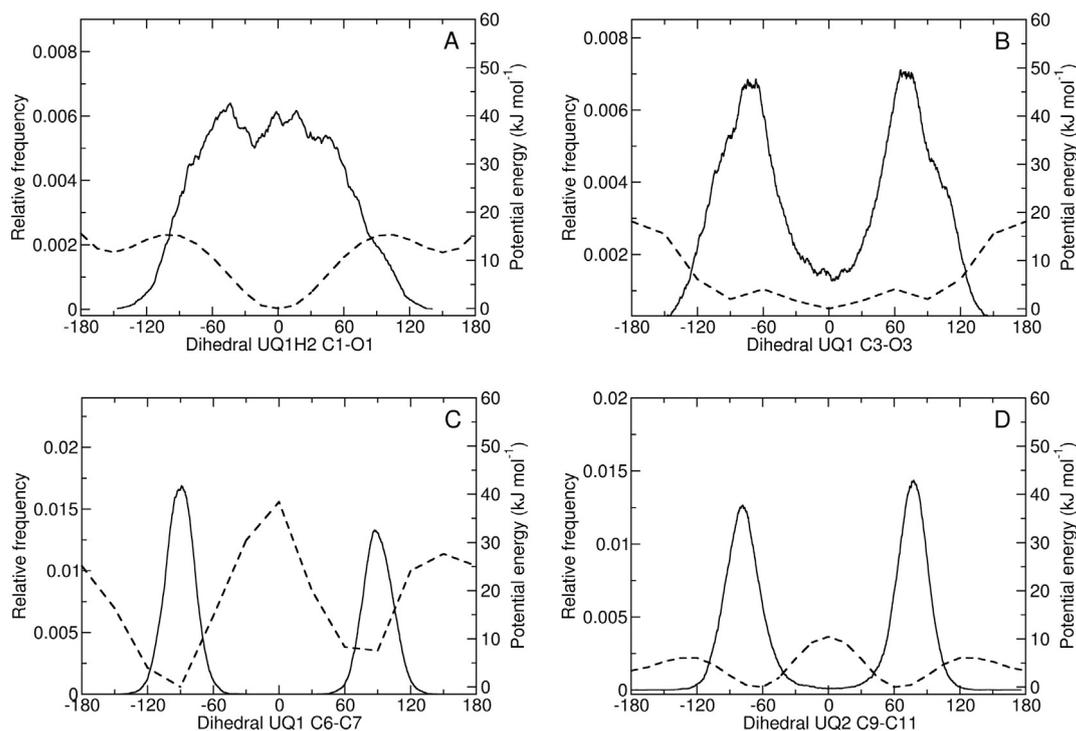


Fig. 11. Dihedral angle distributions of bond torsions. Bond C1–O1 shown in UQ1H2 (panel A), bonds C3–O3 (B) and C6–C7 (C) in UQ1, and bond C9–C11 in UQ2 (D) were calculated over unconstrained MD simulations (full lines). The force-field energy in vacuum is also shown in dashed lines.

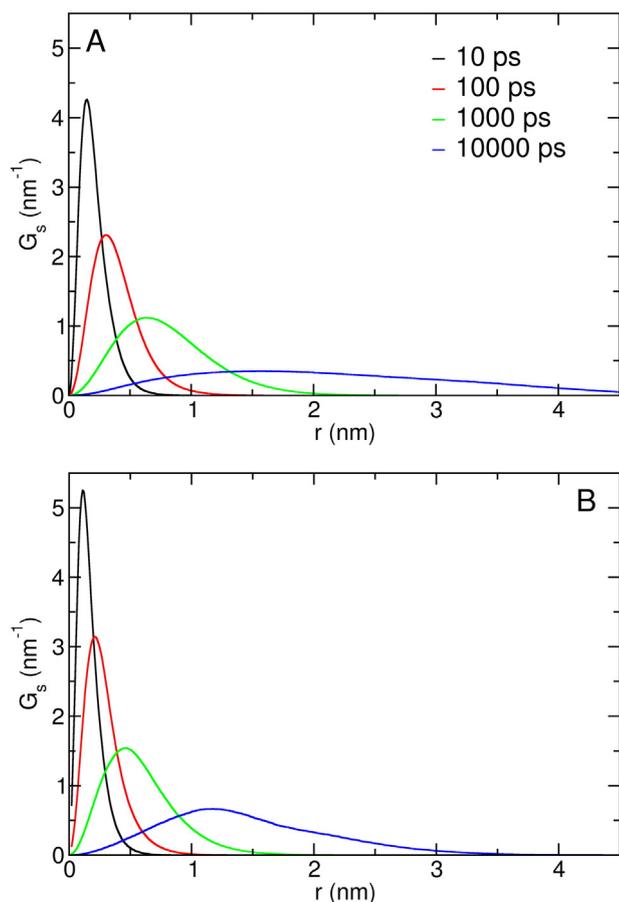


Fig. 12. Self-part of the van Hove correlation function (G_s). Calculations were performed in unconstrained 200 ns MD trajectories for POPC (panel A) and UQ6 (B) at times $t = 0.01, 0.1, 1, \text{ and } 10$ ns.

above, O2 and O3 are available as H-bond acceptors preferentially for these dihedral angles (Fig. 8A and B) [29]. The region around 0° is less populated as these configurations partially block hydrogen bonding to O1/O4. In the bilayer interior, where no H-bond donors are available, bond torsion distributions are closer to what is expected from the force-field profile (Fig. S5A).

Distributions of non-polar bond torsions are not changed from the corresponding torsion potentials, as shown in Fig. 11C and D for the C6–C7 and C9–C11 bonds in the isoprenoid tail. It should be noted that transitions over the high energy barriers around bond C6–C7 were observed during the unconstrained MD simulation time scale (200 ns).

3.4. Ubiquinone mobility and diffusion over the membrane

The dynamics of ubiquinone embedded in the bilayer was investigated by the self part of van Hove correlation function, $G_s(r, \Delta t)$ [71, 72]. This function gives the probability for a particle to show position displacements (r) in Δt time intervals. The van Hove distributions for POPC at 0.01 ns and longer times are gaussian-shaped as shown in Fig. 12. This is expected for a simple diffusion mechanism and is in agreement with a fast characteristic relaxation time (< 10 ps) for lipids

Table 2
Experimental and simulated (MD) lateral diffusion coefficients (in $10^{-7} \text{ cm}^2 \text{ s}^{-1}$) for various ubiquinones complexed to lipid bilayers.^a

	Experiment	MD
UQ1	6[19], 9[14]	4.2
UQ1H2		12.0
UQ2	12[19], 25[14]	4.0
UQ6		4.7
UQ10	13[19], 2.6[20], 5[14]	5.3
POPC	1.5[76]	6.0

^a Experimental coefficients were measured by fluorescence quenching in asolectin vesicles [19,14,20] and by NMR relaxation in pure POPC oriented bilayers [76] in temperatures of ~ 310 K. Simulated values were obtained from the angular coefficient of the linear region of Eq. (3) calculated from unconstrained MD simulations.

in liquid crystalline phase [73,74]. In this regime, the mean square displacement has a linear dependence with respect to time ($\langle x^2(t) \rangle \propto t$). The same behavior is observed for ubiquinone and ubiquinol, but relaxation times are shorter and higher displacements are observed in the same timescale.

In contrast to what is observed for the pure POPC lipid, linearity of squared-displacement $\langle x^2(t) \rangle$ with time is lost for ubiquinones at times longer than 25 ns, already observed from the non-gaussian shape of the $t = 10$ ns curve for UQ6 (Fig. 12B). It may be attributed to a slow relaxation process of the lipid polar group and water solvating ubiquinone at the bilayer interface.

Table 2 shows lateral diffusion coefficients D calculated from the Einstein relation:

$$D = \lim_{t \rightarrow \infty} \frac{\langle x^2(t) \rangle}{4t}. \quad (4)$$

Similar values (not shown) were calculated from fitting the self part of van Hove correlation function to a gaussian function (whereas $G_s(r, \Delta t) = (r/2D\Delta t)\exp[-r^2/4D\Delta t]$) [75].

Calculated lateral diffusion constants are about the same for all ubiquinones (UQ1–UQ10). But, diffusion is three times faster for ubiquinol (UQ1H2). The value calculated for POPC diffusion is in between. This result suggests that the polar region in amphiphathic molecules is the major determinant of their diffusion dynamics. The calculated results should be compared to diffusion coefficients obtained from experiments that probe similar molecular scales. Thus, given the simulated time (ns) and length (nm) scales, methods such as fluorescence collisional quenching and NMR relaxation are the most appropriate for comparisons [22,19]. As shown in Table 2, these methods give experimental diffusion constants that vary up to one order of magnitude ($2.6\text{--}25 \cdot 10^{-7} \text{ cm}^2 \text{ s}^{-1}$) for different ubiquinones and membrane preparations. Results from other methods such as fluorescence recovery after photo-bleaching probe rather different length scales (μm) and yield slower diffusion constants in the order of 10^{-8} to $10^{-9} \text{ cm}^2 \text{ s}^{-1}$ [17, 21]. Thus, we conclude that results for the calculated diffusion constants are in the same order of magnitude to experimental measurements for both POPC pure lipid bilayers [76] and for ubiquinone embedded in the bilayer [18].

4. Conclusions

An accurate energy model of ubiquinone is crucial to describe its conformational flexibility and interfacial behavior. Here, we have parametrized a CHARMM compatible force-field for ubiquinone with special attention to group polarity and torsional barriers involving the isoprenoid tail. The description of water interactions with the ubiquinone polar head was also significantly improved. The proposed force-field and simulation model were validated against experimental data and important insight on the ubiquinone interfacial behavior was obtained by computing free energies for a series of homologs. The convergence of PMFs for long amphiphiles such as ubiquinones interacting with lipid bilayers was tackled by two enhanced sampling methods, umbrella sampling and bias-exchange metadynamics.

We performed a detailed analysis of ubiquinone interaction within the lipid bilayer by addressing structural and dynamical properties. Results are in line with previous experimental studies [11–15] and support the proposal that ubiquinone head is localized in the interfacial region near the lipid glycerol groups and is highly hydrated due to membrane protrusion. The head plane displays a normal orientation with respect to the membrane midplane direction. Isoprenyl tails are extended and packed with the lipid acyl chains. In long tail homologs such as UQ6 and UQ10, the isoprenoid chain interdigitates across the bilayer and its terminal isoprenoid units have high flexibility. The internal torsions of ubiquinone are modulated by the condensed phase. In particular, ubiquinol hydrogen bonds to the lipid phosphate group, to water and

intermolecularly to its methoxide oxygens. It was also found that the diffusion rate of various ubiquinones embedded in a bilayer is similar to that of POPC lipids.

Notably the equilibrium position of ubiquinone head along the bilayer normal z-axis is coincident with the modulus of the average position of ubiquinone redox sites in cytochrome bc_1 and with the entrance of the binding site in NADH:ubiquinone reductase. Given the rather small changes observed here for polar head localization and orientation, and dynamics for ubiquinones with different isoprenoid tail lengths, we can also speculate that, in the complex lipid mixtures observed in bacterial or mitochondrial membranes — which lack cholesterol in their content, changes in bilayer acyl lipid composition will result in small perturbations to the ubiquinone partition and dynamics observed here.

The energy model and molecular dynamics simulations presented here for several ubiquinone homologs provide important insights on their behavior when embedded in a lipid bilayer. These results constitute a benchmark to studies of the transport and complexation mechanisms of this universal charge carrier with proteins involved in cellular respiration and photosynthesis.

Transparency document

The [Transparency document](#) associated with this article can be found, in the online version.

Acknowledgments

Funding from FAPESP (projects 12/02501-4, 12/17833-2 and 14/21900-2) is acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbabbio.2015.08.001>.

References

- [1] P. Mitchell, Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism, *Nature* 191 (4784) (1961) 144–148.
- [2] D. Green, The mitochondrial electron transfer system, *Comp. Biochem.* 14 (1966) 309–326.
- [3] A. Kröger, M. Klingenberg, S. Schweidler, Further evidence for the pool function of ubiquinone as derived from the inhibition of the electron transport by antimycin, *Eur. J. Biochem.* 39 (2) (1973) 313–323.
- [4] C. Heron, C. Ragan, B. Trumppower, The interaction between mitochondrial NADH-ubiquinone oxidoreductase and ubiquinol-cytochrome c oxidoreductase. restoration of ubiquinone-pool behaviour, *Biochem. J.* 174 (1978) 791–800.
- [5] L. Ernster, G. Dallner, Biochemical, physiological and medical aspects of ubiquinone function, *Biochim. Biophys. Acta (BBA) - Mol. Basis Dis.* 1271 (1) (1995) 195–204.
- [6] H. Katsikas, P. Quinn, The polyisoprenoid chain length influences the interaction of ubiquinones with phospholipid bilayers, *Biochim. Biophys. Acta* 689 (1982) 363–369.
- [7] H. Katsikas, P.J. Quinn, Fluorescence probe studies of the distribution of ubiquinone homologues in bilayers of dipalmitoyl glycerophosphocholine, *Eur. J. Biochem.* 131 (1983) 607–612.
- [8] G. Metz, K. Howard, W. van Liemt, J. Prestegard, J. Lugtenburg, S. Smith, NMR studies of ubiquinone location in oriented model membranes: evidence for a single motionally-averaged population, *J. Am. Chem. Soc.* 117 (1995) 564–565.
- [9] M. Jemiola-Rzeminska, J. Kruk, M. Skowronek, K. Strzalka, Location of ubiquinone homologues in liposome membranes studied by fluorescence anisotropy of diphenyl-hexatriene and trimethylammonium-diphenyl-hexatriene, *Chem. Phys. Lipids* 79 (1996) 55–63.
- [10] T. Hauß, S. Dante, T.H. Haines, N.A. Dencher, Localization of coenzyme Q10 in the center of a deuterated lipid membrane by neutron diffraction, *Biochim. Biophys. Acta* 1710 (2005) 57–62.
- [11] P.B. Kingsley, G.W. Feigenson, 1 H-NMR study of the location and motion of ubiquinones in perdeuterated phosphatidylcholine bilayers, *Biochim. Biophys. Acta Bioenerg.* 635 (3) (1981) 602–618.
- [12] M.A. Stidham, T. McIntosh, J.N. Siedow, On the localization of ubiquinone in phosphatidylcholine bilayers, *Biochim. Biophys. Acta* 767 (1984) 423–431.
- [13] B.A. Cornell, M.A. Keniry, A. Post, R.N. Robertson, L.E. Weir, P.W. Westerman, Location and activity of ubiquinone 10 and ubiquinone analogues in model and biological membranes, *Biochemistry* 26 (24) (1987) 7702–7707.

- [14] G. Lenaz, B. Samori, R. Fato, M. Battino, G.P. Castelli, I. Domini, Localization and preferred orientations of ubiquinone homologs in model bilayers, *Biochem. Cell Biol.* 70 (1992) 504–514.
- [15] M. Afri, B. Ehrenberg, Y. Talmon, J. Schmidt, Y. Cohena, A.A. Frimer, Location and mobility of ubiquinones of different chain lengths in artificial membrane vesicles, *Chem. Phys. Lipids* 131 (2004) 107–121.
- [16] G. Lenaz, Quinone specificity of complex I, *Biochim. Biophys. Acta* 1364 (1998) 207–221.
- [17] S. Gupte, E.-S. Wu, L. Hoehli, M. Hoehli, K. Jacobson, A.E. Sowers, C.R. Hackenbrock, Relationship between lateral diffusion, collision frequency, and electron transfer of mitochondrial inner membrane oxidation–reduction components, *Proc. Natl. Acad. Sci. U. S. A.* 81 (1984) 2606–2610.
- [18] R. Fato, M. Battino, M.D. Esposti, G.P. Castelli, G. Lenaz, Determination of partition and lateral diffusion coefficients of ubiquinones by fluorescence quenching of *n*-(9-Anthroyloxy)stearic acids in phospholipid vesicles and mitochondrial membranes, *Biochemistry* 25 (1986) 3378–3390.
- [19] S.D. Bernardo, R. Fato, R. Casadio, P. Fariselli, G. Lenaz, A high diffusion coefficient for coenzyme Q10 might be related to a folded structure, *FEBS Lett.* 426 (1998) 77–80.
- [20] R. Fato, E. Estornell, S.D. Bernardo, F. Pallotti, G.P. Castelli, G. Lenaz, A method for estimating lateral diffusion coefficients in membranes from steady-state fluorescence quenching studies, *Biophys. J.* 51 (1987) 735–744.
- [21] I. Llorente-García, T. Lennic, H. Erhardt, O.L. Harriman, L.-N. Liu, A. Robson, S.-W. Chiu, S. Matthews, N.J. Willis, C.D. Bray, S.-H. Lee, J.Y. Shin, C. Bustamante, J. Liphardt, T. Friedriche, C.W. Mullineaux, M.C. Leak, Single-molecule *in vivo* imaging of bacterial respiratory complexes indicates delocalized oxidative phosphorylation, *Biochim. Biophys. Acta* 1837 (2014) 811–824.
- [22] J.A. Söderhäll, A. Laaksonen, Molecular dynamics simulations of ubiquinone inside a lipid bilayer, *J. Phys. Chem. B* 105 (2001) 9308–9315.
- [23] H.A. Filipe, M.J. Moreno, T. Róg, I. Vattulainen, L.M. Loura, How to tackle the issues in free energy simulations of long amphiphiles interacting with lipid membranes: convergence and local membrane deformations, *J. Phys. Chem. B* 118 (2014) 3572–3581.
- [24] Z. Ghaemi, M. Minozzi, P. Carloni, A. Laio, A novel approach to the investigation of passive molecular permeation through lipid bilayers from atomistic simulations, *J. Phys. Chem. B* 116 (2012) 8714–8721.
- [25] J.P. Jambeck, A.P. Lyubartsev, Exploring the free energy landscape of solutes embedded in lipid bilayers, *J. Phys. Chem. Lett.* 4 (2013) 1781–1787.
- [26] F. Autenrieth, E. Tajkhorshid, J. Baudry, Z. Luthey-Schulten, Classical force field parameters for the heme prosthetic group of cytochrome *c*, *J. Comput. Chem.* 25 (13) (2004) 1613–1622.
- [27] A. Aird, J. Wrachtrup, K. Schulten, C. Tietz, Possible pathway for ubiquinone shuttling in *Rhodospirillum rubrum* revealed by molecular dynamics simulations, *Biophys. J.* 92 (1) (2007) 23–33.
- [28] K. Kaszuba, P.A. Postila, O. Cramariuc, M. Sarewicz, A. Osyczka, I. Vattulainen, T. Róg, Parametrization of the prosthetic redox centers of the bacterial cytochrome *Bc1* complex for atomistic molecular dynamics simulations, *Theor. Chem. Accounts* 132 (1370) (2013) 1–11.
- [29] J.V. Vermaas, A.T. Taguchi, S.A. Dikanov, C.A. Wraight, E. Tajkhorshid, Redox potential tuning through differential quinone binding in the photosynthetic reaction center of *Rhodobacter sphaeroides*, *Biochemistry* 54 (2015) 2104–2116.
- [30] J.A. Nilsson, A. Lyubartsev, L.A. Eriksson, A. Laaksonen, Molecular dynamics simulations of ubiquinone; a survey over torsional potentials and hydrogen bonds, *Mol. Phys.* 99 (21) (2001) 1795–1804.
- [31] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W. E. R. III, B. Roux, M. Schlenkerich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (1998) 3586–3616.
- [32] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. M. Jr., CHARMM General Force Field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, *J. Comput. Chem.* 31 (4) (2010) 671–690.
- [33] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. M. Jr., R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, *J. Phys. Chem. B* 114 (23) (2010) 7830–7843.
- [34] P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision A.1, Gaussian, Inc., Wallingford CT, 2009.
- [35] A.D. Becke, A new mixing of Hartree–Fock and local density-functional theories, *J. Chem. Phys.* 98 (2) (1993) 1372–1377.
- [36] W.J. Hehre, R. Ditchfield, J.A. Pople, Self-consistent molecular orbital methods. XII. Further extensions of GaussianType basis sets for use in molecular orbital studies of organic molecules, *J. Chem. Phys.* 56 (5) (1972) 2257–2261.
- [37] A. Szabo, N. Ostlund, *Modern Quantum Chemistry*, 1st edn Dover, New York, 1989.
- [38] F.B. Van Duijneveldt, J.G.C.M. Van de Rijdt, J.H. Van Lenthe, State of the art in counterpoise theory, *Chem. Rev.* 94 (1994) 1873–1885.
- [39] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935.
- [40] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics* 29 (2013) 845–854, <http://dx.doi.org/10.1093/bioinformatics/btt055>.
- [41] G.M. Arantes, M. Loos, Specific parametrization of a hybrid potential to simulate reactions in Phosphatases, *Phys. Chem. Chem. Phys.* 8 (2006) 347–353.
- [42] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (014101) (2007) 1–7.
- [43] H. Berendsen, J. Postma, W. vanGunsteren, A. DiNola, J. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (8) (1984) 3684–3690.
- [44] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald: an *N*-log(*N*) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (12) (1993) 10089–10092.
- [45] B. Hess, H. Bekker, H.J. Berendsen, J.G. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (12) (1997) 1463–1472.
- [46] C. Anézo, A.H. de Vries, H.-D. Höltje, D.P. Tieleman, S.-J. Marrink, Methodological issues in lipid bilayer simulations, *J. Phys. Chem. B* 107 (2003) 9424–9433.
- [47] N. Kucerka, S. Tristram-Nagle, J.F. Nagle, Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains, *J. Mol. Biol.* 208 (3) (2005) 193–202.
- [48] M. Javanainen, Universal method for embedding proteins into complex lipid bilayers for molecular dynamics simulations, *J. Chem. Theory Comput.* 10 (2014) 2577–2582.
- [49] A. Einstein, *Investigations on the Theory of the Brownian Movement*, Courier Corporation, 1956.
- [50] T.D. Romo, N. Leioatts, A. Grossfield, Lightweight object oriented structure analysis: tools for building tools to analyze molecular dynamics simulations, *J. Comput. Chem.* 35 (32) (2014) 2305–2318.
- [51] G. Torrie, J. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, *J. Comp. Physiol.* 23 (2) (1977) 187–199.
- [52] S. Piana, A. Laio, A bias-exchange approach to protein folding, *J. Phys. Chem. B* 111 (17) (2007) 4553–4559.
- [53] S. Kumar, J.M. Rosenberg, D.B.R.H. Swendsen, P.A. Kollman, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, *J. Comput. Chem.* 13 (8) (1992) 1011–1021.
- [54] J.S. Hub, B.L. de Groot, D. van der Spoel, g_wham—A free weighted histogram analysis implementation including robust error and autocorrelation estimates, *J. Chem. Theory Comput.* 6 (12) (2010) 3713–3720.
- [55] A. Laio, M. Parrinello, Escaping free-energy minima, *Proc. Natl. Acad. Sci. U. S. A.* 99 (20) (2002) 12562–12566.
- [56] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, M. Parrinello, PLUMED: a portable plugin for free-energy calculations with molecular dynamics, *Comput. Phys. Commun.* 180 (2009) 1961–1972.
- [57] A. Laio, F.L. Gervasio, Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science, *Rep. Prog. Phys.* 71 (126601) (2008) 1–22.
- [58] Y. Crespo, A. Laio, G.E. Santoro, E. Tosatti, Calculating thermodynamics properties of quantum systems by a non-Markovian Monte Carlo procedure, *Phys. Rev. E* 80 (015702) (2009) 1–4.
- [59] I.J. General, A note on the standard states binding free energy, *J. Chem. Theory Comput.* 6 (8) (2010) 2520–2524.
- [60] H. Ku, Notes on the use of propagation of error formulas, *J. Res. Natl. Bur. Stand.* 70 (4) (1966) 263–273.
- [61] C. Neale, W.D. Bennett, D.P. Tieleman, R. Pomès, Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers, *J. Chem. Theory Comput.* 7 (2011) 4175–4188.
- [62] C. Neale, C. Madill, S. Rauscher, R. Pomès, Accelerating convergence in molecular dynamics simulations of solutes in lipid membranes by conducting a random walk along the bilayer normal, *J. Chem. Theory Comput.* 9 (8) (2013) 3686–3703.
- [63] M.J. Hinner, S.-J. Marrink, Alex H. de Vries, Location, tilt, and binding: a molecular dynamics study of voltage-sensitive dyes in biomembranes, *J. Phys. Chem. B* 113 (48) (2009) 15807–15819.
- [64] V.E. Kagan, E.A. Serbinova, G.M. Koynova, S.A. Kitanova, V.A. Tyurin, T.S. Stoytchev, P.J. Quinn, L. Packer, Antioxidant action of ubiquinol homologues with different isoprenoid chain length in biomembranes, *Free Radic. Biol. Med.* 9 (1990) 117–126.
- [65] J.A. Platts, M.H. Abraham, D. Butina, A. Hersey, Estimation of molecular linear free energy relationship descriptors by a group contribution approach. 2. Prediction of partition coefficients, *J. Chem. Inf. Comput. Sci.* 40 (1) (2000) 71–80.
- [66] R. Fato, E. Estornell, S.D. Bernardo, F. Pallotti, G.P. Castelli, G. Lenaz, Steady-state kinetics of the reduction of coenzyme Q analogs by complex I (NADH:ubiquinone oxidoreductase) in bovine heart mitochondria and submitochondrial particles, *Biochemistry* 35 (1996) 2705–2716.
- [67] R. Fato, C. Castelluccio, G. Palmer, G. Lenaz, A simple method for the determination of the kinetic constants of membrane enzymes utilizing hydrophobic substrates: ubiquinol cytochrome *c* reductase, *Biochim. Biophys. Acta* 932 (1988) 216–222.
- [68] S. Iwata, J.W. Lee, K. Okada, J.K. Lee, M. Iwata, B. Rasmussen, T.A. Link, S. Ramaswamy, B. Jap, Complete structure of the 11-subunit bovine mitochondrial cytochrome *b*_c complex, *Science* 281 (1998) 64–71.
- [69] C. Lange, J.H. Nett, B.L. Trumpower, C. Hunte, Specific roles of protein–phospholipid interactions in the yeast cytochrome *b*_c complex structure, *EMBO J.* 20 (2001) 6591–6600.
- [70] R. Baradaran, J.M. Berrisford, G.S. Minhas, L.A. Sazanov, Crystal structure of the entire respiratory complex I, *Nature* 494 (2013) 443–448.

- [71] L. van Hove, Correlations in space and time and born approximation scattering in systems of interacting particles, *Phys. Rev.* 95 (1) (1954) 249–3720.
- [72] M.P. Allen, D.J. Tildesley, J.R. Banavar, *Computer Simulation of Liquids*, vol. 42, American Institute of Physics, 2008.
- [73] P. Khalatur, N. Balabaev, A. Pavlov, Molecular dynamics study of a lipid bilayer and a polymer liquid, *Mol. Phys.* 59 (4) (1986) 753–773.
- [74] F.W. Starr, B. Hartmann, J.F. Douglas, Dynamical clustering and a mechanism for raft-like structures in a model lipid membrane, *Soft Matter* 10 (2014) 3036–3047.
- [75] P.S. Niemela, M.S. Miettinen, L. Monticelli, H. Hammaren, P. Bjelkmar, T. Murtola, E. Lindahl, I. Vattulainen, Membrane proteins diffuse as dynamic complexes with lipids, *J. Am. Chem. Soc.* 132 (22) (2010) 7574–7575.
- [76] A. Filippov, G. Orodd, G. Lindblom, Influence of cholesterol and water content on phospholipid lateral diffusion in bilayers, *Langmuir* 19 (2003) 6397–6400.