

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”  
FACULDADE DE ENGENHARIA  
CAMPUS DE ILHA SOLTEIRA**

**GLAUCO GARRIDO ALMEIDA**

**ESTUDO DA DINÂMICA DE SOLVATAÇÃO DE  
PEPTÍDEOS POR QM/MM**

Ilha Solteira

2015

**GLAUCO GARRIDO ALMEIDA**

## **Estudo da dinâmica de solvatação de peptídeos por QM/MM**

**Prof. Dr. João Manuel Marques Cordeiro  
Orientador**

Tese apresentada ao programa de Pós-Graduação em Ciência dos Materiais da Universidade Estadual Paulista – “Julio de Mesquita Filho” – UNESP – Campus Ilha Solteira- como parte das exigências para a obtenção do Título de Doutor em Ciência dos Materiais. Área de conhecimento: Química dos Materiais.

Ilha Solteira

2015

FICHA CATALOGRÁFICA

Desenvolvido pelo Serviço Técnico de Biblioteca e Documentação

A447e Almeida, Glauco Garrido.  
Estudo da dinâmica de solvatação de peptídeos por QM/MM / Glauco Garrido Almeida. -- Ilha Solteira: [s.n.], 2015  
103 f. : il.

Tese (doutorado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Química dos Materiais, 2015

Orientador: João Manuel Marques Cordeiro  
Inclui bibliografia

1. QM/MM. 2. Peptídeos. 3. Alanina. 4. Interação de hidrogênio.

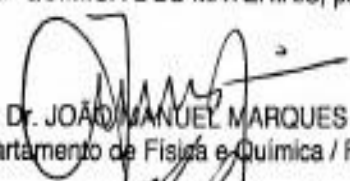
**CERTIFICADO DE APROVAÇÃO**

**TÍTULO:** Estudo da dinâmica de solvatação de peptídeos por QMMM


**AUTOR:** GLAUCO GARRIDO ALMEIDA

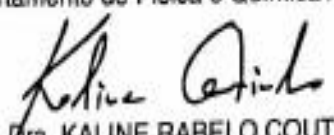
**ORIENTADOR:** Prof. Dr. JOÃO MANUEL MARQUES CORDEIRO

Aprovado como parte das exigências para obtenção do Título de Doutor em Ciência dos Materiais ,  
Área: QUÍMICA DOS MATERIAIS, pela Comissão Examinadora:

  
Prof. Dr. JOÃO MANUEL MARQUES CORDEIRO  
Departamento de Física e Química / Faculdade de Engenharia de Ilha Solteira

  
Prof. Dr. RAFAEL ZADOROSNY  
Departamento de Física e Química / Faculdade de Engenharia de Ilha Solteira

  
Prof. Dr. HAROLDO NAOYUKI NAGASHIMA  
Departamento de Física e Química / Faculdade de Engenharia de Ilha Solteira

  
Prof. Dra. KALINE RABELO COUTINHO  
Departamento de Física Geral / Universidade de São Paulo

  
Prof. Dra. MARÍA ELENA MARTÍN NAVARRO  
Departamento de Química Física / Universidad de Extremadura

Dedico esta tese à minha vó, Maria,  
meus pais, Reginaldo e Glaucia,  
minha irmã ,Regiane  
e, minha companheira Elis Regina

## AGRADECIMENTOS

À Deus, por tudo.

Ao meu orientador e amigo, João, por tudo que fez por mim ao longo desses 6 anos.

Ao grupo de pesquisa QCAMM, em especial à Elena, Manolo e Aurora, por me receberem tão bem e terem tido toda a paciência do mundo comigo.

Aos meus pais, Reginaldo e Glaucia, pelo suporte financeiro e emocional.

À minha irmã, Regiane, e minha companheira Elis Regina, por me apoiarem e ajudarem sempre que precisei.

Aos meus amigos de verdade, que servem de inspiração para que eu continue evoluindo sempre.

À todos meus companheiros de pós-graduação que me ajudaram com as disciplinas.

À todos os docentes

que dividiram seu conhecimento comigo.

À todos aqueles que contribuíram de maneira direta ou indireta para a conclusão deste trabalho.

À CAPES pelo suporte financeiro e por possibilitar um ano do doutorado na Espanha, que foi decisivo para o desenvolvimento do trabalho.( PDSE-CAPES, (nº 99999.014630 / 2013-03)

"A imaginação é mais importante que o conhecimento.  
O conhecimento é limitado. A imaginação envolve o mundo."

Albert Einstein

## RESUMO

Trabalhos experimentais e teóricos desenvolvidos nos últimos anos evidenciaram a capacidade de alcoóis e polialcoóis em mudar características energéticas em proteínas. No entanto, o mecanismo responsável por esse efeito não está totalmente elucidado. Tomando como exemplo o dipeptídeo de alanina, foram realizadas simulações QM/MM em água, etanol e mistura 60-40 % em volume de água-etanol. A molécula de dipeptídeo foi descrita em nível de cálculo quântico MP2/aug-cc-pVDZ. Em solução, apenas conformeros  $\alpha_R$  e PPII foram encontrados na população de equilíbrio. A diferença de energia livre em solução  $\alpha_R \rightarrow$  PPII é determinada pelo balanço entre energia interna do soluto e energia de interação. Para o dipeptídeo de alanina, qualquer fator que aumente a energia de interação soluto-solvente favorece o aumento da população de  $\alpha_R$ . Por outro lado fatores que diminuam esse valor, como a adição de etanol, aumentam a população de PPII. Os resultados indicam solvatação preferencial para o sistema, evidenciado pela formação majoritária da primeira camada de solvatação por moléculas de água (na mistura), embora seja possível encontrar moléculas de etanol em pequenas quantidades ao redor do grupo carboxílico da extremidade N-terminal e, cadeias metálicas laterais. Todavia, o comportamento parece não afetar a diferença na estabilidade de equilíbrio conformacional.

**Palavras-chave:** QM/MM. Alanina. Peptídeos. Interação de hidrogênio. Etanol.

## ABSTRACT

Recent papers, both experimental and theoretical, have highlighted the capacity of alcohols and polyalcohol in modifying the energy landscape in proteins. However, the mechanism underlying this effect is not fully elucidated. Taking as a model-system the alanine dipeptide, QM/MM calculations were performed in water, ethanol and solution 60-40% (volume) water-ethanol. The dipeptide molecule was described by MP2/aug-cc-pVDZ level. In solution, only  $\alpha_R$  and PPII conformers were found in the conformational equilibrium population. The free energy difference in solution  $\alpha_R \rightarrow$  PPII is determined by the interplay between internal energy and the interaction energy. It has been found that, for alanine dipeptide, any factor that increases the solute-solvent interaction energy also promotes an increasing on  $\alpha_R$  stability, moreover factors that decreases this value, such as the addition of ethanol molecules, increases the PPII stability. The results points to a preferential solvation behavior for the system, as evidenced by the composition of the first solvation shell in the mixture mainly populated by water molecules, although it is possible to find small concentration of ethanol molecules around the carboxyl group of the N-terminal end and around methylic side chains. However, this behavior does not seem to affect the differential conformational stability.

**Keywords:** QM/MM. Alanine. Peptides. Hydrogen bond. Ethanol.

## LISTA DE FIGURAS

<b>Figura 1</b>	Representação em new cartoon da proteína citrato sintase destacando pela diferença de cores os diversos resíduos de aminoácidos unidos por ligações peptídicas.....	<b>14</b>
<b>Figura 2</b>	Representação em new cartoon da proteína citrato sintase destacando pela diferença de cores as estruturas secundárias, alfa hélice (amarelo), folhas $\beta$ (laranja) e voltas (vermelho).....	<b>15</b>
<b>Figura 3</b>	Representação em new cartoon da proteína citrato sintase destacando as duas subunidades terciárias que formam o homodímero.....	<b>16</b>
<b>Figura 4</b>	Diagrama de Ramachandran indicando os 9 possíveis conforméromos de acordo com os impedimentos estéricos	<b>21</b>
<b>Figura 5</b>	Conforméromo PPII definido pelos ângulos $\phi$ e $\Psi$ com numeração atômica: Cores: hidrogênios (cinza), carbonos (azul claro), nitrogênio (azul escuro), oxigênio (vermelho).....	<b>21</b>
<b>Figura 6</b>	Conforméromo $\alpha R$ definido pelos ângulos $\phi$ e $\Psi$ com numeração atômica: Cores: hidrogênios (cinza), carbonos (azul claro), nitrogênio (azul escuro), oxigênio (vermelho).....	<b>22</b>
<b>Figura 7</b>	Potencial harmônico linear.....	<b>27</b>
<b>Figura 8</b>	Potencial harmônico angular.....	<b>28</b>
<b>Figura 9</b>	Potencial torsional.....	<b>29</b>
<b>Figura 10</b>	Potencial de Lennard Jones.....	<b>29</b>
<b>Figura 11</b>	Potencial de Coulomb.....	<b>30</b>
<b>Figura 12</b>	Simulação das condições de contorno utilizadas no método.....	<b>32</b>
<b>Figura 13</b>	Representação do raio de corte e convenção de imagem mínima.....	<b>33</b>
<b>Figura 14</b>	Cargas representando o potencial eletrostático médio gerado pela primeira camada de solvatação.....	<b>45</b>
<b>Figura 15</b>	Cargas representando o potencial eletrostático médio gerado pelas cargas além da primeira camada de solvatação.....	<b>45</b>
<b>Figura 16</b>	Esquema básico do método ASEP/MD.....	<b>47</b>
<b>Figura 17</b>	Fração molar local da água em função da distância (em Å) para o sistema AD-água (linha solida). A linha pontilha representa a fração molar de $x_w=0,68$ .....	<b>58</b>

<b>Figura 18</b>	Funções de distribuição radial de pares para o sistema AD-solventes diversos. a) Confórmero $\alpha R$ em etanol puro (azul) e em mistura(vermelho). b) Confórmero $\alpha R$ em água pura (azul) e em mistura (vermelho). c) Confórmeros $\alpha R$ (azul) e PPII (vermelho) em mistura.....	<b>60</b>
<b>Figura 19</b>	Distribuição de probabilidade tridimensional ao redor de $\alpha R$ solvatado na mistura 60-40(em volume) de água-etanol. Código de cores: Magenta: átomos de oxigênio da água. Laranja: átomos de hidrogênio da água. Vermelho: átomos de oxigênio de etanol. Cinza: átomos de hidrogênio de etanol.....	<b>63</b>
<b>Figura 20</b>	Distribuição de probabilidade tridimensional ao redor de PPII solvatado na mistura 60-40(em volume) de água-etanol. Código de cores: Magenta: átomos de oxigênio da água. Laranja: átomos de hidrogênio da água. Vermelho: átomos de oxigênio de etanol. Cinza: átomos de hidrogênio de etanol.....	<b>63</b>
<b>Figura 21</b>	Estrutura do Solvente Solvatado na mistura 60-40% (em volume) de Água +Etanol. Código de cores: Azul: átomos de oxigênio da água. Vermelho: átomos de oxigênio de etanol.....	<b>64</b>
<b>Figura 22</b>	Energia de interação soluto-solvente (Kcal/mol) para a mistura água-etanol em função da distância entre os átomos do soluto e solvente. Código de cores: Azul:PPII. Vermelho: $\alpha R$ .....	<b>65</b>

## LISTA DE TABELAS

<b>Tabela 1</b>	Comparação dos valores de energia livre (kcal/mol) e momento dipolo (D) para os conformeros $\alpha R$ e PPII utilizando-se os conjuntos de base cc-pVDZ e aug-cc-pVDZ em água.....	54
<b>Tabela 2</b>	Diferença de energia livre e suas componentes para a transformação $\alpha R \rightarrow PPII$ em três diferentes solventes. Todas as energias estão em Kcal/mol. Em parênteses estão os valores excluindo a contribuição $\Delta A(\text{rot+vibr})$ .....	54
<b>Tabela 3</b>	Parâmetros geométricos selecionados do conformero PPII em três diferentes solventes. Distâncias estão em angstroms e os ângulos em graus.....	57
<b>Tabela 4</b>	Parâmetros geométricos selecionados do conformero $\alpha R$ em três diferentes solventes. Distâncias estão em angstroms e os ângulos em graus.....	57
<b>Tabela 5</b>	Energia de interação e suas componentes (kcal/mol).....	59
<b>Tabela 6</b>	Número de coordenação ao redor de grupos funcionais selecionados.....	62
<b>Tabela 7</b>	Contribuição dos grupos funcionais para a energia de interação clássica para os conformeros $\alpha R$ e PPII.....	66
<b>Tabela 8</b>	Contribuição dos grupos funcionais para a energia de interação clássica para os conformeros $\alpha R$ e PPI utilizando a distribuição de cargas da fase gasosa.....	67

## LISTA DE ABREVIATURAS E SIGLAS

<b>AD</b>	Dipeptídeo de Alanina (do inglês <i>Alanine Dipeptide</i> )
<b>ASEP/MD</b>	Potencial Eletrostático Médio proveniente de Dinâmica Molecular (do inglês <i>Averaged Elestrotatic Potencial from Molecular Dynamics</i> )
<b>AT</b>	Tripeptídeo de Alanina (do inglês <i>Alanine Tripeptide</i> )
<b>C&amp;T</b>	Ciência e Tecnologia
<b>DNA</b>	Ácido Desoxirribonucleico (do inglês <i>Deoxiribonucleic acid</i> )
<b>fs</b>	Femtosegundo
<b>PDB</b>	Banco de Dados de Proteínas (do inglês <i>Protein Data Bank</i> )
<b>QCAMM</b>	Química Quântica e Modelagem Molecular (do inglês <i>Quantum Chemistry an Molecular Modeling</i> )
<b>QM/MM</b>	Mecânica Quântica - Mecânica Molecular (do inglês, <i>Quantum Mechanics/Molecular Mechanics</i> )
<b>VMD</b>	Dinâmica Molecular Visual (do inglês <i>Visual Molecular Dinamics</i> )

## SUMÁRIO

<b>Capítulo 1 – INTRODUÇÃO .....</b>	<b>13</b>
1.1 A PAISAGEM.....	13
1.2 O CAMINHO .....	19
1.3 O VEÍCULO.....	22
<b>Capítulo 2 – O Veículo: SIMULAÇÃO COMPUTACIONAL .....</b>	<b>24</b>
2.1 MECÂNICA MOLECULAR E CAMPOS DE FORÇA .....	25
2.1.1 Potenciais ligantes .....	27
2.1.2 Potenciais não ligantes .....	29
2.1.3 Condições periódicas de contorno.....	31
2.2 MÉTODO DE DINÂMICA MOLECULAR .....	32
2.3 INTRODUÇÃO À QM/MM .....	34
2.3.1 QM/MM convencional .....	35
2.3.2 QM/MM sequencial .....	37
2.4 MÉTODO DEASEP/MD .....	38
2.4.1 A aproximação de campo médio (MFA).....	40
2.4.2 Detalhes do método ASEP/MD.....	41
2.4.3 Localização dos pontos críticos em superfícies de energia livre.....	47
2.4.4 Cálculo de diferença de energia livre .....	48
2.4.5 Considerações finais .....	49
2.5 DETALHES COMPUTACIONAIS .....	50
<b>Capítulo 3 - O Passeio - RESULTADOS E DISCUSSÃO .....</b>	<b>52</b>
<b>Capítulo 4 - O Aprendizado - CONCLUSÕES.....</b>	<b>67</b>
<b>Capítulo 5 - A Jornada - CONSIDERAÇÕES FINAIS .....</b>	<b>69</b>
<b>REFERÊNCIAS .....</b>	<b>70</b>
<b>ANEXO.....</b>	<b>78</b>

## Capítulo 1 – INTRODUÇÃO

### 1.1 A PAISAGEM

À parte o aspecto financeiro, a ciência tem como principal combustível a curiosidade, seja em compreender a vida e as leis que regem o universo, seja em aplicar os conhecimentos adquiridos em tecnologia, no anseio por evolução. A questão que envolve o comportamento de proteínas frente a fatores externos tangem, certamente, ambos os casos, de tal maneira que os conhecimentos adquiridos na tentativa de compreender os mecanismos de funcionamento dessas moléculas se convertem em evolução tecnológica para as indústrias de bens de consumo, alimentícia e saúde.

As proteínas representam a extensão tridimensional de sequências de 4 tipos diferentes de nucleotídeos (cujos nomes podem ser abreviados para A, T, C, G) dispostos em moléculas de DNA (código genético) inseridas dentro dos núcleos celulares dos organismos, viabilizando a vida e sua manutenção por meio da participação pontual, em essencialmente, todos os processos biológicos. Elas podem atuar no transporte e armazenamento de outras moléculas, suporte mecânico, defesa imunológica, transmissão de impulsos nervosos, além de tornar possível o movimento e controlar a diferenciação e crescimento celulares, o que lhes confere o status de macromoléculas mais versáteis dos sistemas biológicos, e conseqüentemente, indispensáveis à vida.<sup>1</sup>

Cada proteína é um polímero constituído por ao menos 100, dos 20 diferentes tipos de  $\alpha$ -aminoácidos (monômeros), sendo que sua função será determinada de acordo com sua estrutura química final, podendo variar em número, tipo e disposição dos  $\alpha$ -aminoácidos formadores, produzindo diferentes proteínas com funções diversas<sup>2</sup>.

O conceito de que cada sequência de nucleotídeos encontrada na fita de DNA corresponde a uma proteína específica foi desenvolvido a partir de 1953, com o trabalho de Frederick Sanger que determinou a sequência de aminoácidos da insulina. Esse estudo verificou ainda que as ligações peptídicas que uniam os aminoácidos eram compostas apenas por isômeros L dos  $\alpha$ -aminoácidos. Hoje em dia são conhecidas as sequências de nucleotídeos de mais de 2000 proteínas.<sup>1</sup> Para compreender melhor a influência que determinada sequência de aminoácidos tem sobre a conformação final de proteínas, sua

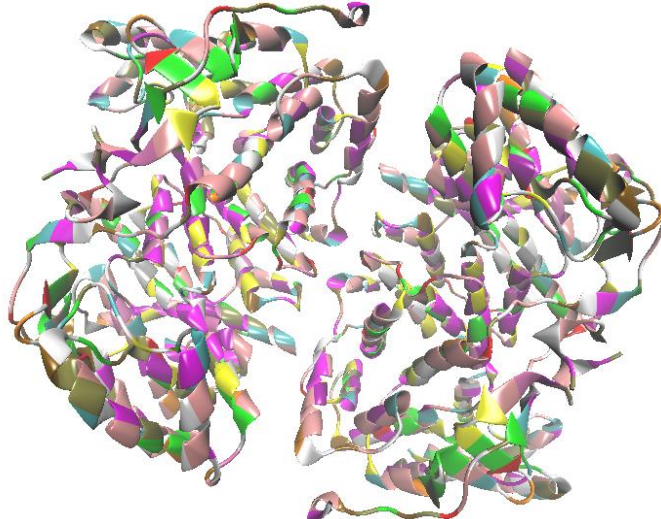
estrutura será explicada mais detalhadamente, indicando as forças envolvidas e as consequências estruturais que seguem:<sup>1-4</sup>

**Estrutura Primária:** A estrutura primária de uma proteína é constituída por uma sequência linear de  $\alpha$ -aminoácidos ligados covalentemente por ligações peptídicas que são formadas em uma reação de condensação entre dois  $\alpha$ -aminoácidos, onde o grupamento amina terminal de um  $\alpha$ -aminoácido se liga ao grupamento carboxílico terminal de outro por meio de um ataque nucleofílico do par de elétrons desemparelhados do átomo de N ao C da carbonila, originando uma ligação peptídica e a liberação de uma molécula de água.

Apesar de ser uma ligação simples, a ligação peptídica possui caráter parcial de ligação dupla devido a ressonância que existe entre ela e a ligação dupla do grupo carboxílico. Esse fator culmina em duas consequências importantes na determinação de suas características bioquímicas: A ligação peptídica é planar e tem caráter polar, onde N e O funcionam como aceptor e doador de elétrons, respectivamente. Portanto, a cadeia polipeptídica pode realizar torções em ângulos de ligações apenas em planos específicos, e há interações por ligações de hidrogênio tanto intramoleculares, quanto intermoleculares envolvidas nas estruturas subsequentes, ajudando na estabilização da molécula de proteína.

Entre todos os níveis estruturais de proteínas, a cadeia polipeptídica é a única que tem natureza 100% covalente, ou seja, a mais fortemente ligada, como poderemos comprovar ao longo do texto. As Figuras 1, 2 e 3 são representações que buscam destacar as estruturas primária, secundária, e, terciária/quaternária em uma proteína. Para exemplificar, usou-se a citrato sintase 1IXE (DOI 10.2210/pdb1ixe/pdb) e as figuras foram editadas com o programa VMD<sup>5</sup>

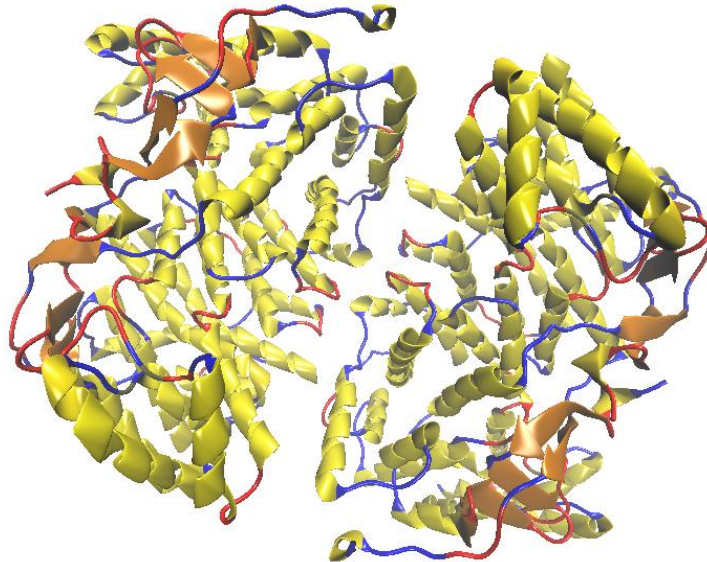
**Figura 1 - Representação em new cartoon da proteína citrato sintase destacando pela diferença de cores os diversos resíduos de aminoácidos unidos por ligações peptídicas.**



Fonte<sup>5</sup>

Estrutura Secundária: Pode assumir forma mais compacta ( $\alpha$ -hélice) que apresenta em média 3,6 resíduos por volta, mais estendida (folhas  $\beta$ , 2 resíduos por volta) ou laços (*turns* e *loops*). As diferentes configurações possíveis são acessadas a partir de interações de hidrogênio intramoleculares que se instalam entre os resíduos de aminoácidos, sendo que não há rotação ao redor da ligação peptídica e os impedimentos estéricos relativos ao volume ocupado pelo  $\alpha$ -aminoácido e por suas cadeias laterais são respeitados. Normalmente, a cadeia polipeptídica irá apresentar uma mescla de estruturas secundárias (Figura ).

**Figura 2 - Representação em new cartoon da proteína citrato sintase destacando pela diferença de cores as estruturas secundárias, alfa hélice (amarelo), folhas  $\beta$  (laranja) e voltas (vermelho e azul).**



Fonte<sup>5</sup>

Estrutura Terciária: As proteínas que possuem esse nível de estruturação molecular são denominadas proteínas globulares e representam o enovelamento da estrutura secundária da cadeia polipeptídica. Se a estrutura primária é constituída inteiramente por ligações covalentes enquanto que a estrutura secundária é constituída inteiramente por interações de hidrogênio intramoleculares, a estrutura terciária apresenta tanto interações não covalentes (interações iônicas, interações de natureza hidrofóbica e interações de hidrogênio), quanto ligações covalentes, por meio das pontes de dissulfeto intramoleculares. A diversidade de interações citadas vai conferir à enzima uma configuração enovelada.

É importante destacar que caso esteja imersa em um meio hidrofílico a tendência é que os grupos polares das cadeias laterais se organizem de maneira a permanecer do lado de fora da molécula, enquanto que os grupos apolares se organizem de maneira a permanecer no interior da estrutura globular formada. Análogo a esse raciocínio é o caso em que a proteína

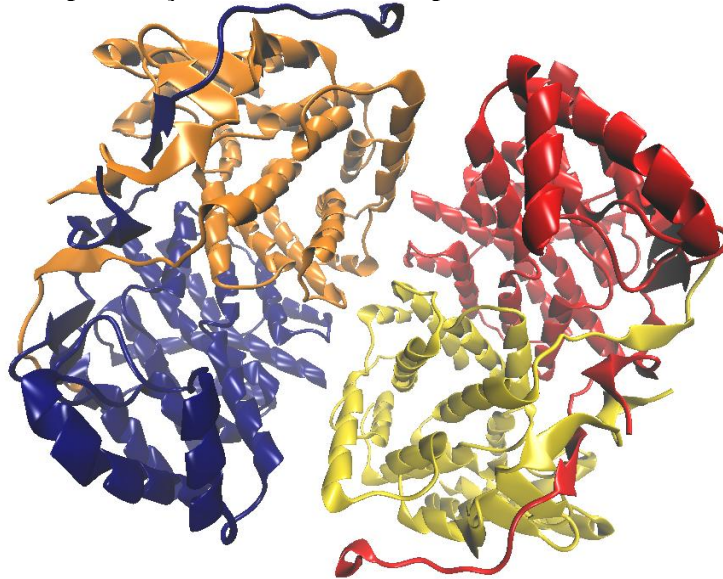
está imersa em um meio apolar, porém, nessa situação os grupos hidrofóbicos se organizam do lado de fora enquanto que os hidrofílicos do lado de dentro da proteína.

Neste ponto a estrutura protéica encontra sua conformação mais estável, ou seja, a estrutura terciária de uma proteína corresponde ao mínimo valor de energia de Gibbs para uma proteína monomérica ou para uma subunidade de proteína oligomérica.

Estrutura Quaternária: Algumas proteínas são denominadas oligoméricas, isso significa que são formadas por mais de uma cadeia polipeptídica, ou seja, mais de uma cadeia polipeptídica assume as estruturas secundárias e terciárias formando subunidades, ou monômeros, que se unem por meio de interações não covalentes (ligações de hidrogênio intermoleculares, interações de natureza iônica e hidrofóbica). A função de uma proteína oligomérica depende de todas as unidades.

O conglomerado de estruturas terciárias que dá origem a denominada estrutura quaternária pode ser formado por subunidades iguais ou distintas, ganhando o prefixo de homo- ou hetero-, respectivamente. Na figura 1 está representado um homodímero.

**Figura 1 - Representação em new cartoon da proteína citrato sintase .**



Fonte<sup>5</sup>

O modelo explanado acima faz alusão às proteínas globulares. Além destas existem as proteínas fibrosas. As diferenças entre elas advêm das distintas funções que desempenham, uma vez que as proteínas fibrosas estão mais relacionadas com funções estruturais, conferindo suporte, forma e proteção externa. As proteínas fibrosas apresentam, normalmente, apenas um tipo de estrutura secundária por cadeia polipeptídica e uma estrutura terciária mais simples e

estendida, enquanto que as proteínas globulares, que estão relacionadas com funções catalíticas e reguladoras, apresentam normalmente mais de uma estrutura secundária por cadeia polipeptídica e uma estrutura terciária mais enovelada e complexa. Apesar de ambas serem essenciais à vida, as proteínas globulares possuem uma relação direta entre a conformação estrutural de suas cadeias polipeptídicas e a função biológica que desempenham nos organismos. Assim sendo, a simples troca de um resíduo por outro na cadeia polipeptídica, que mude a maneira com que a proteína irá se enovelar, ou a exposição a fatores externos que as façam sofrer um processo de perda de suas estruturas tridimensionais nativas (estruturas terciária e quaternárias, no caso de proteínas oligoméricas), acarretará perda de seu papel funcional. O processo de perda da estrutura tridimensional nativa da proteína é denominado denaturação proteica e é exatamente o ponto que tange a busca de conhecimento a respeito dos processos que tornam a vida possível e a aplicação de conhecimento em busca de tecnologias que permitam aumentar o lucro e diminuir o impacto ambiental.

Geralmente, enzimas oligoméricas são naturalmente mais estáveis quando comparadas com monoméricas ou estruturas quaternárias mais simples, devido sua maior rigidez.

Entender os mecanismos do processo pelo qual as proteínas se enovelam ou se denaturam significa ao mesmo tempo ter a capacidade de compreender como organismos extremófilos, como bactérias termófilas e arqueobactérias, são capazes de sobreviver em temperaturas de aproximadamente 100°C, mesmo diferindo tão pouco de suas proteínas homólogas de bactérias como a *Escherichia coli*,<sup>2</sup> ser capaz de propor terapias gênicas que busquem denaturar proteínas que sejam essenciais à reprodução de células cancerígenas e ter a capacidade de utilizar a biocatálise em quaisquer condições reacionais.

Embora os três aspectos sejam importantes, as duas primeiras aplicações necessitam de avanços mais significativos, por serem problemas mais complexos, enquanto que a biocatálise é um campo que está em crescimento, especialmente depois do advento da engenharia genética na década de 1980, que possibilitou a modificação do arranjo atômico nas enzimas e sua produção em maior escala permitindo a utilização em produção de insumos de alto valor agregado.<sup>6,7</sup> Atualmente busca-se fomentar e contribuir na formulação de novas políticas de C&T (ciência e tecnologia) aliando produtividade e sustentabilidade, segundo os 24 princípios que foram estabelecidos para a química verde.<sup>8,9</sup> As características estruturais e bioquímicas das enzimas coadunam-se com essas perspectivas, uma vez que apresentam alta

especificidade na ligação enzima-substrato: regioseletividade (que diz respeito à ligação no sítio ativo da molécula, que leva enzima e substrato a se ligarem sempre na mesma posição), quimioseletividade (que confere à enzima a seleção de determinado grupo funcional presente no substrato (caso haja outros grupos, estes permanecerão inalterados frente à catálise enzimática)) e enancioseletividade, que permite que a enzima reconheça o centro quiral presente na molécula, identificando os pares de enantiômeros.<sup>10</sup> Essa alta especificidade livra os produtos de impurezas emergentes de reações secundárias, poupando o processo das, em muitos casos, complexas e caras operações de purificação, com potenciais reflexos importantes na indústria fármaco-química.<sup>9,11</sup> Além disso as enzimas apresentam alto *turnover* e biodegradabilidade, podendo aumentar a velocidade reacional em fatores de até  $10^{20}$  enquanto outros catalisadores variam a velocidade de uma reação em fatores que vão de  $10^2$  a  $10^4$  vezes.<sup>12</sup>

Para ilustrar a importância da biocatálise na indústria, tomemos como exemplo a cortisona, uma droga esteroide utilizada no tratamento de artrite reumatóide, sintetizada por E. Merck (Darmstadt, Alemanha) que necessitava de 31 passos e era muito ineficiente economicamente, resultando em um balanço que contabilizava 200 dólares por grama da substância. Com a utilização da biocatálise o número de passos foi diminuído para 5 e o valor por grama de produto passou a ser 6 dólares.<sup>6</sup> Este foi apenas um exemplo do potencial que os biocatalisadores apresentam frente à catálise industrial tradicional, haja visto que já em 2012 4000 enzimas eram conhecidas, das quais apenas 20 eram produzidas em escala industrial.<sup>13,14</sup> Mesmo nesse cenário as aplicações abrangiam diversas áreas industriais: têxtil, couro, processamento de alimentos e bebidas, detergentes, alimentação animal, produtos químicos refinados, produtos farmacêuticos, cosméticos e biodiesel com expectativa de crescimento, podendo chegar a cifras de 1.5 bilhões de dólares com enzimas de uso técnico (em especial a área de couro e bioetanol), 727 milhões com alimentação animal, 1,3 bilhões na indústria de alimentos, entre outras, chegando a um montante de 8 bilhões ao final de 2015.<sup>14</sup>

A aplicação de enzimas na indústria passa pela simples assistência à catalisadores convencionais, utilização das biomoléculas para encontrar rotas de síntese alternativas (diferentes rotas reacionais e matérias prima, gerando produtos iguais ou similares)<sup>6</sup> até a criação de metabolismos artificiais (Systems biocatalysis-Sys Biocat)<sup>15,16</sup> que, a exemplo do que ocorre nos organismos, conseguem obter produtos enantiomericamente puros por meio de uma cascata de reações interligadas por várias enzimas.

## 1.2 O CAMINHO

As proteínas evoluíram para atuar em ambientes celulares muito específicos, e quando expostas a condições diferentes podem sofrer uma grande variedade de processos de denaturação, que podem ser revertidos se removida a situação denaturante, ou podem acarretar perda irreversível da atividade.<sup>17</sup>

Vários agentes que promovem mudanças na configuração espacial e na atividade proteica são conhecidos, como por exemplo, temperatura, pH, agentes químicos ou força iônica. A inativação proteica segue um caminho reacional que inclui dissociação de subunidades em enzimas oligoméricas, denaturação de estruturas secundárias ou terciárias, agregação, coagulação e decomposição química.<sup>18</sup> Apesar de vários avanços na compreensão deste problema, há ainda muito que se compreender a fim de poder utilizar o potencial que esse conhecimento pode gerar.

Embora os dados sejam promissores, para que tenha aplicação industrial é fundamental que a enzima seja estável nas condições reacionais específicas de interesse, que podem ser extremas e distintas das encontradas em seu meio biológico. Dessa forma, cada tipo de síntese apresenta diferentes condições que podem causar instabilidade estrutural, favorecendo a denaturação protéica e a perda da atividade catalítica, como por exemplo, a exposição a meios não aquosos, como solventes orgânicos.

Ainda que existam enzimas e complexos multienzimáticos como lipases, esterases, desidrogenases, e aquelas responsáveis pelo metabolismo xenobiótico (desintoxicação), que são cataliticamente ativas em ambientes hidrofóbicos naturais, com eficiência similar ou até superior àquela em soluções aquosas,<sup>19-22</sup> via de regra, a estabilidade e/ou atividade de enzimas nestes sistemas é menor do que em meio aquoso.<sup>23</sup> Explicações para a inativação enzimática em solventes orgânicos baseiam-se na ideia de que a molécula proteica é rodeada por uma camada de hidratação ligada à superfície da enzima por ligações de hidrogênio. Esta camada de hidratação, ou uma parte dela, faz parte da estrutura da proteína e é essencial para a função enzimática.<sup>24,25</sup> Conseqüentemente, o deslocamento da camada aquosa por solvente orgânico resulta numa dramática mudança na estrutura da proteína, levando à denaturação. Então, solventes hidrofóbicos e solventes com alta capacidade de solvatação, como 1,4-dioxano, tetrahidrofurano e etanol, são fortes denaturantes, causando inativação enzimática a concentrações tão baixas quanto 10 a 30% em volume. Por outro lado, solventes hidrofílicos como glicerol, etileno glicol e formamida, podem ser usados a concentrações de 50 a 60%

sem provocar mudanças significativas na atividade.<sup>26</sup> Apesar de se saber que há atividade enzimática em meios não aquosos, e ser possível entender muitos dos mecanismos que estão associados a essa atividade, devido aos avanços nessa área de estudo, ainda há muito que se elucidar sobre esse tipo de catálise, que favorece a maior solubilidade de substratos não polares, o deslocamento do equilíbrio termodinâmico a favor da síntese, a supressão de reações paralelas aquo-dependentes, a diminuição da contaminação microbiológica e o aumento da seletividade, da especificidade e da velocidade de catálise sobre condições industriais.<sup>26</sup>

É importante, então, explorar a estrutura de misturas enzima-solventes orgânicos, tentando desvendar quais fatores, e de que maneira, influenciam a catálise em meio não-aquoso, além de aprofundar o conhecimento sobre o papel da água na catálise.

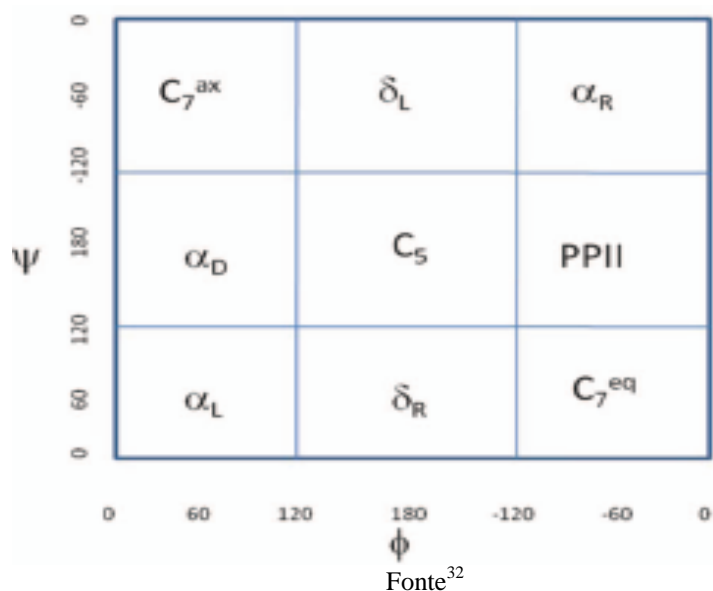
Durante o processo de enovelamento, existem impedimentos estéricos entre as cadeias laterais dos  $\alpha$ -aminoácidos formadores que não permitem certas conformações. Polipeptídeos de alanina são bons modelos por poderem acessar as principais conformações das proteínas (devido a sua cadeia lateral ser um grupamento metil) servindo como um ótimo modelo tanto para o estudo da influência da polaridade do solvente (e do co-solvente) sobre a ligação peptídica, quanto para descrever a qualidade dos diferentes campos de força ou níveis quânticos.

O diagrama de Ramachandran,<sup>1,2</sup> define os valores de ângulos diedros  $\phi$  e  $\psi$  permitidos para os planos formados ao redor da ligação peptídica (que giram em  $C_{\alpha}$ -C e N- $C_{\alpha}$ , respectivamente) e classifica em 9 as regiões conformacionais que podem ser acessadas. Dentre estas, foi identificado experimentalmente que para os polipeptídeos de alanina, a ordem de estabilidade é, em primeiro lugar polyproline II (PPII), seguido pelas formas  $\beta$  e  $\alpha_R$ ,<sup>27</sup> embora no caso do dipeptídeo a forma  $\beta$  sofra uma mudança, fazendo com que as conformações de equilíbrio sejam PPII e  $\alpha_R$  (Figuras 5 e 6).

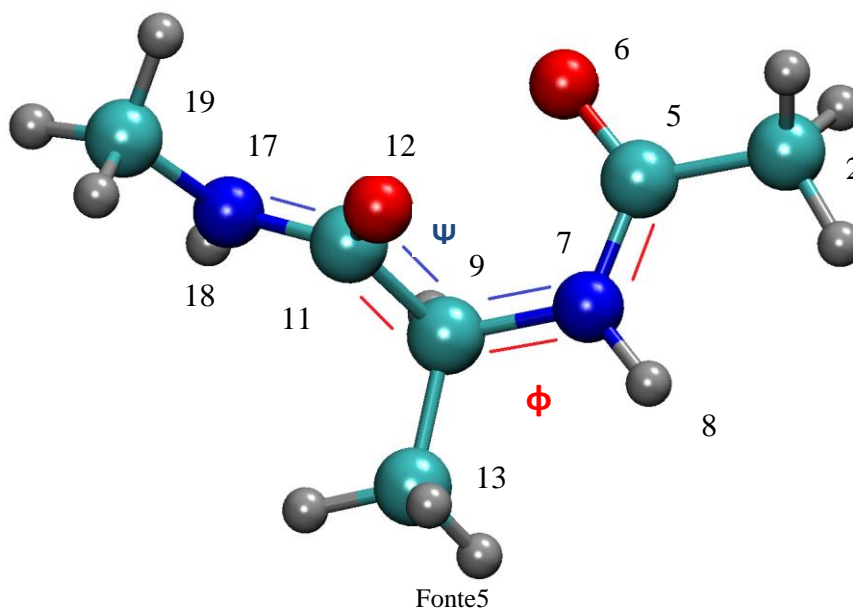
Experimentalmente, as condições impostas ao sistema (especialmente temperatura) podem modificar o panorama das estruturas de equilíbrio, como foi demonstrado em estudos de tripeptídeo de alanina onde a presença de alcoóis como etanol e glicerol, assim como o aumento da temperatura, aumentavam a população de PPII.<sup>28,29</sup> Não se sabe, de qualquer maneira, qual o motivo por trás deste comportamento. Alguns autores indicam que a água estabiliza a estrutura PPII por meio de ligações de hidrogênio entre o carbono adjacente e o grupo amina,<sup>30</sup> enquanto que outros autores propõe que os impedimentos estéricos seriam a força determinante para definir as preferências conformacionais.<sup>31</sup>

A seguir, são apresentadas as figuras relativas ao diagrama de Ramachandran e dos dois conforméros que foram estudados, PPII e  $\alpha_R$

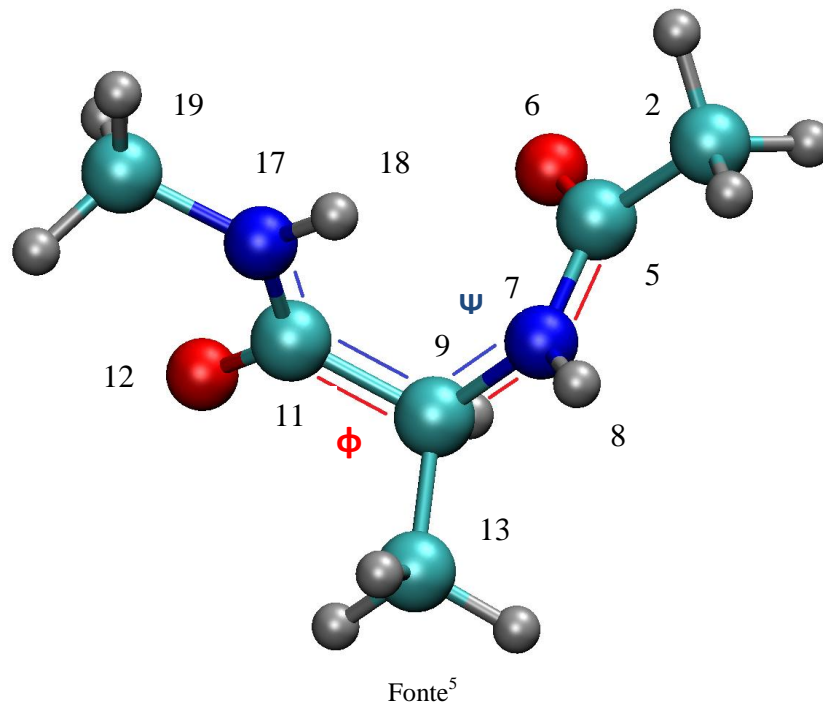
**Figura 4 - Diagrama de Ramachandran indicando os 9 possíveis conforméros de acordo com os impedimentos estéricos.**



**Figura 2 - Conformero PPII definido pelos ângulos  $\phi$  e  $\psi$  com numeração atômica: Cores: hidrogênios (cinza), carbonos (azul claro), nitrogênio (azul escuro), oxigênio (vermelho).**



**Figura 6 - Confórmero  $\alpha R$  definido pelos ângulos  $\phi$  e  $\Psi$  com numeração atômica: Cores: hidrogênios (cinza), carbonos (azul claro), nitrogênio (azul escuro), oxigênio (vermelho).**



### 1.3 O VEÍCULO

Apesar de ensaios biológicos serem indispensáveis, a química computacional é uma excelente ferramenta para auxiliar na compreensão de sistemas enzimáticos, especialmente os métodos QM/MM (mecânica quântica/mecânica molecular), que permitem ao mesmo tempo a descrição eletrônica detalhada do soluto (tratado mecânico-quanticamente) e a representação do solvente, guiada por funções de energia potencial denominadas campos de força (tratado por mecânica molecular), permitindo modelos mais próximos ao comportamento nativo da estrutura tridimensional da enzima imersa em determinado solvente. A utilização de métodos QM/MM aliados a métodos biofísicos não só elucidam fatores responsáveis pelo melhoramento da atividade catalítica, como também ajudam a compreender os meandros dos mecanismos enzimáticos em geral,<sup>7</sup> uma vez que permitem o acesso às informações não disponíveis experimentalmente. Desde 1976, com o trabalho pioneiro de Warshel e Levitt,<sup>33</sup> a aplicação de métodos QM/MM no estudo enzimático vem demonstrando seu valor, estabelecendo-se como ferramenta importante para a enzimologia. Prova disso é que desde

2011 mais de 100 estudos são publicados anualmente (com um montante de 605 trabalhos no total).<sup>34</sup>

A abordagem proposta para este trabalho segue essa linha de raciocínio, quer seja, utilizar um método computacional baseado em QM/MM para estudar dinâmica de solvatação de peptídeos em meios aquoso e não aquoso, a fim de contribuir na elucidação do comportamento enzimático em função do ambiente em que se encontre, na busca de compreensão sobre os aspectos que permitem manter a atividade enzimática em diferentes condições de trabalho. Para isso, foram escolhidos o método QM/MM ASEP/MD (*Averaged electrostatic potential from molecular dynamics*, Potencial eletrostático médio proveniente de dinâmica molecular)<sup>35-37</sup> e o sistema-modelo, dipeptídeo de alanina. A utilização de polipeptídeos ao invés de enzimas propriamente ditas é um recurso muito comum na química computacional, especialmente quando o nível do cálculo quântico aplicado ao sistema é muito alto, como é o caso deste trabalho, em razão do enorme tempo de cálculo envolvido e espaço de armazenamento.

Para a dissertação de mestrado foi utilizado o método estocástico de simulação computacional Monte Carlo<sup>38</sup> e a N-metilformamida como sistema modelo de ligação peptídica, o que deu origem à publicação de dois trabalhos.<sup>39-40</sup> Agora, para a tese de doutorado, estudou-se a atividade de peptídeos de fato como sistemas-modelo, para ser possível a análise dos gráficos de Ramachandran<sup>1,2</sup> além de ter sido utilizada uma metodologia que permite a análise temporal do sistema, de natureza QM/MM, que aumenta o número de informações disponíveis para a compreensão do mesmo problema proposto no mestrado, a compreensão do comportamento enzimático frente a diferentes solventes. No apêndice inseriu-se o trabalho submetido à publicação no JCTC (*journal of chemical theory and computation*) com o título: *Conformational Changes of the Alanine Dipeptide in Water-Ethanol Binary Mixtures* e número de identificação ct-2015009523.

## Capítulo 2 – O VEÍCULO

### SIMULAÇÃO COMPUTACIONAL

A química computacional vêm revolucionando a maneira em lidar com problemas que antes só podiam ser explorados por métodos experimentais, isso porque ela permite tanto o acesso a dados que não estão acessíveis experimentalmente, quanto a predição de novas propriedades ou comportamentos, oferecendo excelentes subsídios para o estudo da complexidade de sistemas químicos.<sup>41</sup>

O desenvolvimento dos métodos utilizados neste âmbito acompanha a evolução tecnológica, de forma que o aumento da capacidade computacional permite além de cálculos com maior número de moléculas e detalhamento, também, o desenvolvimento de novas metodologias teóricas com o intuito de sanar as dificuldades em simular sistemas macroscópicos. Essas dificuldades se relacionam principalmente à qualidade de descrição dos átomos, número de moléculas e tempo de simulação dos eventos dentro do sistema, de maneira que as diferentes metodologias propostas buscam o balanço entre custo computacional e nível de detalhamento.

Dentro deste cenário, o ideal seria tratar todos os átomos do sistema por meio da resolução da equação de Schrödinger,<sup>41</sup> respeitando a natureza quanto mecânica destes. Os métodos *ab initio* e *semiempíricos*<sup>41,42</sup> lançam mão desta abordagem, com a diferença em que no caso dos métodos *semiempíricos* há a inclusão de dados obtidos experimentalmente (empíricos) nos cálculos, enquanto que nos métodos *ab initio* todas as propriedades dos sistema são calculadas do início, ou seja, com exceção das constantes físicas<sup>1</sup> fundamentais ( $e$ ,  $c$ ,  $h$  e  $m$ ) não são utilizados dados empíricos do sistema para se determinar as propriedades físicas desejadas; elas são calculadas totalmente por meio da resolução da equação de Schrödinger, na qual se emprega aproximações adequadas (uma vez que não há, para o caso de sistemas com mais de um elétron a resolução exata da equação).

Todavia, o cálculo envolvendo muitos átomos torna-se proibitivo devido ao alto custo computacional (tanto em tempo de cálculo, quanto em relação à memória necessária

---

<sup>1</sup> **e, c, h, m:**, carga elementar, velocidade da luz no vácuo, constante de Planck, massa.

para armazenamento), portanto, caso haja interesse em simular sistemas maiores uma alternativa é a utilização da mecânica estatística aliada a potenciais clássicos, onde não há a resolução eletrônica do sistema. Nessa classe, podemos citar como metodologias Monte Carlo (estocástico) e Dinâmica Molecular (determinístico).<sup>43</sup>

O método Monte Carlo move os átomos dentro da caixa de simulação aleatoriamente, os movimentos que diminuam a energia potencial total do sistema são mantidos e aqueles que a façam aumentar podem ser mantidos, ou não. Ao final da simulação é obtida a caixa termodinamicamente equilibrada da qual serão extraídas as propriedades desejadas pela aplicação da mecânica estatística. A dinâmica molecular utiliza os mesmo potenciais clássicos e também lança mão da mecânica estatística para a determinação das propriedades do sistema em equilíbrio, porém neste caso, a evolução temporal (trajetórias atômicas no tempo) é descrita por meio da integração das equações de movimento de Newton (correspondente clássica à equação de Schrödinger dependente do tempo, na mecânica quântica).

Uma abordagem alternativa, quando há o interesse em se estudar sistemas grandes e ao mesmo tempo determinar propriedades eletrônicas destes, são os métodos híbridos, que lançam mão de mecânica quântica para um número reduzido de átomos do sistema enquanto que o restante é tratado por cargas elétricas pontuais ou campos de força clássicos, por isso ganham a denominação de QM/MM, e é exatamente esta, a abordagem utilizada neste trabalho. Existem diversas maneiras de aplicar os métodos QM/MM ao sistema de estudo, aqui, apenas a molécula do soluto foi tratada por mecânica quântica, enquanto que o solvente foi tratado por dinâmica molecular. Mais detalhes sobre o método QM/MM utilizado serão dados na secção 2.4, antes será apresentada um breve introdução sobre mecânica molecular e os campos de força que são os responsáveis por determinar a correlação entre os pares de átomos nesses métodos (secção 2.1), em 2.2 a teoria que rege a dinâmica molecular e em 2.3, uma visão geral de métodos QM/MM mais comuns.

## 2.1 MECÂNICA MOLECULAR E CAMPOS DE FORÇA

Os métodos de mecânica molecular são de natureza empírica com parâmetros ajustados para reproduzir dados experimentais, como por exemplo, calor de vaporização e/ou densidade do material. Para isso utilizam-se ensembles estatísticos, como NpT ou NVT, em

que as condições nas quais ocorrem as simulações são mantidas constantes, permitindo acessar um conjunto de microestados capaz de representar o macroestado desejado, onde, para cada tipo de ensemble existirá uma respectiva distribuição de probabilidade, que permitirá ao programa o cálculo das propriedades médias do sistema.

O tratamento clássico de sistemas microscópicos lança mão de um campo de força empírico (que é na realidade uma função da energia potencial), que permite que seja calculada a energia potencial total do sistema,  $V(r)$ . A função que descreve  $V(r)$  é uma somatória de vários termos de energia, incluindo os termos para átomos ligados (comprimentos e ângulos de ligação, ângulos diedros) e os termos para átomos não-ligados (interações de van der Waals e de Coulomb).<sup>44-45</sup>

A portabilidade de campos de força é um fator determinante para a boa representação clássica de sistemas microscópicos, devendo ser parametrizados de acordo com o sistema que se deseja estudar. Em geral, um campo de força pode ser representado pela seguinte equação:

$$V_{total} = V_{ligante} + V_{não-ligante} \quad (1)$$

que pode ser escrita da seguinte maneira:

$$V_{total} = V_d + V_\theta + V_\varphi + V_{LJ} + V_{coulomb} \quad (2)$$

onde,

$$\left. \begin{array}{l} V_d = \text{Potencial Harmônico Linear} \\ V_\theta = \text{Potencial Angular} \\ V_\varphi = \text{Potencial Torcional} \end{array} \right\} V_{ligante}$$

$$\left. \begin{array}{l} V_{LJ} = \text{Potencial de Lennard-Jones} \\ V_C = \text{Potencial de Coulomb} \end{array} \right\} V_{não-ligante}$$

Outros termos podem ser incluídos na determinação do potencial, como potenciais cruzados entre ligações, ligações e ângulos ou ligações e torções em ângulos diedros, além de potenciais para descrever interações não harmônicas e ligações de hidrogênio, entre outros,

dependendo do tipo de campo de força utilizado. Entretanto, limitar-se-á a descrição do potencial em sua forma mais simples, representada na equação 2, mais detalhes de termos adicionais a estes podem ser encontrados na bibliografia.<sup>44-45</sup>

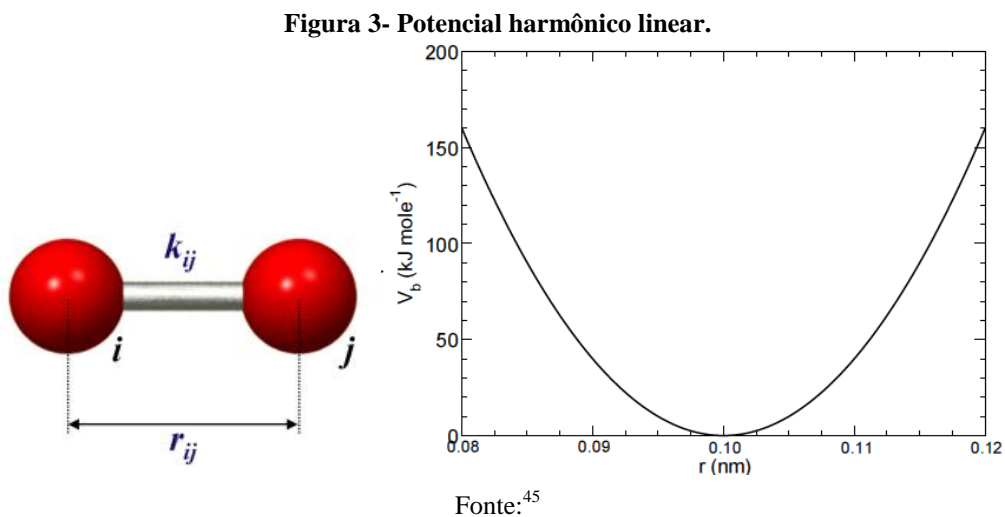
Os valores de referência e as constantes presentes nas funções que descrevem os termos do potencial são parâmetros inseridos nos campos de força e assumem valores determinados experimentalmente ou valores próximos a estes a fim de reproduzir o comportamento macroscópico dos arranjos atômicos.

### 2.1.1 Potenciais ligantes

- **Potencial Harmônico Linear ( $V_d$ )**

Representa a energia associada à variação na distância de ligação entre dois átomos ligados covalentemente, segundo a equação 3:

$$V_d = \frac{1}{2} K_d (d - d_0)^2 \quad (3)$$



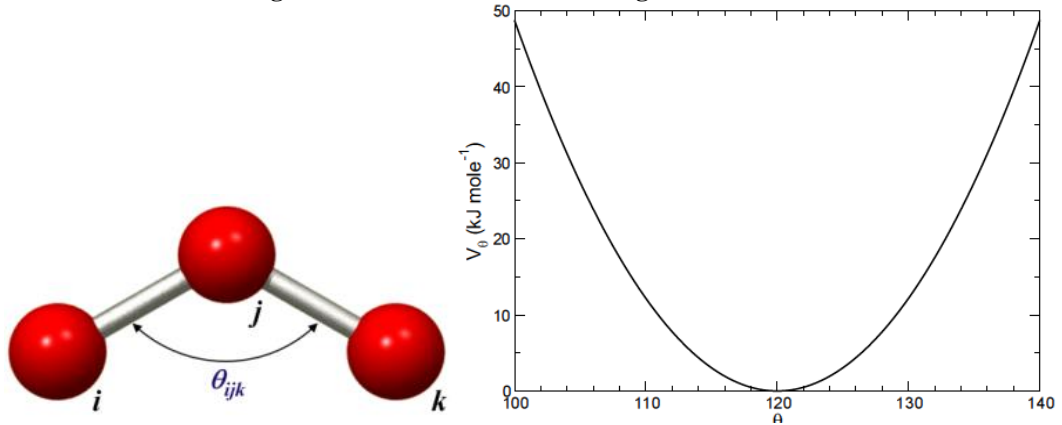
Onde,  $d$  é o comprimento de ligação entre dois átomos  $i$  e  $j$ ,  $d_0$  é valor de referência do comprimento de equilíbrio da ligação e  $K_d$  é a constante elástica da força.

- **Potencial Harmônico Angular ( $V_\theta$ )**

Representa a energia associada à variação dos ângulos de ligação entre três átomos ligados consecutiva e covalentemente, segundo a equação 4:

$$V_{\theta} = \frac{1}{2} K_{\theta} (\theta - \theta_0)^2 \quad (4)$$

Figura 4 - Potencial harmônico angular.



Fonte:<sup>44,45</sup>

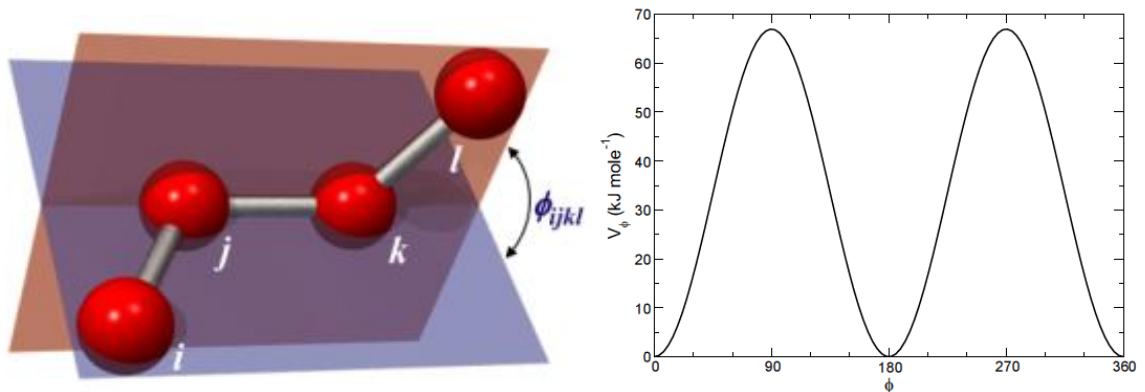
Onde,  $\theta$  é o ângulo definido pelos átomos i, j, k,  $\theta_0$  é o valor de referência do ângulo  $\theta$  em equilíbrio e  $K_{\theta}$  é a constante elástica da força.

#### • Potencial Torsional

Representa a energia associada à torção em uma das ligações covalentes onde quatro átomos encontram-se consecutivamente ligados, i,j,k,l, sendo gerado um potencial torsional em torno do eixo j,k, onde ocorre a torção. Esse potencial é dado pela equação 5:

$$V_{\phi(ijkl)} = K_{\phi} [1 + \cos(n\phi - \phi_s)] \quad (5)$$

Figura 5 - Potencial torsional.



Fonte:<sup>45</sup>

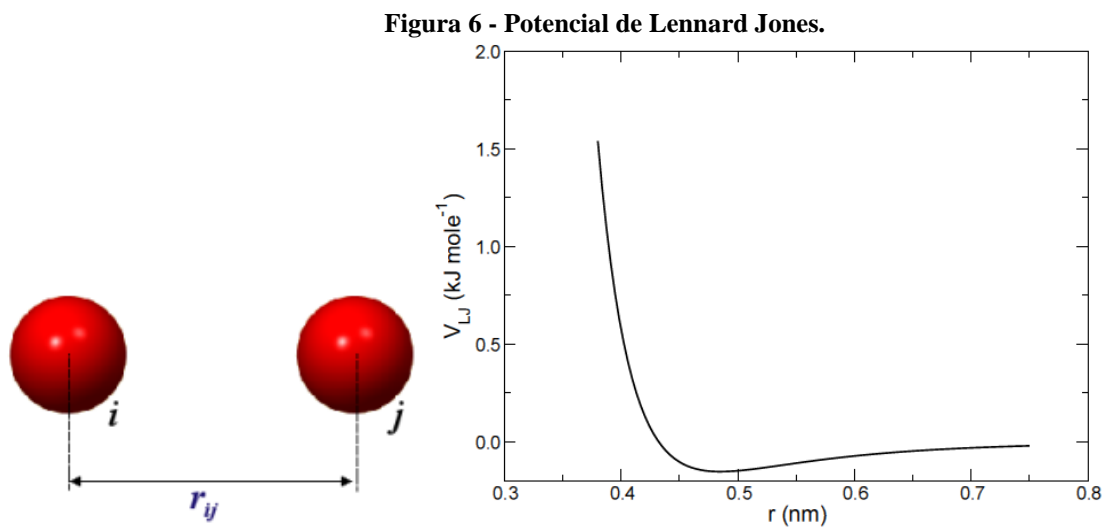
Onde  $K_\phi$  é constante que define a barreira de energia para que ocorra a torção,  $n$  é o número de máximos (ou mínimos) de energia em uma torção completa,  $\phi$  é o ângulo diedro de referência e  $\phi_s$  é o ângulo de fase (defasagem no ângulo diedro que pode gerar um ponto de mínimo ou de máximo na posição  $\phi = 0$ ).

### 2.1.2 Potenciais não ligantes

#### • Potencial de Lennard-Jones

Representa as energias decorrentes das interações por *van der Waals*, segundo a equação 6:

$$V_{ij} = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (6)$$



Fonte:<sup>45</sup>

Onde,  $\varepsilon$  representa o mínimo de energia do poço de potencial (ponto em que ocorre a máxima interação entre os átomos),  $\sigma$  se refere à média aritmética dos raios dos átomos envolvidos no cálculo, e,  $r_{ij}$  simboliza a distância entre o átomo  $i$  da molécula  $a$ , e, o átomo  $j$  da molécula  $b$ , de forma que o argumento que está elevado a 12 refere-se ao potencial repulsivo que ocorre quando há sobreposição de densidade eletrônica com a aproximação dos átomos  $i$  e  $j$ . Já o argumento elevado a 6, descreve o potencial atrativo decorrente da formação de dipolos induzidos, instantâneos, originados por flutuações na densidade de carga eletrônica dos átomos.

### • Potencial de Coulomb

Representa a energia associada às interações eletrostáticas entre os átomos, segundo a equação 7:

$$V_{elet} = \frac{q_i q_j}{4\pi\epsilon r_{ij}} \quad (7)$$

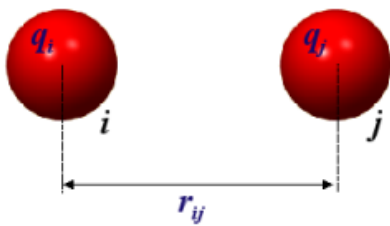
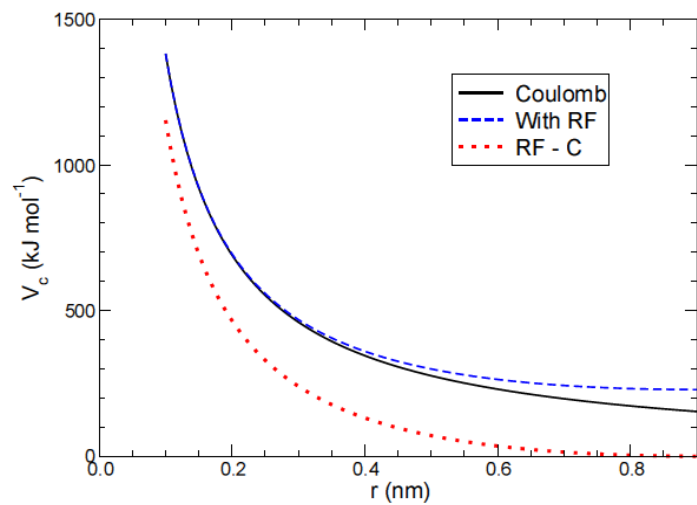


Figura 7 - Potencial de Coulomb.



Fonte:<sup>45</sup>

Onde,  $r_{ij}$  simboliza a distância entre o átomo  $i$  e o átomo  $j$ ,  $q_i q_j$  representam as cargas dos respectivos átomos,  $\epsilon_0$  representa a constante dielétrica do vácuo.

A partir do modelo molecular é possível por meio da mecânica estatística, utilizando-se equações físico-matemáticas e um método adequado, obter propriedades macroscópicas do sistema. As equações que são utilizadas em geral dependem do *ensemble* utilizado no sistema em questão.

Aqui, é importante destacar que a energia potencial do sistema é calculada a partir de interações de pares efetiva, ou seja, apenas as correlações entre pares de átomos são computados explicitamente, a função que soma todos os termos da energia potencial é truncada neste ponto, de forma que os termos que correlacionam interações entre 3 ou mais corpos será inserida no potencial de pares de maneira indireta, na tentativa de obter um campo de força que calcule a energia potencial total do sistema de maneira mais eficiente computacionalmente, sem deixar de lado as interações entre vários corpos.

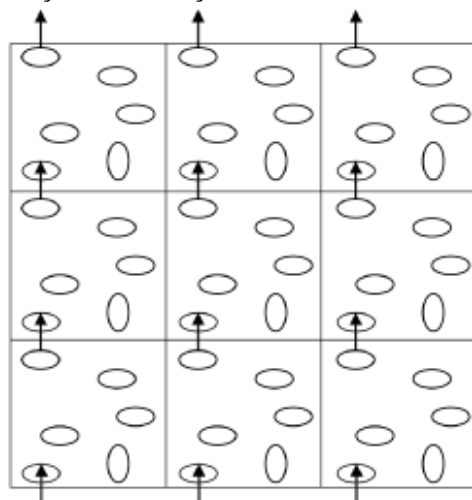
### 2.1.3 Condições periódicas de contorno

Mesmo com o avanço computacional, tentar simular o número real de átomos contidos no sistema não só é muito caro computacionalmente, como também desnecessário, tendo em vista que ótimos modelos propostos computacionalmente reproduzem com fidelidade dados experimentais mesmo utilizando um número reduzido de átomos. Isso é possível graças a aplicação das condições periódicas de contorno. A caixa onde o sistema será simulado é replicada infinitamente no espaço (eixos x, y e z) de forma que quando uma molécula se move na caixa central, suas imagens periódicas também se movem nas caixas replicadas, ou seja, caso uma molécula saia da caixa por um lado, outra entrará pelo lado oposto, eliminando os efeitos de superfície e mantendo o número de moléculas dentro da caixa constante, porém sem restrições de movimento devido aos limites da caixa (Figura 8).

Como dito na secção 2.1.2 a função de energia potencial não-ligante aplicada à simulação molecular neste trabalho é dada pelo potencial de pares efetivo, que considera apenas as interações intermoleculares por meio da soma dos potenciais clássicos de *Lennard-Jones* e *Coulomb*, segundo a equação 14:

$$U_{ij}(\mathbf{X}) = 4\epsilon_o \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \quad (14)$$

Figura 8 - Simulação das condições de contorno utilizadas no método.

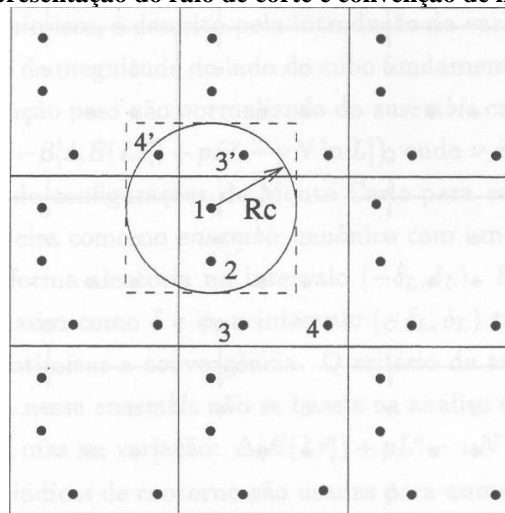


Fonte<sup>43</sup>

Porém, como as caixas são replicadas infinitamente, há a necessidade de se limitar o cálculo dessa energia, caso contrário haveriam infinitos termos a serem considerados o que tornaria os cálculos impraticáveis além de refletir uma periodicidade estrutural artificialmente criada. Para isso é utilizado o raio de corte (círculo na Figura 13), para que apenas as interações mais significativas sejam consideradas.

É importante salientar que, para não haver a possibilidade da molécula interagir com sua própria imagem, ou, com uma mesma molécula duas vezes, o raio de corte deve ter uma magnitude máxima de  $L/2$ , onde  $L$  é o tamanho da aresta da caixa de simulação.

**Figura 9 - Representação do raio de corte e convenção de imagem mínima.**



Fonte<sup>43</sup>

## 2.2 MÉTODO DE DINÂMICA MOLECULAR

Um dos métodos que utiliza funções de energia potencial para o tratamento clássico no estudo de sistemas microscópicos é a dinâmica molecular. Cada átomo dentro da caixa de simulação apresenta uma trajetória definida por meio da evolução temporal do sistema, baseada nas integrais das equações clássicas de movimento de Newton. Caso a molécula não seja considerada rígida, torna-se necessário a inclusão de termos ao potencial referentes às vibrações moleculares, torções em ângulos de ligação e diédricos, entre outros. Na maioria dos campos de força estas contribuições para o potencial total do sistema são representadas por funções harmônicas, tais como, estiramento de ligações e ângulos de ligações, e, por série de Fourier, como as energias de torção de ângulos diédricos. Esse potencial, portanto, é

dividido em partes, cada qual construída e parametrizada para representar aspectos da realidade física do problema e possui valores específicos do campo de força escolhido.

Uma vez definido o potencial total de interação para o sistema, é possível calcular as forças que agem sobre cada átomo a partir da derivada primeira da energia potencial em relação às posições atômicas, segundo a equação :

$$F_i(t) = - \frac{\nabla V(r_1, r_2, \dots, r_n)}{\partial r_i} \quad (8)$$

Onde,  $F_i(t)$  é a força sobre o átomo  $i$  no instante  $t$ ,  $\nabla V (r_1, r_2, \dots, r_n)$  é o potencial total de interação, e  $r_i$  é a posição atômica do átomo  $i$ .

Dessa forma, a partir da força que age sobre um determinado átomo em um instante  $t$ , é possível saber qual é a aceleração que age sobre esse átomo nesse mesmo instante, de acordo com a equação de Newton que em casos onde a massa é constante pode ser escrita de acordo com a equação:

$$F_i(t) = m_i \frac{d^2 r_i}{dt^2} \quad (9)$$

Uma vez determinada a aceleração sobre cada partícula, e considerando que essa seja constante no intervalo de tempo entre  $t$  e  $t + dt$ , pode-se determinar sua velocidade a partir da integração das equações de movimento. Dessa forma, a partir da integração da velocidade, é possível calcular a mudança de posição de cada átomo. Sabendo-se as novas velocidades e posições para cada partícula, são obtidas as energias potenciais e cinéticas do sistema. Por fim, esse processo se repete ao longo de toda a simulação e o conjunto de diferentes velocidades e posições para cada átomo são armazenadas, gerando as trajetórias moleculares. Nesse ponto, vale a pena ressaltar que como não existe solução analítica para a resolução das equações de movimento, diferentes métodos numéricos são utilizados para se fazer essa integração. O método mais conhecido é o do algoritmo de Verlet, do qual vários outros métodos derivam, que utiliza as posições e acelerações no tempo  $t$  e as posições do passo anterior,  $r(t-\Delta t)$  para calcular as posições em  $t + \Delta t$ , de acordo com as equações abaixo:

$$r(t + \Delta t) = r(t) + \Delta t v(t) + \frac{1}{2} \Delta t^2 a(t) + \dots \quad (10)$$

$$r(t - \Delta t) = r(t) - \Delta t v(t) + \frac{1}{2} \Delta t^2 a(t) - \dots \quad (11)$$

Somando as equações (10) e (11), obtemos:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \Delta t^2 a(t) \quad (12)$$

Os valores das velocidades não são calculados automaticamente no algoritmo de Verlet, o que implica em um menor custo computacional. Porém, caso se deseje obter essas velocidades, existem diversas maneiras, como por exemplo dividir a diferença entre  $t + \Delta t$  e  $t - \Delta t$  por  $2 \Delta t$ :

$$v(t) = \frac{[r(t + \Delta t) - r(t - \Delta t)]}{2\Delta t} \quad (13)$$

O algoritmo de integração utilizado neste trabalho foi o *Leap Frog* que vem implementado para o programa *Gromacs*. Esse algoritmo calcula as velocidades na metade do tempo de integração e usa essas velocidades para encontrar as novas posições, segundo as equações:<sup>43</sup>

$$v\left(t + \frac{1}{2} \Delta t\right) = v\left(t - \frac{1}{2} \Delta t\right) + \Delta t a(t) \quad (14)$$

$$r(t + \Delta t) = r(t) + \Delta t v\left(t + \frac{1}{2} \Delta t\right) \quad (15)$$

O intervalo de integração,  $\Delta t$ , deve ser menor do que o menor grau de liberdade que existe no sistema, normalmente esse valor corresponde a vibração do átomo de hidrogênio (por ser leve em relação aos outros átomos) que é  $\Delta t = 1\text{fs}$ .

### 2.3 INTRODUÇÃO À QM/MM

A utilização de potenciais clássicos para determinar o comportamento atômico é muito útil para estudos na química computacional; porém desde o advento da mecânica quântica, busca-se a implementação de métodos computacionais que permitam o estudo de sistemas microscópicos de acordo com sua natureza física dualística. No caso de sistemas em fase gasosa essa implementação é facilitada, considerando-se a aproximação de não interação

dos constituintes do fluido. Situação semelhante ocorre no estudo de sistemas em fase sólida que, embora não possam ter suas interações negligenciadas, têm suas interações simplificadas e podem ser estudados neste nível devido à periodicidade na disposição atômica dentro dos cristais. O problema para a implementação da mecânica quântica em métodos computacionais ocorre quando esse sistema envolve líquidos, isso porque os líquidos possuem características intermediárias entre os estados gasoso e sólido, impossibilitando as aproximações de não interação dos constituintes e por simetria, uma vez que possuem densidades altas (comparado à fase gasosa) e estrutura desorganizada (comparado ao estado sólido). Além disso, devido às suas características, existem diferentes configurações microscópicas que conferem o mesmo estado macroscópico em determinada temperatura, culminando em uma natureza mecânico estatística inerente ao sistema.

Aliando isto ao problema em ter de lidar com muitos átomos, sistemas onde há moléculas imersas em solvente ou mesmo moléculas muito grandes deram origem à implementação de métodos híbridos, chamados QM/MM, que tratam parte do sistema com mecânica molecular e parte com mecânica quântica.

O restante da secção 2.3 tem o intuito de situar o leitor sobre as principais alternativas encontradas dentro da metodologia QM/MM, no entanto, cabe ressaltar que neste trabalho limita-se a apresentação apenas ao suficiente para que o leitor esteja a par das principais abordagens e compreenda a discussão sobre o método utilizado pelo programa ASEP feita no capítulo seguinte. Portanto, para maiores detalhes utilizar as referências que serviram como base para esta secção.<sup>46,47</sup>

### **2.3.1 - QM/MM convencional**

Em QM/MM convencional o sistema é subdividido em diferentes regiões que são tratadas ao mesmo tempo com níveis de cálculo distintos, indo do clássico ao quântico, de acordo com o interesse na região de estudo. Nesse sentido, mesmo as regiões consideradas quanticamente podem ser tratadas com cálculos mais ou menos robustos, analogamente ao que se propõe para a divisão entre a parte clássica e quântica.

Esse sistema subdividido será regido por um Hamiltoniano efetivo que pode ser escrito como segue, na equação 15:

$$H_{eff} = H_{QM} + H_{MM} + H_{QM/MM} \quad (16)$$

Onde,  $H_{eff}$  representa o Hamiltoniano efetivo,  $H_{QM}$  o sistema tratado quanticamente,  $H_{MM}$  o sistema tratado classicamente, e,  $H_{QM/MM}$  a interface entre os dois sistemas. O termo que irá diferenciar os métodos de QM/MM convencionais é justamente o Hamiltoniano da interação entre a parte clássica e quântica ( $H_{QM/MM}$ ). Além do componente eletrostático do sistema poder ser acoplado de maneira clássica (acoplamento mecânico) ou de maneira quântica (acoplamento eletrostático), outra questão que diferencia os métodos é o ponto de intersecção entre a região de  $H_{QM}$  e a região de  $H_{MM}$ . Há duas possibilidades para essa divisão, no primeiro caso há a separação das moléculas inteiras (i.e. uma molécula pequena imersa em um solvente, onde a molécula é tratada quanticamente e o solvente classicamente), enquanto que no segundo caso a divisão acontece dentro da própria molécula (i.e. uma proteína solvatada tem, por exemplo, seu sítio ativo tratado quanticamente, enquanto que o restante da molécula é tratado classicamente). Ao contrário do que ocorre na primeira situação há o rompimento de ligações covalentes, o que resulta em complicações adicionais ao cálculo.

Na abordagem em que as moléculas inteiras são tratadas quanticamente e não há, portanto, rompimento de ligações covalentes, a interação entre a parte clássica e quântica é feita apenas por interações intermoleculares, o que facilita os cálculos. De acordo com a maneira que os termos intermoleculares são representados pode-se ter a situação em que não há polarização nem do solvente pelo soluto, nem do soluto pelo solvente (interação despolarizada):

$$H_{QM/MM} = \sum_a \sum_b 4\epsilon_{ab} \left[ \left( \frac{\sigma_{ab}}{r_{ab}} \right)^{12} - \left( \frac{\sigma_{ab}}{r_{ab}} \right)^6 \right] + \sum_m \sum_b \frac{Z_m q_b}{r_{mb}} - \sum_i \sum_b \frac{q_b}{r_{ib}}, \quad (17)$$

Nesse caso, apesar da interação entre a parte clássica e quântica, não há influencia de uma sobre a outra. Uma situação alternativa é aquela em que há polarização do soluto pelo solvente (através do acoplamento eletrostático) mas não o contrário, ou o caso onde há acoplamento total entre as regiões clássica e quântica, sendo o soluto polarizado pelo solvente e vice-versa. Para isso, mais termos são adicionados à equação 16, de maneira que a polarização do soluto pelo solvente, ou entre ambos seja descrita corretamente. Porém cada novo termo adicionado implica, além de uma maior detalhamento do sistema, um aumento no

custo computacional. Dessa maneira, a situação mais comum em estudos por QM/MM convencional em que as moléculas são tratadas inteiras, é o caso em que o soluto é polarizado pelo solvente mas não o inverso, por apresentar um ganho considerável para o modelo pela representação da polarizabilidade do soluto, sem aumentar de maneira expressiva o custo computacional.

No entanto pode haver a necessidade da divisão de uma parte específica da molécula (no caso de enzimas, por exemplo, caso deseje-se estudar seu sítio ativo) implicando no rompimento de ligações covalentes. Isso leva a um problema mais complicado, por envolver ligações químicas de fato, ao invés de interações intermoleculares. Neste caso, há duas principais abordagens para realizar a interação entre a parte quântica e clássica: i) A molécula é cortada em alguma ligação pouco polar e pouco polarizável (como uma ligação simples C-C) e adiciona-se um átomo que ligue as duas partes, denominado átomo de ligação (*link atom*); ii) A aproximação dos orbitais congelados, que tenta resolver o maior problema dos átomos de ligação, que é a representação pontual das cargas nas proximidades da fronteira entre os sistemas. Neste caso, os orbitais congelados tentam representar essas cargas de maneira mais real - uma distribuição de cargas conseguidas quanticamente, que deve permanecer fixa uma vez calculada, sem sofrer polarização pelos átomos do sistema MM da fronteira.

### 2.3.2 QM/MM sequencial

Os métodos de QM/MM sequenciais são baseados no tratamento do sistema em duas etapas, na primeira todo o sistema é tratado classicamente, depois, faz-se um cálculo quântico na parte do sistema em que se deseja obter as características eletrônicas. A abordagem em duas etapas proposta pelo QM/MM sequencial elimina alguns problemas da abordagem QM/MM convencional, como por exemplo, obter resultados que representem médias estatisticamente convergidas, o que permite a análise das propriedades do soluto de maneira mais completa ao final da simulação, pois, ao contrário do que ocorre nos métodos convencionais, é possível, caso se deseje, obter dados sobre uma propriedade que não fora definida inicialmente sem a necessidade de um novo cálculo quântico. Para isso é necessário que o método sequencial leve em conta apenas termos estatisticamente decorrelacionados para a convergência de energia do sistema.

No entanto, representar a polarização do soluto pelo solvente por meio da abordagem sequencial é um problema maior do que o encontrado para a abordagem convencional, tendo em vista que nesse caso a parte clássica e quântica são realizadas em diferentes etapas, o que não permite o acoplamento imediato das duas partes para a determinação da polarização do soluto (como acontece no método convencional).

O método utilizado neste trabalho lança mão exatamente do tratamento QM/MM sequencial, por isso, a secção seguinte será dedicada a explicar de maneira mais detalhada tal abordagem com a descrição da teoria e fluxograma do próprio método.

## 2.4 MÉTODO ASEP/MD

A evolução computacional das últimas décadas permitiu um aprofundamento nos métodos focados em entender e prever como a estrutura, propriedades, reatividade e comportamento das moléculas são afetados pelo meio em que elas se encontram inseridas.<sup>48</sup> O que se deseja é que cada vez mais os cálculos realizados para sistemas em soluções se aproximem da qualidade de descrição encontrada para os métodos quânticos em fase gasosa. No entanto, a adição de solventes (de qualquer natureza) ao sistema implica em alguns problemas práticos adicionais em relação a estudos análogos em fase gasosa. Entre eles podemos destacar:<sup>49</sup>

1. O grande número de moléculas envolvidas na descrição dos efeitos de polarização do solvente. Molécula localizada a longas distâncias tem efeitos não negligenciáveis nas propriedades do soluto. Em geral, no estudo de efeitos de solvente, e dependendo dos tipos de interações soluto-solvente envolvidas, se faz necessário incluir diversas camadas de solvatação nos cálculos.

2. A possibilidade de interações específicas, principalmente ligações de hidrogênio, entre as moléculas de soluto e solvente presentes na primeira camada de solvatação. A descrição adequada destas interações torna obrigatória a utilização de modelos de solventes microscópicos.

3. Finalmente, existe um grande número de configurações de soluto-solvente termicamente acessíveis em solução. Diferentes moléculas de soluto estarão inseridas em diferentes ambientes e conseqüentemente, terão propriedades ligeiramente diferentes. Sendo

assim, para se conseguir resultados estatisticamente significativos é necessário que se incluam centenas ou milhares de diferentes configurações.

Uma dificuldade adicional advém de que em solução a energia relevante a ser considerada é a energia livre e é, portanto, necessário que tenhamos à disposição métodos que permitam o cálculo dessa quantidade de maneira efetiva e computacionalmente factível. Todos esses problemas: grande número de moléculas de solvente, possível existência de interações específicas, grande número de diferentes configurações do solvente, necessidade de determinar as diferenças de energia livre, e assim por diante, têm como consequência um enorme aumento do custo computacional, no estudo de efeitos de solvente.

Uma alternativa para amenizar esse custo computacional é a utilização de métodos híbridos de simulação computacional. Ainda assim, na maior parte das abordagens desse tipo há que se resolver a equação de Schrödinger do soluto para cada nova configuração do solvente, o que implica em milhares de cálculos quânticos e acaba por limitar a qualidade da descrição do sistema em dois principais aspectos: número de configurações incluídas e nível do cálculo quântico, devido ao grande custo computacional associado a eles.

O ASEP/MD,<sup>29-31</sup> acrônimo para *Averaged Solvent Electrostatic Potencial from Molecular Dynamics* (Potencial Eletrostático Médio do Solvente proveniente de Dinâmica Molecular) é um método de QM/MM sequencial que lança mão de MFA,<sup>50,51</sup> *Mean Field APPROXimation* (aproximação de campo médio) para tentar sanar essas dificuldades, aliando um método de *continuum* ao QM/MM para estudar os efeitos de solvente. Como nos métodos de *continuum*<sup>52-55</sup> ou dipolo de Langevin,<sup>56</sup> a perturbação gerada pelo solvente é introduzida na Hamiltoniana do soluto por meio de uma função de distribuição contínua, sendo que, no ASEP/MD, essa função é obtida através de simulações de dinâmica molecular.

O método desenvolvido pelo grupo de pesquisa QCAMM (Quantum Chemistry and Molecular Modeling) da Universidade de Extremadura, na cidade de Badajoz-Espanha, serve como interface, garantindo fluidez no acoplamento de programas de dinâmica molecular e cálculos quânticos (no caso deste trabalho, GROMACS<sup>45</sup> e GAUSSIAN,<sup>43</sup> respectivamente) reduzindo drasticamente o número de cálculos por meio da utilização de MFA, o que permite uma maior atenção à descrição eletrônica do soluto sem perdas consideráveis na descrição do solvente. Em resumo, o método ASEP/MD é ideal para sistemas onde se deseja obter simultaneamente uma boa descrição das correlações eletrônicas do soluto e das interações

entre soluto e solvente, com um custo computacional reduzido em relação aos métodos mais tradicionais de QM/MM, permitindo além de estudos como o que será apresentado neste trabalho, sobre efeitos de solvente, estudos de reatividade química<sup>57</sup> e espectro VIS-UV.<sup>58-62</sup>

#### 2.4.1 A Aproximação de campo médio (MFA)

Um grande dilema, no campo da química teórica é estabelecer a medida ideal para balancear o detalhamento na descrição do sistema ao custo computacional despendido, tanto em tempo de cálculo quanto em espaço de armazenamento. De maneira geral, o aumento no número de átomos, vem acompanhado da diminuição no nível de detalhamento na descrição do sistema (como explicado anteriormente). Nesse sentido, ainda que tenham havido avanços, para muitos sistemas o número de átomos envolvidos ainda é muito grande, o que inviabiliza a abordagem puramente quântica, como para o caso da *dinâmica ab initio* que descreve mecânico-quanticamente ambos, solvente e soluto. Isso implica na resolução da equação de Schrödinger para centenas ou milhares de moléculas em distintas configurações, culminando na obrigatoriedade da diminuição do nível de descrição, seja pelo detalhamento das funções de onda, número de moléculas ou número de configurações do solvente.<sup>63</sup>

Para lidar com esse problema, o ASEP/MD faz uso de duas aproximações, a primeira consiste em tratar apenas uma parte do sistema mecânico-quanticamente (no caso deste trabalho, apenas o soluto) enquanto o restante é representado de maneira simplificada. A segunda aproximação é a utilização do MFA, que propõe considerar o potencial do solvente em relação ao soluto de maneira média, ou seja, o potencial que é introduzido na Hamiltoniana do soluto como perturbação é gerado por uma única configuração do solvente que representa a média entre todas as configurações termicamente acessíveis. No caso do ASEP/MD a estrutura do solvente ao redor do soluto é conseguida por meio de simulações de dinâmica molecular e a perturbação é descrita usando-se um potencial de cargas ajustadas.

A utilização do MFA reduz drasticamente o número de cálculos quânticos necessários para o estudo dos efeitos de solvente já que não há a necessidade de uma nova resolução da função de onda para cada nova configuração do solvente. No entanto, como citado anteriormente, cada aproximação feita torna o estudo do sistema mais factível do ponto de vista do custo computacional mas traz consigo um prejuízo na descrição do mesmo. Nesse

caso, o preço a ser pago é a denominada *componente Stark*,<sup>64-65</sup> que representa o negligenciamento do termo de energia da correlação associada à resposta da distribuição de cargas do soluto às mudanças instantâneas do solvente devido a agitação térmica, ou seja, métodos que utilizem o MFA (como é o caso) só terão validade caso a contribuição energética do valor da componente Stark para a energia total de interação entre soluto-solvente seja negligenciável. Estudos feitos anteriormente, tanto teoricamente<sup>51</sup> quanto experimentalmente,<sup>66</sup> indicam ser esse normalmente o caso.

#### 2.4.2 Detalhes do método ASEP/MD

Como de costume em métodos QM/MM o método ASEP/MD divide a Hamiltoniana em 3 diferentes termos:<sup>28,67-72</sup>

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (18)$$

Onde  $\hat{H}_{QM}$  é a Hamiltoniana correspondente ao subsistema quântico,  $\hat{H}_{MM}$  é a energia do subsistema clássico, calculado por meio de campos de forças de mecânica molecular, e,  $\hat{H}_{QM/MM}$  é o termo referente a interação entre os subsistemas quântico e clássico. Por conveniência computacional o termo de interação entre os dois subsistemas é normalmente expresso como a soma de dois componentes:

$$\hat{H}_{QM/MM} = \hat{H}_{QM/MM}^{eletrostático} + \hat{H}_{QM/MM}^{vdw} \quad (19)$$

Onde,  $\hat{H}_{QM/MM}^{eletrostático}$  é o componente eletrostático da interação entre a parte quântica e a clássica (aqui, soluto e solvente respectivamente) enquanto  $\hat{H}_{QM/MM}^{vdw}$  representa a energia de Van der Walls referente a interação entre os dois subsistemas.

Como o termo  $\hat{H}_{MM}$  não diz respeito ao subsistema quântico, ele pode ser retirado da Hamiltoniana e computado separadamente. O restante é incluído na equação de Schrödinger efetiva:

$$(\hat{H}_{QM} + \hat{H}_{QM/MM})|\Psi\rangle = E|\Psi\rangle \quad (20)$$

Onde o termo E, inclui as energias referentes ao subsistema quântico e à interação entre o subsistema quântico e clássico.

Normalmente nos métodos de QM/MM a equação (17) é resolvida para cada diferente configuração do solvente, o que significa centenas ou mesmo milhares de vezes, e o valor final da energia ou de qualquer outra propriedade do sistema é calculada fazendo-se a média sobre todas as configurações acessadas. É nesse ponto que a utilização do MFA se mostra eficiente em diminuir o custo computacional, pois ao invés de realizar o cálculo dessa maneira faz-se uma média de todas as configurações, calcula-se a energia dessa configuração média e essa energia é introduzida como perturbação na equação de Schrodinger efetiva. Outro ponto a se destacar é que se supõe que o termo  $\hat{H}_{QM/MM}^{vdw}$  não vai influenciar na função de onda do soluto (contribuindo apenas para o valor final da energia) e por isso é representado por potenciais clássicos que dependem apenas das coordenadas nucleares (normalmente pelo potencial de Lennard Jones).

Sendo assim, a Hamiltoniana referente à interação entre os dois subsistemas irá representar apenas a parte eletrostática da interação, como uma média sobre todas as configurações, tomando a seguinte forma:

$$\langle \hat{H}_{QM/MM}^{eletrico} \rangle = \int dr. \hat{\rho} \cdot \langle V_s(r, \rho) \rangle \quad (21)$$

Onde  $\hat{\rho}$  é o operador de densidade de carga do soluto e o termo  $\langle V_s(r, \rho) \rangle$  representa o potencial eletrostático médio gerado pelo solvente no ponto r (ASEP), onde as chaves denotam uma média estatística sobre todas as configurações de equilíbrio vindas da dinâmica molecular, em que o soluto tem mantida sua geometria fixa e é representado por um conjunto de cargas  $\rho$ . Assim, temos a equação de Schrödinger efetiva em sua forma final, para o método ASEP:

$$\left( \hat{H}_{QM} + \langle \hat{H}_{QM/MM}^{eletrico} \rangle \right) |\Psi\rangle = \bar{E} |\Psi\rangle \quad (22)$$

Onde o valor de energia não é o mesmo da equação (17), mas um valor obtido na presença da perturbação média causada pelo solvente no soluto, ou seja,  $\langle E \rangle \approx \bar{E}$ .

É importante salientar que apenas configurações separadas por ao menos 0.05 ps são incluídas para a determinação da configuração média do solvente ao redor do soluto. Isso, para que diminua a correlação estatística entre as configurações selecionadas e não haja desperdício no tempo de cálculo. Também, para manter o número de cargas "manuseáveis", apenas as moléculas presentes na primeira camada de solvatação têm o efeito de suas cargas considerado explicitamente, o efeito das cargas das demais moléculas é descrito por um potencial de cargas ajustado.

Por fim, por conveniência computacional, o potencial eletrostático gerado pela média das configurações, ASEP, é tornado discreto e representado por um conjunto de pontos de cargas  $\{q_i\}$  que simulam o potencial gerado pela distribuição contínua do solvente:

$$\langle V_s(r, \rho) \rangle = \sum_i \frac{q_i}{r - r_i} \quad , \quad (23)$$

O conjunto de cargas  $q_i$  é obtido em 3 passos<sup>49</sup>:

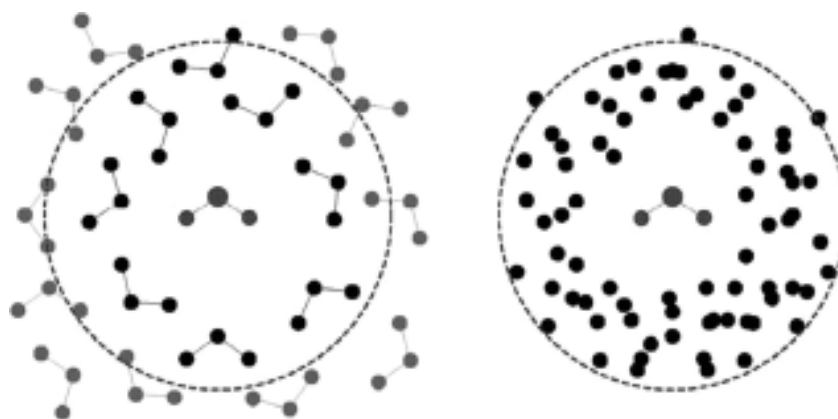
**1.** Cada configuração selecionada é transladada e rotacionada de maneira que todas as coordenadas do solvente possam ser referenciadas a um sistema de referência centrado no centro de massa do soluto, com o eixo das coordenadas paralelo ao principal eixo de inércia do soluto.

**2.** Na sequência, são incluídos explicitamente no ASEP as cargas pertencentes às moléculas de solvente que, em qualquer das configurações da dinâmica molecular selecionadas, estejam dentro de uma esfera de raio  $a$  e que inclua ao menos a primeira camada de solvatação (Figura 10). O valor de cada carga é então dividido pelo número de configurações de solvente incluídas na determinação do ASEP. Depois para diminuir o número de cargas, todas as cargas que estejam suficientemente próximas umas das outras (geralmente até 0,5 angstroms de distância), são somadas.

**3.** Por fim, um segundo conjunto de cargas é adicionado para representar o efeito das moléculas de solvente da segunda camada de solvatação (Figura 11). Essas cargas são obtidas pelo ajuste dos mínimos quadrados dos valores do ASEP originados pelas moléculas do solvente externas, em uma rede de pontos tridimensional definida no volume ocupado pela molécula do soluto. O volume do soluto é definido por meio de um conjunto de esferas

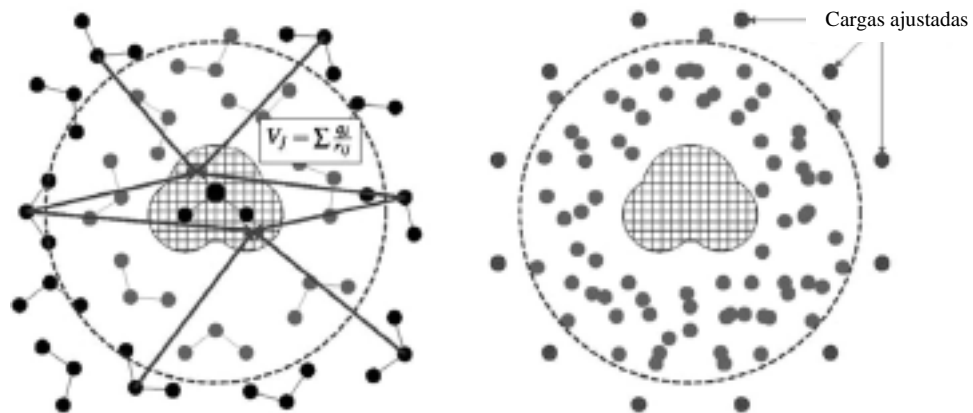
interligadas de raios  $fR_{vdw}$ , onde  $f$  é um fator numérico próximo a unidade, e,  $R_{vdw}$  são os raios das ligações.

**Figura 10** - Cargas representando o potencial eletrostático médio gerado pela primeira camada de solvatação.



Fonte<sup>49</sup>

**Figura 11 - Cargas representando o potencial eletrostático médio gerado pelas cargas além da camada de solvatação.**



Fonte<sup>49</sup>

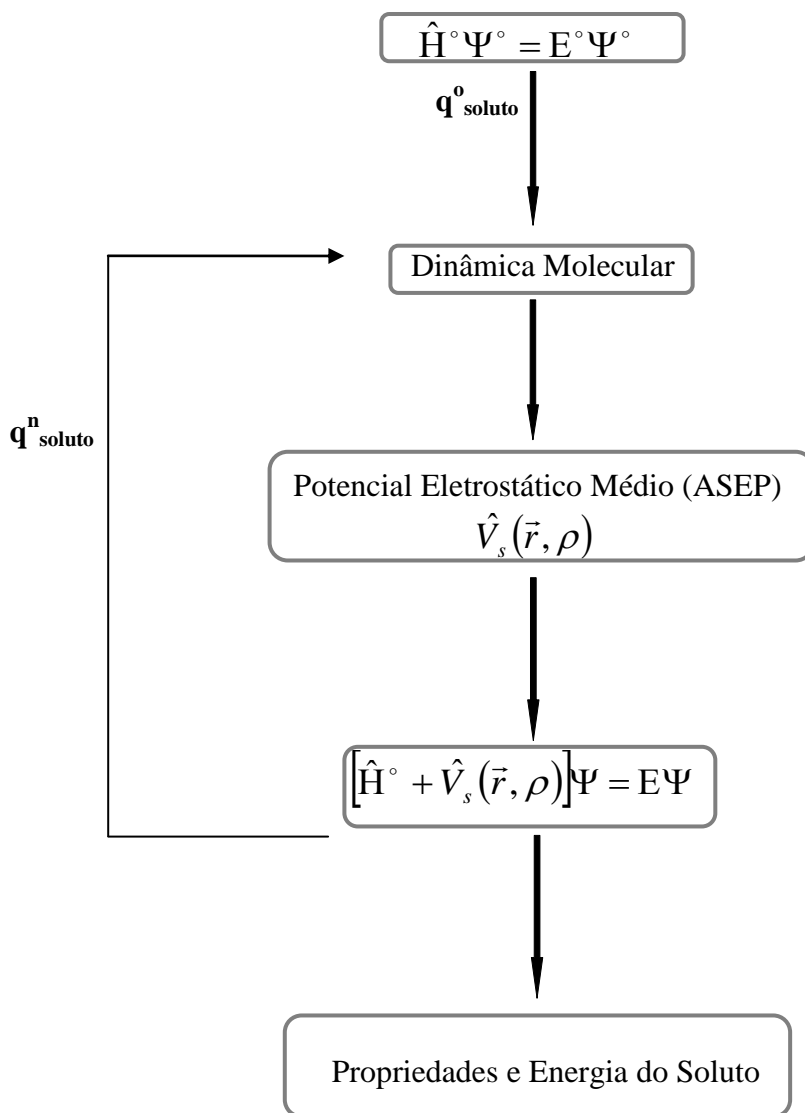
O número total de cargas introduzidas dentro da Hamiltoniana que representa a perturbação que o solvente causa ao soluto em solução está, normalmente, entre 25.000 e 35.000.

O esquema básico do método ASEP/MD é muito simples e está esquematizado na

Figura 12. Resumidamente, a simulação inicia-se com um cálculo quântico da molécula de soluto na fase gasosa. A distribuição de cargas obtidas por meio deste cálculo será inserida como parâmetro inicial da simulação por dinâmica molecular, onde o soluto terá sua estrutura representada por campos de força clássicos (com exceção das cargas obtidas no primeiro passo), e fixada, tanto em relação ao solvente (sem se rotacionar ou translacionar) quanto para distâncias ou ângulos de ligação. A partir desse ponto, sucessivas dinâmicas moleculares são realizadas a fim de se obter diversas configurações possíveis para a organização das moléculas de solvente ao redor da molécula do soluto, para que seja possível por meio da média sobre todas as configurações incluídas (segundo os critérios citados anteriormente neste capítulo) a obtenção de uma configuração que represente a distribuição média das moléculas do solvente, e se possa calcular o potencial eletrostático médio (ASEP) gerado por essa na molécula do soluto. Esse potencial é então adicionado à Hamiltoniana do soluto e um novo cálculo quântico é realizado, onde, a partir da resolução da equação de Schrodinger efetiva (equação 19) é obtida a equação de onda do soluto perturbado pela presença do solvente. A nova distribuição de cargas para o soluto é então introduzida como parâmetro de entrada, novamente, para um novo ciclo de cálculos de dinâmica molecular, ou

seja, os dados de saída proveniente dos cálculos quânticos serão sempre os dados de entrada dos cálculos de dinâmica molecular e vice-versa, sendo que o processo se repetirá de maneira interativa até que haja convergência nas cargas do soluto e a estrutura do solvente ao redor deste. Normalmente a convergência é obtida em 4-5 ciclos, porém, é conveniente que os cálculos continuem por mais 10-15 ciclos para que os resultados finais possam ter o erro estatístico associado da média dos últimos 5 ciclos.

Figura 12 - Esquema básico do método ASEP/MD



Fonte<sup>49</sup>

### 2.4.3 Localização dos pontos críticos em superfícies de energia livre

Como citado anteriormente, é necessário que um modelo que estude soluções permita o cálculo de energia livre e é, portanto, necessário que tenhamos à disposição métodos que permitam o cálculo dessa quantidade de maneira efetiva e computacionalmente factível.

A superfície de energia livre é formada pela média temporal das forças atuando em cada átomo de determinada molécula. Uma maneira de se otimizar a geometria de uma molécula de soluto sendo solvatada (como no presente trabalho) é utilização do método de gradiente de energia livre na busca da estrutura mais estável, de modo a alcançar o equilíbrio mútuo entre as cargas do soluto e o potencial gerado pelo solvente.

O ASEP/MD faz o uso de uma variante deste método,<sup>73-76</sup> onde a força média  $\langle F \rangle$  e a Hessiana  $\langle H \rangle$  sentidas pelos átomos da molécula de soluto (com geometria e carga fixos) são utilizadas para a otimização de geometria juntamente com MFA, ou seja, ao invés de se calcular a derivada primeira, e segunda, da energia média de todas as configurações para determinar a força e a Hessiana, respectivamente, utiliza-se a energia da configuração média. Uma aproximação adicional diz respeito ao negligenciamento dos termos relacionados as flutuações térmicas da força para a Hessiana, já que aqui, ela é utilizada apenas para acelerar o processo de otimização, guiando a direção do mínimo do gradiente. Essas aproximações permitem uma considerável economia em número de cálculos quânticos, introduzindo, em contrapartida, erros menores que 5%<sup>77</sup>. As equações que descrevem  $\langle F \rangle$  e  $\langle H \rangle$  no método ASEP/MD seguem abaixo:

$$F(r) = -\frac{\partial G}{\partial(r)} \approx -\frac{\partial E(r)}{\partial(r)}, \quad (24)$$

$$H \approx \frac{\partial^2 E}{\partial r \partial r}, \quad (25)$$

Onde,  $G$  é a energia livre do sistema, e  $E$  é a energia relativa à configuração média do soluto ao redor do solvente, obtida na equação (19).

A localização dos pontos críticos e mínimos em superfície de energia livre são obtidos por meio do algoritmo RFO<sup>78</sup> uma vez que ambos, força e Hessiana, estejam determinados.

#### 2.4.4 Cálculo de diferença de energia livre

Para diversos interesses práticos a diferença de energia livre é uma informação imprescindível, como na determinação de estados excitados e elementares nos processos foto-físicos, para produtos, reagentes e estados de transição nos processos que envolvem reações químicas, e como, a exemplo deste trabalho, no estudo da ordem de estabilidade dos conformeros definidos no mapa de *Ramachandran* em determinado solvente.

Para a determinação da diferença de energia livre de dois estados distintos é utilizada a equação:

$$\Delta G_s = \Delta E_{soluto} + \Delta G_{interação} + \Delta ZPE_{soluto}, \quad (26)$$

Onde,  $\Delta G_s$  é a diferença de energia livre entre os estados A e B,  $\Delta E_{soluto}$  é a diferença de energia interna mecânico-quantica,  $\Delta G_{interação}$  é a diferença de energia livre para a interação soluto-solvente, enquanto o termo  $\Delta ZPE_{soluto}$  inclui a diferença de energia entre os dois estados no ponto zero de energia, as contribuições térmicas e de entropia para a energia quanto mecânica livre do soluto.

As duas estruturas são otimizadas em solução e tanto a distribuição de cargas quanto a energia interna do soluto são determinadas na presença do solvente. Assim, o termo  $\Delta E_{soluto}$  é definido como :

$$\Delta E_{soluto} = E_B - E_A = \langle \Psi_B | \hat{H}_B^\circ | \Psi_B \rangle - \langle \Psi_A | \hat{H}_A^\circ | \Psi_A \rangle \quad (27)$$

Onde  $\hat{H}_B^\circ$  é a Hamiltoniana para a estrutura B na fase gasosa, e  $\Psi_B$  é a função de onda eletrônica do sistema (mesma correspondência para a estrutura A), ou seja, integra-se a Hamiltoniana do soluto em vazio para todo o intervalo onde há a perturbação deste pelo

solvente. Sendo que  $\Psi_B$  (consequentemente  $\Psi_A$ ) é obtido pela resolução da equação 19, onde  $E_B$  e  $E_A$  são calculados utilizando-se as geometrias otimizadas em solução.

O termo  $\Delta G_{\text{interação}}$  é calculado pelo método de teoria de perturbação,<sup>79</sup> levando em conta o conjunto das configurações termicamente acessíveis do sistema. Nesse processo, a geometria da molécula de soluto é considerada rígida (e função do parâmetro de perturbação  $\lambda$ ), enquanto que as moléculas de solvente em que ela está inserida podem se mover livremente. O parâmetro de perturbação pode assumir valores entre 0 e 1, sendo que quando se assume que este seja igual a 0, os valores de geometria, cargas e dos parâmetros de *Lennard Jones* (soluto-solvente) são os referentes à molécula inicial, enquanto que os valores referentes a molécula final vão estar presentes na molécula quando o parâmetro de perturbação  $\lambda$  for igual a 1. Para valores intermediários faz-se uma interpolação linear.

O soluto, que para a determinação das energias, geometrias e das distribuições de cargas iniciais e finais é representado mecânico-quânticamente, para o cálculo de  $\Delta G_{\text{interação}}$  é representado classicamente, ou seja, uma vez estabelecidos esses valores, o soluto será inserido em uma simulação de dinâmica molecular, representado por um conjunto de pontos de cargas. Essa aproximação permite uma grande economia em tempo de cálculo, e, caso seja usada uma boa distribuição de cargas, não se espera melhoria alguma com a representação quântico-mecânica.<sup>80</sup>

#### 2.4.5 Considerações finais

O método ASEP/MD tem quatro principais características que fazem com que ele seja uma ferramenta muito útil para a química computacional:<sup>49</sup>

1. Número reduzido de cálculos quânticos, o que permite utilizar níveis tão robustos quanto os utilizados em fase gasosa.

2. Facilidade em inserir termos de interações específicas e de longo alcance, uma vez que o solvente é descrito pelos potenciais da mecânica molecular.

3. Ao final do procedimento a função de onda do soluto e a estrutura do solvente estão mutuamente equilibrados, ou seja, o soluto está polarizado pelo solvente e a estrutura do solvente está em equilíbrio com a distribuição de cargas polarizada do soluto.

4. O método permite o cálculo da energia livre de maneira eficiente.

## 2.5 DETALHES COMPUTACIONAIS

No intuito de estudar o equilíbrio conformacional da molécula de dipeptídeo de alanina (AD) em água, etanol e na mistura água etanol, foram realizados cálculos MP2/aug-cc-pVDZ. Esse nível de cálculo quântico proporciona a descrição correta da conformação espacial de AD e já foi utilizado anteriormente para estudar o sistema em fase gasosa,<sup>81</sup> o que permite uma análise direta sobre os efeitos de solvente sobre a geometria e distribuição de cargas de AD. A utilização de um conjunto de bases de funções difusas no estudo de efeitos de solvente pode ser problemático, uma vez que, em alguns casos, elas podem resultar em uma super polarização das cargas do soluto.<sup>82</sup> Para verificar se esse problema poderia afetar o sistema proposto, cálculos com a base cc-pVDZ foram realizados.

Todos os cálculos em solução foram realizados pelo método ASEP/MD (descrito na secção anterior). Como foi apontado, o método permite, por meio da combinação de dinâmica molecular e cálculos *abinito*, uma descrição correta das interações soluto-solvente, com um menor número de cálculos quânticos de alto nível do que os utilizados em metodologias convencionais de QM/MM. Os cálculos *ab initio* foram realizados com o pacote de programas Gaussian09<sup>42</sup> enquanto que, as simulações por dinâmica molecular foram realizadas pelo pacote de programas GROMACS.<sup>44</sup> Cálculos com solvente rígido foram realizados com Moldy.<sup>83</sup> As simulações contêm uma molécula rígida de AD imersa em três diferentes soluções, em 1497 moléculas de água, 657 moléculas de etanol e em uma mistura de 606 moléculas de água e 280 moléculas de etanol. Como está demonstrado na literatura que a combinação de campos de força OPLS-AA/TIP3P reproduz adequadamente a estrutura de AD em água,<sup>32</sup> os parâmetros de Lennard-Jones utilizados na dinâmica molecular foram retirados dos campos de força OPLS-AA,<sup>84</sup> para AD e etanol, e TIP3P<sup>85</sup> para a água. As cargas atômicas foram atualizadas a cada ciclo ASEP/MD pelo método CHELPG.<sup>86,87</sup>

As condições periódicas de contorno foram aplicadas nos três eixos dimensionais. As interações eletrostáticas de curto alcance tiveram um raio de corte de 13 angstroms, enquanto que as de longo alcance foram calculadas com o método de malha de partículas de Ewald (PME).<sup>88</sup> A temperatura foi fixada a 298 K pelo termostato de Nosé-Hoover.<sup>89</sup> Todas as simulações foram realizada no *esemble* NVT por 500 ps, com intervalo (time step) de 1ps entre as configurações, onde os primeiros 200ps eram utilizados para equilibrar o sistema e os últimos 300 ps para produção de dados. Os resultados finais do estudo das soluções foram obtidos pela média dos últimos 5 ciclos ASEP/MD, e, por isso, representam 1.5 ns de média

Por fim, para cada cálculo de energia livre (por teoria de perturbação) que conecta dois conformeros, A e B, 100 dinâmicas moleculares de 500 ps cada uma foram realizadas. Em todos os casos a energia livre de transformação direta A-B, difere em menos de 0,2 kcal/mol da energia livre de transformação inversa, B-A. O resultado final é o valor médio entre a energia direta e reversa.

A numeração atômica e os ângulos diedros representativos da molécula AD estão representados nas Figuras 1 e 2. Como de costume, a disposição da cadeia central é referenciada de acordo com os valores dos ângulos diedros centrais  $\Phi$  e  $\Psi$  ( $C_5N_7C_9C_{11}$  e  $N_7C_9C_{11}N_{17}$ )

### Capítulo 3 - O PASSEIO - RESULTADOS E DISCUSSÃO

Apesar do uso significativo de co-solventes alcoólicos para estabilização protéica, não há um consenso quanto aos fatores que são determinantes nesse processo. Interações de hidrogênio<sup>4</sup>, impedimentos estéricos<sup>36</sup> e solvatação preferencial<sup>90</sup> foram propostos. Estudos utilizando peptídeos pequenos concluíram que mesmo a presença de pequenas quantidades de alcoóis muda a conformação de equilíbrio, favorecendo o conformero PPII, enquanto que outros estudos argumentam sobre se há ou não a exclusão das moléculas de álcool da primeira camada de solvatação. Nesse sentido, o espectro vibracional de tripeptídeo de alanina(AT)<sup>33</sup> revelou haver interações entre suas cadeias laterais alifáticas e moléculas de álcool em misturas binárias. Em busca de esclarecer esses pontos e outros que tangem o problema de estabilidade enzimática, foi utilizado o AD como sistema modelo em três sistemas distintos: água, etanol, e mistura de 60-40% água-etanol em volume, que corresponde a uma fração molar de 0.68.

Começamos a análise esclarecendo alguns detalhes técnicos. Como foi citado anteriormente, o uso de funções difusas para o estudo de moléculas em solução por métodos de QM/MM pode ser problemático devido a possibilidade de super polarização da molécula de soluto.<sup>88</sup> Na maior parte dos métodos QM/MM apenas o componente eletrostático da interação soluto-solvente é incluído na Hamiltoniana do soluto. Nesse caso o efeito da troca de elétrons entre soluto-solvente é completamente negligenciado pela função de onda do soluto. Como consequência a sobreposição da função de onda do soluto com as moléculas do solvente não é penalizada e, a energia de atração entre os elétrons do soluto por cargas positivas do solvente pode ser super estimada. Esse efeito não-físico pode ser especialmente importante quando se utiliza funções difusas, uma vez que elas se estendem por distâncias maiores dentro do solvente. Em função disso, na Tabela 1 é apresentada a comparação dos resultados para energia livre e momento dipolo dos conformeros definidos pelo diagrama de Ramachandram,  $\alpha$ -R e PPII, em água.

**Tabela 1-Comparação dos valores de energia livre (kcal/mol) e momento dipolo (D) para os conforméromos  $\alpha_R$  e PPII utilizando-se os conjuntos de base cc-pVDZ e aug-cc-pVDZ em água.**

	$\alpha_R \rightarrow \text{PPII}$			
	$\Delta\mu$	$\Delta E_{\text{soluta}}$	$\Delta G_{\text{int}}$	$\Delta G_s$
cc-pVDZ	-3,8	-1,5	2,0	0,5
aug-cc-PVDZ	-4,0	-2,4	3,1	0,7

A mudança de energia livre associada a transformação  $\alpha_R \rightarrow \text{PPII}$  em água é semelhante para os dois conjuntos de bases: 0.5 e 0.7 Kcal/mol para cc-pVDZ e aug-cc-PVDZ, respectivamente. Os dipolos induzidos são maiores para o conjunto de bases mais extenso, não por apresentar valores superestimados, mas como consequência da melhor descrição da polarização do AD. Em geral, como se pode verificar na tabela 1, para o sistema em estudo o uso de funções difusas parece não produzir artefatos e a superpolarização do soluto pode ser descartada. Tendo em vista que a base aug-cc-pVDZ foi utilizada para o estudo teórico do AD em água, para facilitar a comparação dos resultados, o restante dos cálculos foram realizados utilizando-se essa base.

A Tabela 2 apresenta a diferença de energia livre e suas componentes para a transformação  $\alpha_R \rightarrow \text{PPII}$  nos três solventes considerados. As energias livre foram calculadas, como citado na secção 2.3.3, com um método dualístico, hora tratando o sistema mecânico-quanticamente (para a estimativa da contribuição do soluto), hora classicamente (para o tratamento das interações soluto-solvente). Um sinal positivo representa maior estabilidade do conforméromo  $\alpha_R$  em relação ao PPII.

**Tabela 2 - Diferença de energia livre e suas componentes para a transformação  $\alpha_R \rightarrow \text{PPII}$  em três diferentes solventes. Todas as energias estão em Kcal/mol. Em parênteses estão os valores excluindo a contribuição  $\Delta A(\text{rot+vibr})$ .**

	Água	Água+Etanol	Etanol
$\Delta G_{\text{int}}$	3,1	3,1	3,4
$\Delta E_{\text{soluta}}$	-2,4	-2,9	-3,3
$\Delta ZPE_{\text{soluta}}(\text{rot+vibr})$	-0,08	+0,09	-0,42
$\Delta G_s$	0,6 (0,7)	0,2 (0,3)	-0.3 (0,1)
$\Delta G_s(\text{solvente rígido})$	(0,2)	-	(-0,8)

Para os três sistemas estudados a conformação de equilíbrio é dominada pelos conforméros  $\alpha_R$  e PPII, portanto, a população de qualquer outro conforméroro pode ser considerada negligenciável e não foi incluída na tabela. Esse comportamento é o inverso que se constata em polipeptídeos mais longos de alanina, onde o conforméroro  $C_5$ (forma  $\beta$ ) pode contribuir de forma substancial para a população de equilíbrio. Aqui, na busca de otimizar a estrutura  $\beta$  o que se verificou foi uma evolução para a estrutura  $\alpha_R$ , em todas as tentativas. Em um trabalho anterior, em que foi utilizada a aproximação de corpo rígido para o solvente e uma otimização de geometria baseada no gradiente de energia potencial, foi possível obter a geometria  $C_5$  otimizada. Essa mesma geometria foi utilizada para se medir a diferença de energia livre em relação ao conforméroro PPII, assumindo-se o solvente flexível. Sob estas condições e, em água, o conforméroro  $C_5$  apresentou uma energia de 3,3 Kcal/mol maior que PPII, isso justifica a baixa estabilidade deste conforméroro à temperatura ambiente.

Utilizando-se nível de cálculo quântico MP2-aug-cc-pVDZ, o conforméroro  $\alpha_R$  se mostra ligeiramente mais estável que o PPII em água e na mistura, porém o comportamento inverso é apresentado em etanol, onde o conforméroro PPII é ligeiramente mais estável que o  $\alpha_R$ . Os resultados para a água divergem de um trabalho anterior que identificou o conforméroro PPII o mais estável.<sup>90</sup> É provável que a diferença entre os trabalhos tenha origem na mudança da flexibilidade do solvente (anteriormente rígido, aqui flexível) e do método de otimização de geometria utilizado (anteriormente baseado no gradiente, agora na Hessiana). A utilização de moléculas flexíveis de água intensifica a energia de interação solvente-soluto e estabiliza o conforméroro  $\alpha_R$ , enquanto que moléculas rígidas de solvente favorecem o conforméroro PPII (ver tabela5). De qualquer maneira, a diferença de energia livre entre eles é muito pequena e o resultado final pode favorecer uma ou outra espécie dependendo do tipo de solvente utilizado, ou as cargas utilizadas durante a simulação. Como conclusão geral a energia interna,  $\Delta E_{QM}$ , favorece o conforméroro PPII, enquanto que a energia de interação,  $\Delta A_{int}$ , favorece o  $\alpha_R$ . As contribuições vibracionais e rotacionais são completamente negligenciáveis para água e para a mistura, mas de alguma maneira, não para o etanol. Qual o conforméroro mais estável é consequência da resultante entre energia interna e energia de interação soluto-solvente, ou seja, pequenas mudanças na geometria, cargas, parâmetros, solvente, podem favorecer o equilíbrio hora de um conforméroro, hora de outro. Em geral, qualquer fator que diminua a energia de interação soluto-solvente vai aumentar a estabilidade relativa do PPII, isso faz com que a população de PPII aumente com a adição de etanol, ainda que a diferença de

estabilidade seja pequena ( em torno de 1 Kcal/mol) passando de 0,6 Kcal/mol em água para -0,3 Kcal/mol em etanol, caso seja usado solvente flexível, e de 0,2 a -0,8 Kcal/mol caso o solvente seja considerado rígido. De acordo com os dados coletados, utilizando solvente flexível, a população de PPII passa de 27% à 62% quando o AD passa da água para o etanol. No intuito de entender mais a fundo esse comportamento, fez-se outras análises relativas à interação AD-solvente.

Primeiro, apresentados nas Tabelas 3 e 4, estão os parâmetros geométricos para os comprimentos de ligação e ângulos diedros dos confórmeros  $\alpha_R$  e PPII otimizados por MP2/aug-ccpVDZ em água, etanol e em solução de água-etanol.

A comparação direta com a fase gasosa não é possível, levando em consideração que os confórmeros PPII e  $\alpha_R$  não são estáveis em fase gasosa e  $\alpha_R$  é uma modificação da estrutura de equilíbrio  $\delta_R$  em solução. De qualquer maneira, as mudanças estruturais mais significativas introduzidas pelo efeito do solvente dizem respeito ao comprimento das ligações peptídicas, e não pela diferença nos valores em obtidos em um caso ou outro, mas pela tendência geral apresentada, em que a ligação C-O aumenta com o aumento da polaridade, enquanto que a distância da ligação C-N diminui (esse comportamento é condizente com a natureza de ligação dupla conferida à ligação peptídica devido a sua estabilização por ressonância), além de aumentar também a distância na ligação N-H. O aumento nas distâncias de ligação N-H e C-O apontam para a formação de interações de hidrogênio intermoleculares com as moléculas do solvente. Nota-se uma mudança maior dos comprimentos de ligação do AD em água do que a apresentada em etanol. Os valores apresentados para a mistura e para água são os mesmos, o que pode indicar solvatação preferencial. É dito que existe solvatação preferencial quando a composição local de mistura é diferente de sua composição global, esse comportamento é originado normalmente por interações específicas, porém, mesmo na ausência destas, um solvente polar pode induzir uma mudança na composição do seu ambiente por meio do chamado "enriquecimento dielétrico".

O efeito dos solvente sobre os ângulos diedros  $\Phi$  e  $\Psi$  são pequenos.

**Tabela 3 - Parâmetros geométricos selecionados do conformero PPII em três diferentes solventes. Distâncias estão em angstroms e os ângulos em graus.**

	Água	Água+Etanol	Etanol
C <sub>5</sub> O <sub>6</sub>	1,26	1,26	1,25
C <sub>5</sub> N <sub>7</sub>	1,34	1,34	1,35
N <sub>7</sub> H <sub>8</sub>	1,04	1,04	1,03
C <sub>11</sub> O <sub>12</sub>	1,26	1,26	1,25
C <sub>11</sub> N <sub>17</sub>	1,33	1,33	1,34
N <sub>17</sub> H <sub>18</sub>	1,04	1,04	1,03
C <sub>5</sub> N <sub>7</sub> C <sub>10</sub> C <sub>11</sub> (Φ)	-55,9	-54,2	-54,5
N <sub>7</sub> C <sub>10</sub> C <sub>11</sub> N <sub>17</sub> (Ψ)	152,5	150,9	150,5

**Tabela 4 - Parâmetros geométricos selecionados do conformero αR em três diferentes solventes. Distâncias estão em angstroms e os ângulos em graus.**

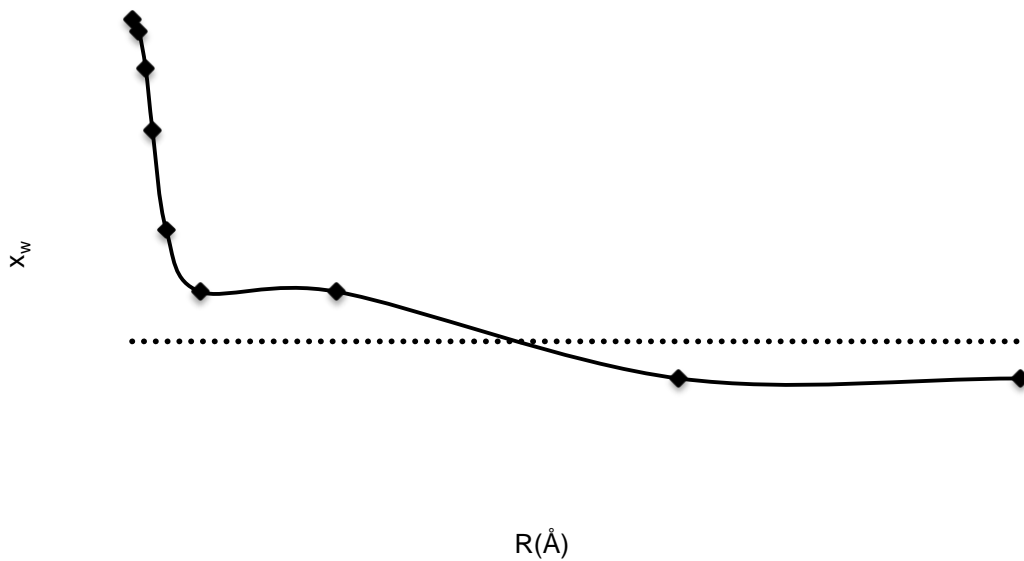
	Água	Água+Etanol	Etanol
C <sub>5</sub> O <sub>6</sub>	1,26	1,26	1,25
C <sub>5</sub> N <sub>7</sub>	1,34	1,35	1,35
N <sub>7</sub> H <sub>8</sub>	1,04	1,04	1,03
C <sub>11</sub> O <sub>12</sub>	1,26	1,26	1,25
C <sub>11</sub> N <sub>17</sub>	1,33	1,33	1,34
N <sub>17</sub> H <sub>18</sub>	1,04	1,04	1,03
C <sub>5</sub> N <sub>7</sub> C <sub>10</sub> C <sub>11</sub> (Φ)	-62,21	-61,29	-58,95
N <sub>7</sub> C <sub>10</sub> C <sub>11</sub> N <sub>17</sub> (Ψ)	-32,34	-32,08	-34,82

Na Tabela 5, pode-se notar um aumento de cerca de 60% da energia de interação,  $E_{int}$ , quando o dipeptídeo passa de etanol para mistura, e de apenas cerca de 7% quando passa de mistura à água. Para os três solventes a  $E_{int}$  é maior em αR do que em PPII, sendo que essa diferença diminui com a diminuição da polaridade do solvente. A interação soluto-solvente desvia, claramente, do comportamento ideal (linear), na realidade, pode-se inclusive, definir uma "fração molar efetiva", segundo a equação 25:

$$\chi_w = \frac{E_{\text{int}}^{W/E} - E_{\text{int}}^E}{E_{\text{int}}^W - E_{\text{int}}^E} \quad (25)$$

Onde, W, E e W/E se referem à água, etanol e água/etanol, respectivamente. Utilizando os valores de  $E_{\text{int}}$  apresentados na Tabela 5 para os conformémeros  $\alpha_R$  e PPII chega-se aos valores 0,85 e 0,83 (respectivamente), valores claramente maiores do que os 0,68, confirmando a presença de solvatação preferencial, conforme se pode notar na figura 17.

**Figura 13 - Fração molar local da água em função da distância (em Å) para o sistema AD-água (linha sólida). A linha pontilhada representa a fração molar de  $x_w=0,68$ .**



A Tabela 5 apresenta também, as energias de distorção,  $E_{dis}$ , que de acordo com a equação 26, é calculada como a diferença entre a energia interna em solução e a energia mínima correspondente em fase gasosa:

$$\Delta E = \langle \Psi | \hat{H}_{QM} | \Psi \rangle - \langle \Psi^\circ | \hat{H}_{QM} | \Psi^\circ \rangle \quad (26)$$

Todavia, como indicado acima, nenhum dos dois conformémeros têm geometrias definidas para fase gasosa, por isso, essas energias foram calculadas utilizando-se a geometria

otimizada em solução, o que implica no fato destas representarem apenas a energia despreendida para a polarização das cargas do AD, não incluindo o efeito de distorção de geometria propriamente dito. O alto valor deste componente evidencia a grande polarização do sistema durante o processo de solvatação. Esse fato é confirmado pelo aumento do valor de momento dipolo da molécula de AD. As diferenças entre  $\alpha_R$  e PPII variam de -2,7 e -3,2 Kcal/mol dependendo do solvente considerado. Como esperado, quanto maior a energia de interação, maior será a energia de distorção, porém, com sinais opostos.

**Tabela 5 - Energia de interação e suas componentes (kcal/mol)**

	Água		Água+Etanol		Etanol	
	$\alpha_R$	PPII	$\alpha_R$	PPII	$\alpha_R$	PPII
$E_{int}$	-138,8	-132,8	-130,2	-123,2	-83,9	-76,2
$E_{QM}^{\circ}$ (a.u)	-494,52099	-494,52037	-494,521217	-494,52082	-494,52325	-494,52373
$E_{QM}$ (a.u)	-494,46870	-494,47246	-494,4726	-494,47727	-494,49361	-494,49889
$E_{dist}$	32,8	30,1	30,5	27,3	18,6	15,6
$\mu$	12,8	8,8	12,4	8,4	10,9	7,1
$\mu_0^*$	7,8	5,2	7,8	5,1	7,5	4,9

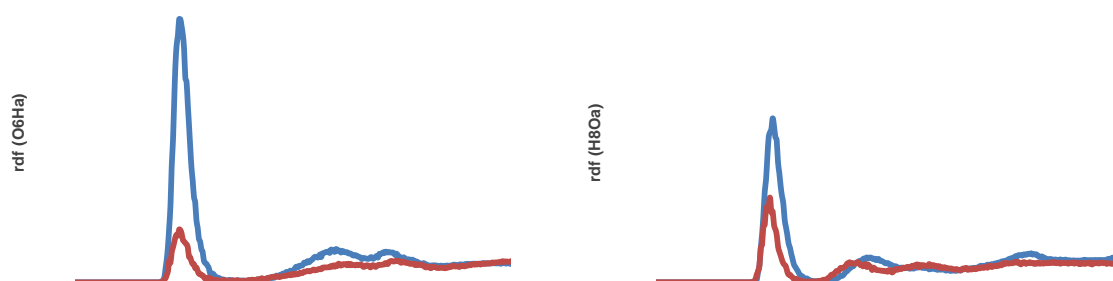
Na sequência é apresentado, a partir de funções de distribuição radial de pares (rdf), números de coordenação e mapas de densidade tridimensional do solvente, o estudo das diferenças estruturais do solvente nos dois conformeros.

A Figura 18 apresenta o diferentes rdf ao redor dos átomos de oxigênio da carbonila e de hidrogênio da amina. Aqui, nota-se que a altura dos primeiros picos para as correlações oxigênio(AD)-hidrogênio(etanol) e hidrogênio(AD)-oxigênio(etanol) são menores do que as encontradas para o etanol puro (Figura 18a), enquanto que o oposto é observado nos rdfs envolvendo moléculas de água (Figura 18b), fato que corrobora a hipótese de solvatação

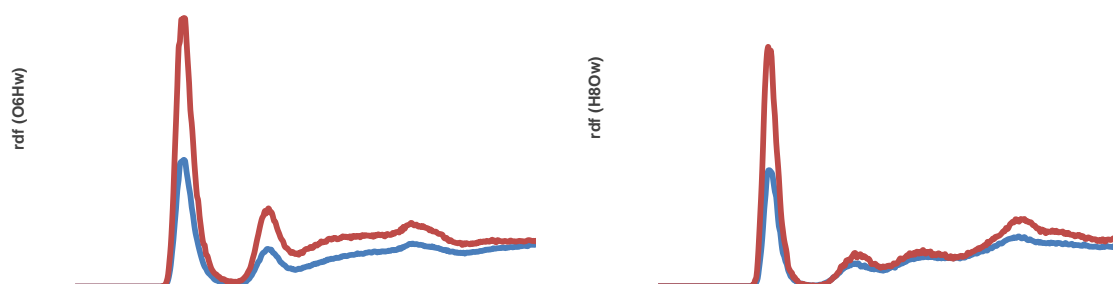
preferencial. Por fim, a Figura 18c, destaca a semelhança entre a primeira camada de solvatação dos conformeros, com uma alta densidade de moléculas de água circundando os grupos carboxílico e amina. Além disso, nota-se que há mais diferenças na distribuição de moléculas de etanol do que de água.

**Figura 14 - Funções de distribuição radial de pares para o sistema AD-solventes diversos. a) Conformero  $\alpha_R$  em etanol puro (azul) e em mistura (vermelho). b) Conformero  $\alpha_R$  em água pura (azul) e em mistura (vermelho). c) Conformeros  $\alpha_R$  (azul) e PPII (vermelho) em mistura.**

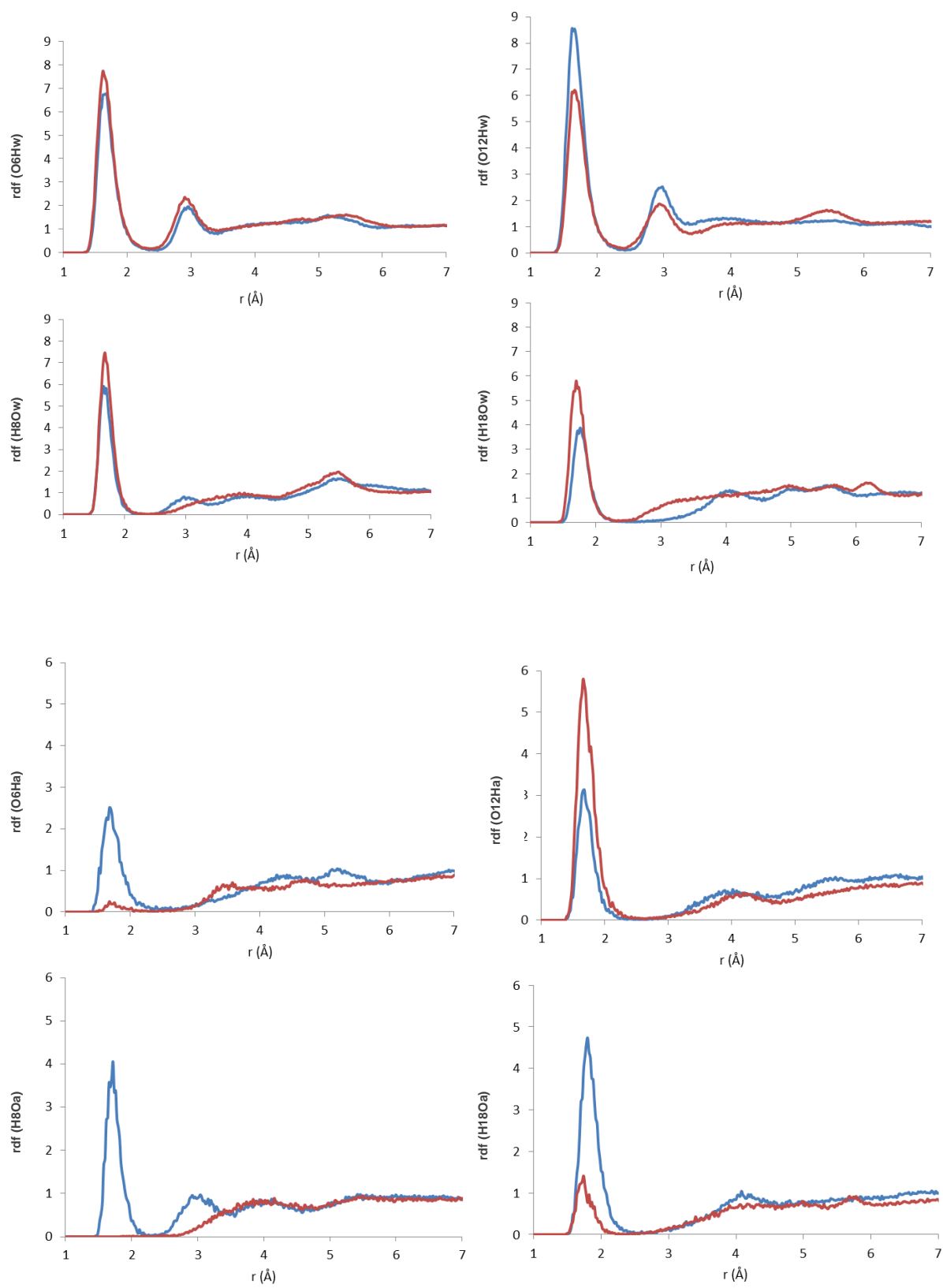
a)



b)



c)



Na tabela 6 estão dipostos os números de coordenação ao redor de alguns átomos da molécula de AD:

**Tabela 5 - Número de coordenação ao redor de grupos funcionais selecionados.**

	Água		Água+Etanol		Etanol	
	$\alpha_R$	PPII	$\alpha_R$	PPII	$\alpha_R$	PPII
O <sub>6</sub> H <sub>w</sub>	2,70	2,75	2,20	2,50		
O <sub>12</sub> H <sub>w</sub>	3,00	2,80	2,75	2,25		
H <sub>8</sub> O <sub>w</sub>	1,00	1,00	0,75	1,00		
H <sub>18</sub> O <sub>w</sub>	1,00	1,00	0,60	0,90		
O <sub>6</sub> H <sub>a</sub>	-	-	0,25	0,00	1,75	1,80
O <sub>12</sub> H <sub>a</sub>	-	-	0,25	0,50	2,00	1,80
H <sub>8</sub> O <sub>a</sub>	-	-	0,25	0,00	1,00	1,00
H <sub>18</sub> O <sub>a</sub>	-	-	0,40	0,10	1,00	1,00

Diversas conclusões podem ser extraídas desta tabela: 1) O número de coordenação total diminui com o aumento da concentração de etanol, 2) Na mistura água-etanol aproximadamente 85% das moléculas na primeira camada de solvatação são de água, sendo que, em média apenas uma molécula de etanol está próxima à molécula de AD (em contraste com as 6 moléculas de etanol presentes na primeira camada de solvatação para o líquido puro). 3) O número de moléculas de solvente na primeira camada de solvatação do conformero  $\alpha_R$  é ligeiramente maior do que o encontrado para o PPII, apesar das diferenças serem pequenas (em torno de 0,3 - 0,4 moléculas). A fração molar local da água, na primeira camada de solvatação, para  $\alpha_R$  varia entre 0,93 e 0,85 dependendo da distância considerada (entre 2,1 e 2,3 Angstroms respectivamente), com valores menores para PPII. Essas diferenças vêm das moléculas próximas ao grupo carboxílico, onde, o grupo carboxílico da extremidade C-terminal favorece PPII, enquanto que a carbonila do N-terminal favorece  $\alpha_R$ , como poderá se observar adiante.

As Figuras 19 e 20 apresentam mapas de densidade das moléculas de solvente ao redor dos conformeros  $\alpha_R$  e PPII para a mistura água-etanol, elas foram geradas propositalmente com densidades atômicas distintas para ressaltar diferentes aspectos estruturais do solvente.

Figura 15 - Distribuição de probabilidade tridimensional ao redor de  $\alpha_R$  solvatado na mistura 60-40(em volume) de água-etanol. Código de cores: Magenta: átomos de oxigênio da água. Laranja: átomos de hidrogênio da água. Vermelho: átomos de oxigênio de etanol. Cinza: átomos de hidrogênio de etanol.

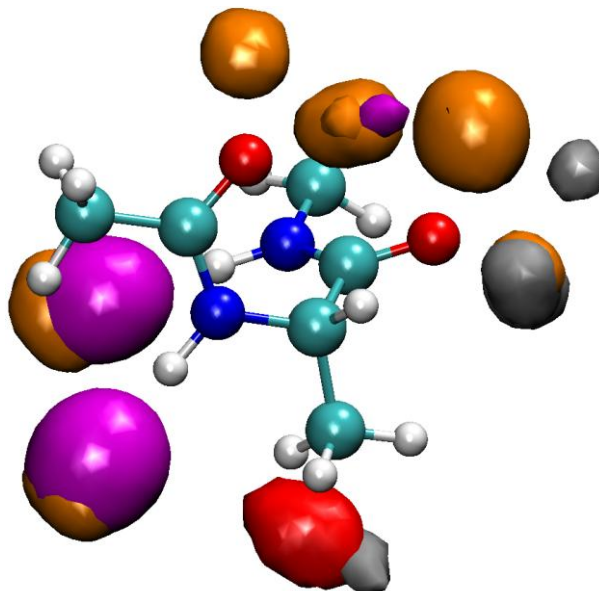
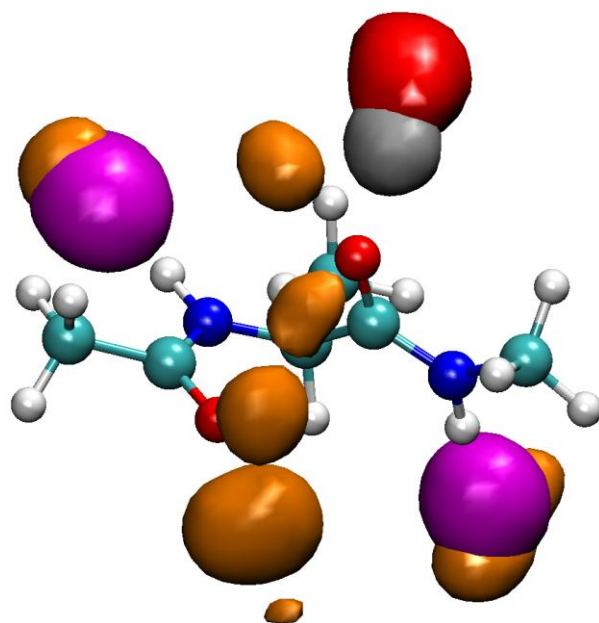
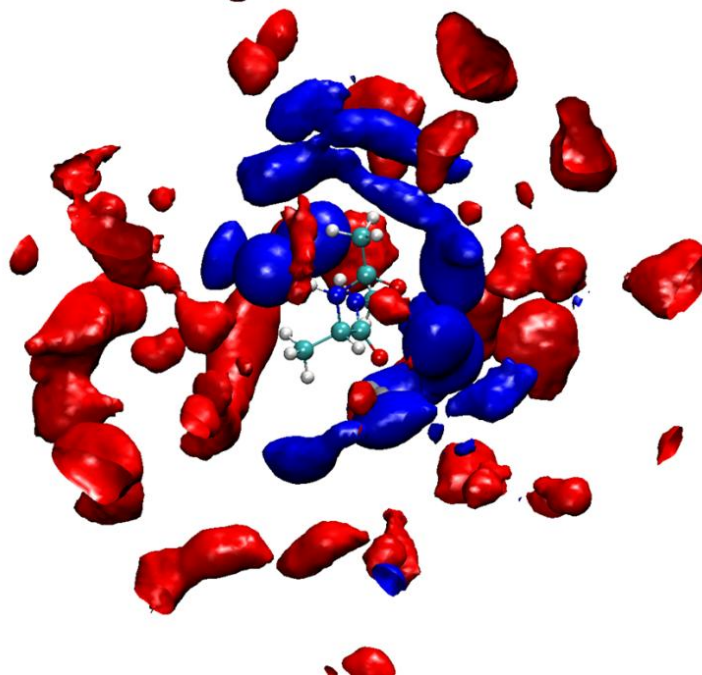


Figura 16 - Distribuição de probabilidade tridimensional ao redor de PPII solvatado na mistura 60-40(em volume) de água-etanol. Código de cores: Magenta: átomos de oxigênio da água. Laranja: átomos de hidrogênio da água. Vermelho: átomos de oxigênio de etanol. Cinza: átomos de hidrogênio de etanol.



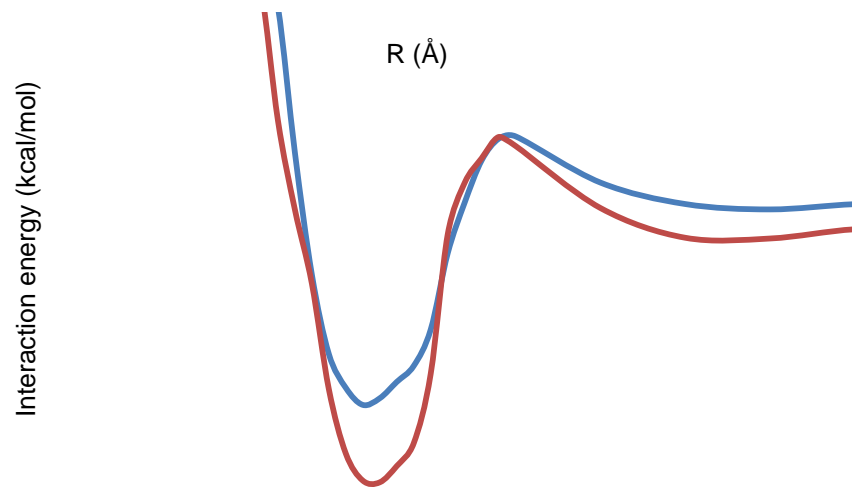
A primeira camada de solvatação é muito semelhante entre  $\alpha_R$  e PPII, formada majoritariamente por moléculas de água, indo de encontro com o ilustrado nas Figuras 18 e 19. Apenas ao redor do grupo carboxílico na extremidade N-terminal são encontradas moléculas de etanol contribuindo de maneira efetiva para a primeira camada de solvatação, sendo que, há maior densidade de moléculas de etanol ao redor do conformero PPII do que em  $\alpha_R$ . Nota-se que no caso do PPII há moléculas de água entre os grupos carboxílico e amínico. Um ponto importante a se esclarecer aqui, é se as moléculas de água formam ligações de hidrogênio entre os grupos (um dos motivos propostos para a estabilidade do conformero PPII em solução aquosa) ou formam ligações de hidrogênio com outras moléculas de água. Neste caso, conforme pode-se observar na Figura 16, não foi encontrada evidência de interação de hidrogênio entre os grupos funcionais, ao contrário, observa-se que os dois grupos têm camadas de solvatação completamente separadas. Por fim, a Figura 17 ilustra a fração molar efetiva obtida a partir dos dados de  $E_{int}$  da tabela 5, conferindo uma visão geral da estrutura do solvente ao redor do conformero  $\alpha_R$ , onde se nota a presença de moléculas de etanol próximas aos grupos metis das cadeias laterais (ver também Figura 15), fato que foi destacado experimentalmente.<sup>28</sup>

**Figura 17 - Estrutura do Solute Solvatado na mistura 60-40% (em volume) de Água +Etanol. Código de cores: Azul: átomos de oxigênio da água. Vermelho: átomos de oxigênio de etanol.**



Uma vez que foi confirmada a presença de solvatação preferencial na molécula de AD, é importante determinar a influência disto na diferença de estabilidade entre os conforméros  $\alpha_R$  e PPII. Com essa finalidade a Figura 18 apresenta a energia de interação em função da distância entre os átomos do solvente e qualquer átomo do AD (apenas o sistema água-etanol é representado, uma vez que os outros solventes apresentam características similares a estas).

**Figura 18 - Energia de interação soluto-solvente (Kcal/mol) para a mistura água-etanol em função da distância entre os átomos do soluto e solvente. Código de cores: Azul:PPII. Vermelho: $\alpha_R$ .**



A diferença de  $E_{\text{int}}$   $\alpha_R$  e PPII depende fortemente da distância. O valor máximo para a diferença é atingido entre 2,1 e 2,2 Å, coincidindo com a primeira camada de solvatação, onde há uma interação muito maior com o solvente para o conforméro  $\alpha_R$  do que para o PPII. Então, entre 2,9 e 3,1 Å, próximo à segunda camada de solvatação, a diferença de  $E_{\text{int}}$  entre eles chega a 0. Por fim, a diferença vai aumentando gradativamente até atingir o seu valor final, que favorece o conforméro  $\alpha_R$ . Em resumo, a primeira camada de solvatação favorece claramente  $\alpha_R$ , efeito que é cancelado pela segunda camada de solvatação. Assim, conclui-se que o valor maior de  $\alpha_R$  é um efeito de longo alcance associado a uma característica mais polar (maior momento dipolar) deste conforméro.

Complementarmente e correlacionado com a disposição do solvente ao redor dos diferentes grupos funcionais é interessante analisar a contribuição desses grupos para a energia de interação clássica soluto-solvente, dispostos abaixo.

**Tabela 6 - Contribuição dos grupos funcionais para a energia de interação clássica para os conforméros  $\alpha_R$  e PPII.**

	Água		Água+Etanol		Etanol	
	$\alpha_R$	PPII	$\alpha_R$	PPII	$\alpha_R$	PPII
CH <sub>3</sub>	-0,5	-2,3	-0,2	-2,0	0,4	-0,8
C <sub>5</sub> O <sub>6</sub>	-54,23	-47,23	-50,12	-43,4	-34,0	-26,6
N <sub>7</sub> H <sub>8</sub>	-12,1	-15,2	-12,7	-17,2	-8,2	-11,1
CHCH <sub>3</sub>	-0,1	-2,8	0,4	-1,9	0,1	-2,4
C <sub>11</sub> O <sub>12</sub>	-53,6	-47,8	-50,5	-42,1	-30,2	-24,3
N <sub>17</sub> H <sub>18</sub>	-14,9	-15,4	-15,5	-15,2	-11,7	-11,1
CH <sub>3</sub>	-1,1	-0,5	-0,04	-0,2	0,1	0,3
Total	-136,5	-131,2	-128,7	-122,0	-83,4	-76,0

A molécula de AD foi dividida de acordo com os grupos funcionais: CH<sub>3</sub>, CO, NH, CHCH<sub>3</sub>, e se calculou a  $E_{\text{int}}$  para cada um dos grupos. As carbonilas contribuem com cerca de 78% da energia total de interação para o conformémero  $\alpha_R$  e 70% para PPII. O grupo NH contribui com cerca de 20% em ambos conformémeros, sendo praticamente negligenciável para a contribuição final na diferença de estabilidade entre  $\alpha_R$  e PPII, o que nos leva a conclusão que o maior valor na energia de interação soluto-solvente em  $\alpha_R$  quando comparado ao PPII é atribuída principalmente a solvatação dos grupos carboxílicos. No intuito de se entender melhor a origem deste comportamento a  $E_{\text{int}}$  foi recalculada utilizando-se o mesmo conjunto de cargas para os dois conformémeros (note que neste modelo as cargas são calculadas mecânico-quânticamente e, conseqüentemente, elas são dependentes da configuração). Ao se calcular (recalcular) a  $E_{\text{int}}$  do sistema água-  $\alpha_R$  utilizando o conjunto de cargas do PPII, seu valor decresce de -136,5 para -133,7 Kcal/mol. Dado que  $\Delta E_{\text{int}}$  é igual a 5,3 Kcal/mol em água, isso significa que 2,8 Kcal/mol corresponde ao uso de conjuntos de cargas distintos para os dois conformémeros e 2,5 Kcal/mol está relacionado com a diferença da estrutura ao redor deles.

Por fim, a partir da comparação entre os dados da Tabela 7 e da Tabela 8 é possível analisar a influência que a polarização do soluto tem na distribuição de cargas e na conformação de equilíbrio em solução.

**Tabela 7 - Contribuição dos grupos funcionais para a energia de interação clássica para os conforméros  $\alpha_R$  e PPII utilizando a distribuição de cargas da fase gasosa.**

	Água		Água+Etanol		Etanol	
	$\alpha_R$	PPII	$\alpha_R$	PPII	$\alpha_R$	PPII
CH <sub>3</sub>	0,5	-0,9	0,6	-0,9	0,8	-0,3
CO	-37,9	-31,1	-35,4	-28,9	-26,5	-20,3
NH	-6,0	-11,7	-6,3	-12,8	-3,3	-8,0
CHCH <sub>3</sub>	1,3	-1,6	1,5	-1,3	0,6	-1,7
CO	-34,6	-29,5	-33,1	-26,7	-22,4	-18,2
CN	-9,6	-11,1	-10,2	-11,2	-7,6	-7,6
CH <sub>3</sub>	-0,3	-0,1	0,4	0,2	0,4	0,4
Total	-86,6	-85,9	-82,6	-81,6	-57,7	-55,7

O aumento na energia de interação que se verifica para o soluto solvatado (valores da Tabela 7 em relação aos da Tabela 8) é guiado majoritariamente pela polarização dos grupos carboxílicos. A carga do oxigênio aumenta em cerca de 40% e a do carbono cerca de 25% em solução.

Em água a energia dos conforméros  $\alpha_R$  e PPII têm os valores, respectivamente, de -86,5 e -85,9, apresentando uma diferença de apenas 0,6 Kcal/mol. Nessas condições o conforméromo PPII se torna o conforméromo mais estável. Analisando-se a contribuição de cada grupo e a relação com os conformémeros é possível notar que os grupos carboxílicos favorecem  $\alpha_R$  enquanto que os grupos aminas favorecem ao PPII. Quando a distribuição de cargas do soluto se torna polarizada pelo solvente (Tabela 7) a contribuição dos grupos carboxílicos permanece praticamente constante, enquanto que a contribuição dos grupos amina tende a diminuir, ou seja, a polarização pelo solvente favorece as interações que conferem maior estabilidade ao conforméromo  $\alpha_R$  em detrimento do PPII.

Independentemente da polarização das cargas do soluto pelo solvente, a adição de etanol sempre favorece a estabilidade do conforméromo PPII.

## Capítulo 4- O APRENDIZADO - CONCLUSÕES

A comparação das estruturas conformacionais de equilíbrio do dipeptídeo de alanina em água, etanol e em mistura água-etanol, permite entender melhor o papel desempenhado pelos alcoóis na estabilização de peptídeos e proteínas. Os resultados teóricos deste trabalho confirmaram dados experimentais, indicando que solventes orgânicos como o etanol mudam o panorama energético do dipeptídeo, favorecendo o conformero PPII em relação ao conformero  $\alpha_R$ . Na tentativa de esclarecer a origem deste comportamento, foram analisados os valores dos principais componentes da energia livre, tendo-se observado que as energias de interação soluto solvente e energia interna trabalham de maneira antagônica, enquanto a energia interna é mais baixa, favorecendo o conformero PPII, a energia de interação AD-solvente favorece o conformero  $\alpha_R$ . Os efeitos do conjunto de bases (aug-ccpVDZ x ccpVDZ), flexibilidade do solvente (flexível x rígido) e distribuição de cargas do dipeptídeo (polarizado x não-polarizado) foram analisados, e, apesar desses fatores afetarem o valor absoluto de energia livre, eles não afetam as conclusões finais: 1) A presença de etanol favorece a população de PPII. 2) A diminuição da energia de interação entre soluto e solvente está correlacionada com a diminuição na população  $\alpha_R$ . 3) Existe solvatação preferencial para o sistema, no entanto, esse fato parece não contribuir de maneira decisiva para a definição da estabilidade das populações relativas de equilíbrio, uma vez que o valor de energia de interação ao redor da segunda camada de solvatação é igual para ambos conformeros.

Foi observado que o sistema não apresenta comportamento linear em dois diferentes aspectos. Primeiro, a mistura água-etanol é um exemplo claro de solvatação preferencial (como citado acima), uma vez que a primeira camada de solvatação é constituída majoritariamente por moléculas de água, o que resulta em energias de interação soluto-solvente muito similares em água e em mistura. Apesar deste comportamento estar de acordo com resultados experimentais de IR, foi encontrada interação direta entre moléculas de etanol e o dipeptídeo, sendo o grupo carboxílico do N-terminal e, as cadeias metílicas laterais, as áreas mais prováveis dessa interação ocorrer. O segundo ponto em que o sistema não se comporta de maneira ideal é quanto ao valor de energia livre calculado em etanol, que distoa do esperado ( $\Delta A_{int} = \frac{1}{2} \Delta E_{int}$ ), na realidade, o solvente etanol é aquele que apresenta o valor

mais baixo de  $\Delta A_{\text{int}}$  apesar de exibir a maior diferença de  $\Delta E_{\text{int}}$  PPII-  $\alpha_R$ . Para água e a mistura, o comportamento linear esperado é observado.

## Capítulo 5 - A JORNADA - CONSIDERAÇÕES FINAIS

O trabalho iniciou com o desenvolvimento das rotinas de dinâmica molecular, implementadas no programa GROMACS, o que se realizou por um período de aproximadamente 2 anos, no Laboratório de Química Computacional do DFQ. Durante esse tempo se adquiriu a expertise para realizar os cálculos relacionados à dinâmica de solvatação do peptídeo. Posteriormente, em função de concessão de bolsa sanduiche da CAPES, o trabalho teve continuidade ao longo de um ano junto ao grupo de pesquisa da Universidade de Extremadura, na cidade de Badajoz. Além do artigo sobre o dipeptídeo de alanina (anexo), espera-se ter iniciado uma colaboração que se estenda futuramente. Neste sentido, cálculos sobre o tripeptídeo (forma zwitteriônica) de alanina, utilizando os mesmos parâmetros que foram utilizados para o dipeptídeo, estão sendo realizados e os resultados serão comparados além do trabalho com o dipeptídeo, a dados experimentais. Com isso, pretende-se investigar se o comportamento observado para o dipeptídeo se mantém no caso do tripeptídeo e analisar a qualidade do método em reproduzir dados experimentais, afim de melhorar em nível de detalhes a compreensão do observado macroscopicamente por métodos experimentais.

## REFERÊNCIAS

- 1 BERG, J. M.; TYMOCZKO, J. L.; STRYER, L. **Bioquímica**. 6. ed. Rio de Janeiro: Guanabara Koogan, 2010. p. 25-61.
- 2 NELSON, D. L.; COX, M. M. **Princípios de bioquímica de Lehninger**. 5. ed. Porto Alegre: Artmed, 2011. p.72-103.
- 3 MORAN, L. et al. **Bioquímica**. 5. ed. São Paulo: Pearson Education do Brasil, 2013. p. 56-135.
- 4 QUINTAS, A.; ROSA, A. P.; HALPERN, M. J. **Bioquímica: organização molecular da vida**. Lisboa: Lidel-Edições Técnicas, 2008. p.139-293.
- 5 KANAMORI, E. et al. **Crystal structure of citrate synthase from thermus thermophilus HB8**. [S.l.: s.n., 200] Disponível em: <<http://www.rcsb.org/pdb/explore/explore.do?structureId=1IXE>>. Acesso em: 14 fev. 2016.
- 6 YURYEV, R.; LIESE, A. Biocatalysis: the Outcast. **Chem.Cat.Chem.**, Berlin, v. 2, n. 1, p. 103-107, 2010.
- 7 REETZ, M. T. Biocatalysis in organic chemistry and biotechnology: past, present, and future. **J. Am. Chem. Soc.**, Marburg, v. 34, n. 135, p. 12480-12496, 2013.
- 8 MACHADO, A. A. S. C. From the first to the second twelve principles of green chemistry. **Quím. Nova**, São Paulo, v. 35, n. 6, p. 1250-1259, 2012.
- 9 Braga, A. L. et al. Catalise assimétrica no Brasil: Desenvolvimento e potencialidades para o avanço da indústria química brasileira. **Quím. Nova**, São Paulo, v. 36, n. 10, p. 1591-1599, 2013.
- 10 CAMPBELL, M. K.; FARRELL, S. O. **Bioquímica**. 5. ed. Rio de Janeiro: Combo, 2009. p. 40-160.
- 11 ITO, T. et al. Identification of primary target of thalidomide teratogenicity. **Science**, Washington, v. 327, n. 5971, p. 1345-1350, março 2010.
- 12 DIAS, F. R. F.; FERREIRA, V. F.; CUNHA, A. C. Uma visão geral dos diferentes tipos de catálise em síntese orgânica. **Rev. Virtual Quím**, Niterói, v. 4, n. 6, p. 840-871, 2012.
- 13 SHUANG, L. et al. Technology prospecting on enzymes: aPplication, marketing and engineering. **Comput. Struct.l Biotech. J.**, Beijing , v. 2, n. 3, p. 1-11, 2012.
- 14 JEGANNATHAN, K. R.; NIELSEN, P. H. Environmental assessment of enzyme use in industrial production - a literature review. **J.Clean. Product.**, Amsterdam, v.42, p. 228-240, março 2013.

- 15 DE GRAAFF, C.; RUIJTER E.; ORRU, R. V. A. Recent developments in asymmetric multicomponent reactions. **Chem. Soc. Rev**, Cambridge, v. 41, n. 10, p. 3969-4009, 2012.
- 16 RUIJTER, E.; SCHEFFELAAR, R.; ORRU, R. V. A. Multicomponent reaction design in the quest for molecular complexity and diversity. **Angew. Chem. Int. Ed**, Weinheim, v. 50, n. 28, p. 6234 – 6246, 2011.
- 17 FAGAIN, O. Understanding and increasing protein stability. **Biochimica et Biophysica Acta**, Amsterdam, v. 1, n. 1252, p. 1-14, 1995.
- 18 LADERO, M. et al. Thermal and pH inactivation of an immobilized thermostable betagalactosidase from thermus SP strain T2: comparison to the free enzyme. **Biochemical Engineering Journal**, Amsterdam, v. 31, n. 1, p. 14-24, 2006.
- 19 LIMA, A. W. O.; ANGNES, L. Biocatálise em meios aquo-restritos: fundamentos e aplicações em química analítica. **Química Nova**, São Paulo, v. 22, n. 2, p. 229-245, 1999.
- 20 DORDICK, J. S. Designing enzymes for use in organic solvents. **Biotechnol. Prog**, Hoboken, v. 8, n. 4, p. 259-267, 1992.
- 21 HALLING, P. J. Biocatalysis in low-water media: understanding effects of reaction conditions. **Curr. Opin. Chem. Biol**, Kidlington, v. 4, n. 1, p. 74-80, 2000.
- 22 KLIBANOV, A. M. Why are enzymes less active in organic solvents than in water?. **Trends Biotechnology**, Maryland Heights, v. 15, n. 3, p. 97-101, 1997.
- 23 PETERS, G. H. et al. Dynamics of proteins in different solvent systems: analysis of essential motion in lipases. **Biophysical Journal**, St. Louis, v. 71, n. 5, p. 2245-2255, 1996.
- 24 HALLING, P. J. What can we learn by studying enzymes in non-aqueous media?. **Phil. Trans. R. Soc. Lond. B**, London, v. 359, n. 1448, p. 1287–1297, 2004.
- 25 OGINO, H.; ISHIKAWA H. Enzymes which are stable in the presence of organic solvents; **J. Biosci. Bioeng**, Osaka, v. 91, n. 2, p. 109-116, 2001.
- 26 IYER, P. V.; ANANTHANARAYAN, L. Enzyme stability and stabilization-aqueous and non-aqueous environment. **Process Biochemistry**, Oxford, v. 43, n. 10, p. 1019-1032, 2008.
- 27 Graf, J. et al. Structure and dynamics of the homologous series of alanine peptides: a joint molecular dynamics/NMR study. **J. Am. Chem. Soc.**, Washington, v. 129, n. 5, p. 1179-1189, 2007.
- 28 TOAL, S.; AMIDI, O.; SCHWEITZER-STENNER, R. Conformational changes of trialanine induced by direct interactions between alanine residues and alcohols in binary mixtures of water with glycerol and ethanol. **J. Am. Chem. Soc.**, Washington, v. 133, n. 32, p. 12728-12739, 2011.
- 29 VAGENENDE, V.; YAP, M.G.S.; TROUT, B.L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. **Biochemistry**, Washington, v. 48, n. 46, p. 11084-11096, 2009.

- 30 Han, W.-G. et al. Theoretical study of aqueous n-acetyl-l-alanine n'-methylamide: structures and raman, VCD, and ROA spectra. **J. Phys. Chem. B**, Washigton, v.102, n. 14, p. 2587-2602, 1998.
- 31 DROZDOV, A.N.; GROSSFIELD, A.; PAPPU, R.V. Role of solvent in determining conformational preferences of alanine dipeptide in water. **J. Am. Chem. Soc.**, Washigton, v. 126, n. 8, p. 2574-2581, 2004.
- 32 García-Prieto F.F.et al. E. Study on the conformational equilibrium of the alanine dipeptide in water solution by using the averaged solvent electrostatic potential from molecular dynamics methodology. **J. Chem. Phys.**, Weinheim, v. 135, n. 19, p. 194502-194509, 2011.
- 33 WARSHEL, A.; LEVITT, M. Theoretical studies of enzymatic reactions: dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysoenzyme. **J. Mol. Bio**, Cambridge, v. 103, n. 2, p. 227-249, 1976.
- 34 WEB OF KNOWLEDGE (Org.). **Levantamento**. [S.l.: s.n., 200]. Disponível em: <[https://apps.webofknowledge.com/Search.do?product=UA&SID=1AtirOPfgPPGvuiuYjK&search\\_mode=GeneralSearch&prID=16a93f7e-363d-4fa0-a47e-b3417bbdf8b3](https://apps.webofknowledge.com/Search.do?product=UA&SID=1AtirOPfgPPGvuiuYjK&search_mode=GeneralSearch&prID=16a93f7e-363d-4fa0-a47e-b3417bbdf8b3)>. Acesso em: 11 nov. 2015.
- 35 FDEZ GALVÁN, I. et al. ASEP/MD: a program for the calculation of solvent effects combining QM/MM methods and the mean field aPProximation. **Comput. Phys. Comm.**, Cambridge, v. 155, n. 3, p. 244-259, 2003.
- 36 SANCHEZ, M. L.; AGUILAR, M. A.; OLIVARES DEL VALLE, F. J. Study of solvent effects by means of averaged solvent electrostatic potentials obtained from molecular dynamics data. **J. Comput. Chem.**, Hoboken, v. 18, n. 3, p. 313-322, 1997.
- 37 SANCHEZ, M. L.; AGUILAR, M. A.; OLIVARES DEL VALLE, F. J. Solvent effects by means of averaged solvent electrostatic potentials: coupled method **J. Comput. Chem** , Hoboken, v. 21, n. 9, p. 705-715, 2000.
- 38 CORDEIRO, J. M.; BOSSO, A. R. S. A. Monte Carlo studies of n-methylformamide–dimethyl sulfoxide mixtures. **Journal of Molecular Liquids**, Amsterdam, v. 154, n.1, p. 36-40, 2010.
- 39 ALMEIDA, G. G.; CORDEIRO, J. M. M. A Monte Carlo revisiting of N-methylformamide and acetone. **Braz. Chem. Soc**, São Paulo, v. 22, n. 11, p. 2178-2185, 2011.
- 40 ALMEIDA, G. G; BORGES, A.; CORDEIRO, J. M. M. On the hydrogen bonding between Nmethylformamide and acetone and tetrahydrofuran **Chemical Physics**, Amsterdam, v. 434, p. 25-29 2014.
- 41-Levine, I. N. **Quantum chemistry**. 4. ed. New York: Prentice Hall, 1991. p. 629.
- 42 FRISCH, M. J. et al.**Gaussian 09**: revision A.1. Wallingford: Gaussian, 2009.
- 43 ALLEN, M. P.; TILDESLEY, D. J. **Computer simulation of liquids**. Oxford: Clarendon, 1987. p. 408.

- 44 BERENDSEN, H.J.C.; VAN DER SPOEL, D.; VAN DRUNEN, R. Gromacs - a message-passing parallel molecular-dynamics implementation. **Comp. Phys. Comm.**, Cambridge, v. 91, n. 1-3, p. 43-56, 1995.
- 45 SPIJKER, P. et al. **Potentials**. [S.l.: s.n.], 2007. Disponível em <<http://cbio.bmt.tue.nl/pumma/index.php/Theory/Potentials>> acesso em: 13 nov. 2015.
- 46 COUTINHO, K.; MORGON, N. H. **Métodos de química teórica e modelagem molecular**. Rio de Janeiro: Guanabara Koogan, 2007. p. 320-400.
- 47 SENN, H. M; THIEL, W. QM/MM methods for biomolecular systems. **Agew. Chem. Int. Ed.**, Weinheim, v. 48, n. 7, p. 1198-1229, 2009.
- 48 PETERS, G. H. et al. Dynamics of proteins in different solvent systems: analysis of essential motion in lipases. **Biophys.**, St. Louis, v. 71, n. 5, p. 2245-2255, nov. 1996.
- 49 CANUTO, S. (Ed.). **Solvation effects on molecules and biomolecules: an effective hamiltonian method from simulations: ASEP/MD**. Berlin: Springer Verlag, 2008. p. 580-591.
- 50 Fdez, I. G. et al. Use of the average solvent potential approach in the study of solvent effects **Advances in Quantum Chemistry**, Maryland Heights, v. 59, p. 59-97, 2010.
- 51 SANCHEZ, M. L. et al. Theoretical calculation of the stark component of the solute-solvent interaction energy. Validity of the mean field approximation in the study of liquids and solutions. **J. Phys. Chem. B.**, Washington, v. 106, n. 18, p. 4813-4817, 2002.
- 52 TAPIA, O.; GOSCINSKI, S. Self-consistent reaction field theory of solvent effects. **Mol. Phys.**, London, v. 29, n. 6, p. 1653-1661, 1975.
- 53 TAPIA, O.; SUSSMAN, F.; POULAIN, E. J. Environmental effects on H-bond potentials: A SCRF MO CNDO/2 study of some model systems. **Theor. Biol.**, London, v. 71, n. 1, p. 49-72, 1978.
- 54 TOMASI, J. et al. J. Theoretical chemistry in solution. some results and perspectives of the continuum methods and in particular of the polarizable continuum model. **J. Mol. Struct. Theochem**, Amsterdam, v. 234, p. 401-424, 1991.
- 55 TOMASI, J.; PERSICO, M. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. **Chem. Rev.**, Washington, v. 94, n. 7, p. 2027-2094, 1994.
- 56 CRAMER, C. J. et al. (Ed.). **Reviews in Computational Chemistry**. New York: VCH, 1995. v. 6, p. 1.
- 57 WARSHEL, A. **Computer modelling of chemical reactions in enzymes and solutions**. 6. ed. New York: Wiley-Interscience, 1991. p. 236.

- 58 FDEZ GALVÁN, I.; AGUILAR, M. A.; RUIZ LÓPEZ, M. F. Theoretical study of the 1,3-hydrogen shift of triazene in water. **J. Phys. Chem B.**, Washington, v. 109, n. 48, p. 23024-23030, 2005.
- 59 MARTÍN, M. E. et al. A multiconfiguration self-consistent field/molecular dynamics study of the  $(n \rightarrow \pi^*)$  transition of carbonyl compounds in liquid water. **J. Chem. Phys.**, Weinheim, v. 113, n. 15, p. 6308-6315, 2000.
- 60 MARTÍN, M. E. et al. Multiconfigurational self-consistent and molecular mechanics simulation of solvent effects on the  $n \rightarrow \pi^*$  blue shift of pyrimidine. **J. Mol. Struct. Theochem**, Amsterdam, v. 537, n. 1-3, p. 213-222, 2001.
- 61 MARTÍN, M. E. et al. Theoretical study of solvent effects on the  $1(n \rightarrow \pi^*)$  electron transition in acrolein. **J. Chem. Phys.**, Weinheim, v. 121, n. 8, p. 3710, 2004.
- 62 MUÑOZ LOSA, A. et al. Solvent effects on the low-lying excited states of a model of retinal. **J. Phys. Chem. B**, Washington, v. 110, n. 36, p. 18064-18071, 2006.
- 63 CAR, R.; PARRINELLO, M. Unified approach for molecular dynamics and density-functional theory. **Phys. Rev. Lett.**, College Park, v. 55, n. 22, p. 2471-2474, 1985.
- 64 LINDER, B. Reaction-field techniques and their applications to intermolecular forces. **Adv. Chem. Phys.**, Weinheim, v. 12, p. 225-282, 1967.
- 65 KARLSTRÖM, G; HALLE, B. A fluctuation approach to solvation in polar fluids. **J. Chem. Phys.**, Weinheim, v. 99, n. 10, p. 8056-8062, 1993.
- 66 GHONEIM, N.; SUPPAN, P. Solvatochromic shifts of non-dipolar molecules in polar solvents. **Spectrochim Acta**, Amsterdam, v. 51, n. 6, p. 1043-1050, 1995.
- 67 FIELD, M. J.; BASH, P. A.; KARPLUS, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. **J. Comput. Chem.**, Hoboken, v. 11, n. 6, p. 700-733, 1990.
- 68 LUZHKOV, V.; WARSHEL, A. Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. **J. Comput. Chem.**, Hoboken, v. 13, n. 2, p. 199-213, 1992.
- 69 GAO, J. Absolute free energy of solvation from Monte Carlo simulations using combined quantum and molecular mechanical potentials. **J. Phys. Chem.**, Washington, v. 96, n. 2, p. 537-540, 1992.
- 70 VASILYEV, V. V.; BLIZNYUK, A. A.; VOITYUK, A. A. A combined quantum chemical/molecular mechanical study of hydrogen-bonded systems. **Int. J. Quantum Chem.**, New York, v. 44, n. 5, p. 897-930, 1992.
- 71 THÉRY, V. et al. Quantum mechanical computations on very large molecular systems: the local self-consistent field method. **J. Comput. Chem.**, Hoboken, v. 15, n. 3, p. 269-282, 1994.

- 72 THOMPSON, M. A.; GLENDENING, E. D.; FELLER, D.; The nature of K<sup>+</sup>/crown ether interactions: a hybrid quantum mechanical-molecular mechanical study. **J. Phys. Chem.**, Washington, v. 98, n. 41, p. 10465-10476, 1994.
- 73 OKUYAMA-YOSHIDA, N.; NAGAOKA, M.; YAMABE, T. Transition-state optimization on free energy surface: toward solution chemical reaction ergodography. **Int. J. Quantum Chem.**, New York, v. 70, n. 1, p. 95-103, 1998.
- 74 OKUYAMA-YOSHIDA, N. et al Structure optimization via free energy gradient method: aPplication to glycine zwitterion in aqueous solution. **J. Chem. Phys.**, Weinheim, v. 113, n. 9, p. 3519-3524, 2000.
- 75 HIRAO, H.; NAGAE, Y.; NAGAOKA, M. Transition-state optimization by the free energy gradient method: aPplication to aqueous-phase Menshutkin reaction between ammonia and methyl chloride. **Chem. Phys. Lett.**, Weinheim, v. 348, n. 3-4, p. 350-356, 2011.
- 76 ZHANG, Y.; LIU, H.; YANG, W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined *ab initio* QM/MM potential energy surface. **J. Chem. Phys.**, Weinheim, v. 112, n. 8, p. 3483-3492, 2000.
- 77 FDEZ. GALVÁN, I. et al. Geometry optimization of molecules in solution: Joint use of the mean field approximation and the free-energy gradient method. **J. Chem. Phys.**, Weinheim, v. 118, n. 1, p. 255-263, 2003.
- 78 BANERJEE, A. et al. Search for stationary points on surfaces. **J. Phys. Chem**, Washington, v. 89, n. 1, p. 52-57, 1985.
- 79 ZWANZIG, W. R. High-temperature equation of state by a perturbation method. I. nonpolar gases **J. Chem. Phys.**, Weinheim, v. 22, n. p. 1420-1426, 1954.
- 80 Fdez. Galván, I. et al. Comparison of three effective Hamiltonian models of increasing complexity: triazene in water as a test case. **J. Chem. Phys.**, Weinheim, v. 124, n. 21, p. 214504-214508, 2006.
- 81 VARGAS, R. et al. Conformational study of the alanine dipeptide at the mp2 and dft levels. **J. Phys. Chem. A**, Washington, v. 106, n. 13, p. 3213-3218, 2002.
- 82 SÁNCHEZ, M. L.; AGUILAR, M. A.; OLIVARES DEL VALLE, F. J. A mean field aPproach that combines quantum mechanics and molecular dynamics simulation: the water molecule in liquid water. **J. of Molecular Structure Theochem**, Amsterdam, v. 426, n. 1-3, p. 181-190, 1998.
- 83 REFSON, K. Moldy: a portable molecular dynamics simulation program for serial and parallel computers. **Comput. Phys. Commun.**, Cambridge, v. 126, n. 3, p. 310-329, 2000.
- 84 JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. J. Development and testing of the opl's all-atom force field on conformational energetics and properties of organic liquids. **Am. Chem. Soc.**, Washignton, v. 118, n. 45, p. 11225-11236, 1996.

- 85 Jorgensen, W. L. et al. Comparison of simple potential functions for simulating liquid water. **J. Chem. Phys.**, Weinheim, v. 79, n. 2, p. 926-935, 1983.
- 86 CHIRLIAN, L. E.; FRANCL, M. M. J. Atomic charges derived from electrostatic potentials: a detailed study. **Comput. Chem.**, Hoboken, v. 8, n. 6, p. 894-905, 1987.
- 87 BRENEMAN, C. M.; WIBERG, K. B. J. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis **Comput. Chem.**, Hoboken, v. 11, n. 3, p. 316-373, 1990
- 88 DARDEN, T.; YORK, D.; PEDERSEN L. Particle mesh Ewald: an  $N \cdot \log(N)$  method for Ewald sums in large systems. **J. Chem. Phys.**, Weinheim, v. 98, n. 12, p. 10089-10092, 1993.
- 89 HOOVER, W. G. Canonical dynamics: equilibrium phase-space distributions. **Phys. Rev. A.**, Washington, v. 31, n. 3, p. 1695-1697, 1985.
- 90 GEKKO, K.; TIMASHEFF, S. N. Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures. **Biochemistry**, Washington, v. 20, n.16, p. 4667-4676, 1981.

**EM ANEXO SEGUE A VERSÃO SUBMETIDA À PUBLICAÇÃO SOBRE O ESTUDO REALIZADO COM O DIPEPTÍDEO DE ALANINA**

## 8-ANEXO

### Conformational Changes of the Alanine Dipeptide in Water-Ethanol Binary Mixtures.

*Glauco G. Almeida,<sup>1</sup> João M. M. Cordeiro,<sup>1</sup> M. Elena Martín<sup>2</sup> and Manuel A. Aguilar<sup>2</sup>*

Correspondence to: M.A. Aguilar (e-mail: maguilar@unex.es)

<sup>1</sup>*Faculdade de Engenharia, Universidad Estadual Paulista "Julio de Mesquita Filho", Ilha Solteira, (Brasil)*

<sup>2</sup>*Área de Química Física, University of Extremadura, Avda. Elvas s/n, Edif. José M<sup>a</sup> Viguera Lobo 3<sup>a</sup>, planta Badajoz, 06006, (Spain)*

#### Abstract

Experimental work developed in the last years has evidenced the capacity of alcohols and polyalcohols to modify the energy landscape of proteins. However, the mechanism underlying this effect is not clear. Taking as a model system the alanine dipeptide (AD) we perform a QM/MM study in water, ethanol and a 40-60 % in volume water-ethanol binary mixture. The AD molecule was described at MP2/aug-cc-pVDZ level. In polar solution, only  $\alpha_R$  and PPII conformers contribute in aPPreciable way to the conformational equilibrium. The final in solution  $\alpha_R$ -PPII free energy difference is determined from the interplay between the internal energy of the dipeptide and the solute-solvent interaction energy. Internal energy favors the formation of PPII whereas, on the contrary, solute-solvent interaction is favorable to  $\alpha_R$ , so any factor that decreases the solute-solvent interaction energy will increase the PPII population. The addition of ethanol increases the stability of the PPII conformer. Our results point to the presence of preferential solvation in this system, the composition of the first solvation shell in the binary mixture being dominated by water molecules. Remarkably, this fact does not affect to the differential conformational stability that is controlled by long-range interactions. From the analysis of solvent density maps it is concluded that ethanol molecules are more likely found around the carbonyl group of the N-terminal end and around methylic side chains.

## I.- Introduction

Despite the doubts expressed by some authors about the utility of using small peptides as model systems for describing the structure and dynamics of proteins,<sup>1</sup> they have been profusely studied<sup>2-37</sup>, as the general belief is that they can provide insight into important aspects of the conformational equilibrium. Alanine polypeptides span the main conformations of protein and are therefore a good laboratory where one can check the influence of the solvent polarity or the presence of a co-solvent on the structure as well as the relative quality of the different force fields or the quantum mechanical level necessary to adequately describe the system, among other questions.

In alanine polypeptides, the polyprolyne II conformer (PPII), has been identified as the most stable in aqueous solution,<sup>34</sup> followed by the  $\beta$  strand form, and  $\alpha_R$ , although in dipeptides the relative stability order of  $\beta$  strand and  $\alpha_R$  are interchanged and the conformational equilibrium is dominated by the PPII and  $\alpha_R$  conformers. Experimentally, the relative populations vary depending on the conditions (temperature and composition, mainly) imposed to the sample. Recent studies<sup>38,39</sup> on the alanine tripeptide (AT) have concluded that the presence of alcohols as ethanol or glycerol as well as the increase of the temperature can modify the conformational equilibrium favoring the presence of PPII. It is not clear, however, which are the reasons behind this behavior. Some authors<sup>40</sup> indicate that water stabilizes PPII by forming a bridge between adjacent carbonyl and amide groups through hydrogen bonding, while other authors<sup>41</sup> propose that steric interactions are the decisive forces to determine conformational preferences.

Theoretical studies have confirmed the prevalence of the PPII and  $\alpha_R$  conformers, although the differences in the relative stabilities are very small,<sup>42</sup> and the populations very sensitive to the calculation level. These differences can become even smaller<sup>42</sup> when correlation effects beyond the MP2 level are included in the calculations. In these conditions even vibrational and rotational contributions to the free energy could determine the relative stability. In a recent paper, Nerenberg and Head-Gordon<sup>43</sup> analyzed the accuracy of the TIP3P and TIP4P-Ew water models, among others, to generate conformational ensembles for several tripeptides and they found that the use of improved water models can yield more accurate conformational ensembles. The effect that the protein force field has on the conformational preference of short peptides has also been analyzed. It has been found<sup>44</sup> that most force fields overestimate the stability of the helical conformation. Consequently, several authors have proposed modifications of the force field, especially in the values associated to the  $\phi$  dihedral angle (see Figure 1).

It is well established that solvent modifies the energy landscape of alanine dipeptide (AD).<sup>20,45</sup> Thus, while in gas phase cyclic structures forming intramolecular hydrogen bonds are found to be more stable, in particular C7<sub>eq</sub> and C5 ( $\beta$  strand), in aqueous solution there is a destabilization of these structures in favor of extended forms as they permit a better

interaction with the solvent molecules. The same trend has been found in N-methylformamide, another molecule used as model for the peptide bond.<sup>46-48</sup> In aqueous solutions, PPII and  $\alpha_R$  become the most stable forms. In these conformers the intramolecular hydrogen bonds are replaced by intermolecular hydrogen bonds with the water molecules. The calculated relative stability order is very sensitive to the force-field parameters used, mainly the charge values. The C5 conformer has been detected in some experiments, albeit in small concentrations.<sup>13</sup> It deserves special attention as its formation is not easy to predict from a theoretical point of view. Its presence depends on the quantum calculation level (DFT or MP2) or the solvent model used (rigid or flexible). The difficulty of finding C5 in solution lies in its evolution toward  $\alpha_R$  form. Other conformers, including C7<sub>eq</sub> or  $\delta_R$ , lie at too high energy to contribute significantly to the in solution conformational equilibrium at room temperature.

The competition between intramolecular and intermolecular hydrogen bonds plays an important role in determining the relative stability of the different conformers. Consequently, simplified solvent models (continuum, Langevin dipoles, etc.) that neglect specific interactions seem, a priori, not to be adequate for this type of studies. Additionally, it is important to properly account for thermal fluctuations in order to explain temperature and entropic effects. QM/MM methods permit the inclusion of the aforementioned factors and, consequently, their use seems to be especially adequate for this task. Furthermore, they permit to get the polypeptide geometry without relying on intramolecular parameters. In what follows, the effect that the addition of an alcohol as ethanol has on the AD conformational equilibrium is studied using a QM/MM method. With this study, we expect to get insight into the possible presence of preferential solvation and about the competition between bulk solvent effects and specific interactions in determining the relative stability order.

The rest of the paper is organized as follows: in section II the main features of the solvation method used will be explained. As we will show, the method permits the combination of high-level quantum calculations in the description of the solute with a detailed description through molecular mechanics force fields of the solvent. In section III the computational aspects will be detailed. Then in section IV the conformational equilibrium of AD in water, ethanol and a 40-60% water-ethanol mixture will be discussed paying special attention to the structure of the first solvation shell and to the presence or not of preferential solvation. Finally, the main conclusions are reported in section V.

## II.- Method

Solvent effects on the conformational equilibrium of the AD were studied with the average solvent electrostatic potential/molecular dynamics (ASEP/MD) method<sup>49-52</sup> developed in our laboratory. ASEP/MD is a QM/MM method<sup>53-55</sup> that uses the mean field aPProximation. In the last years, the original method and related proposal<sup>56-57</sup> have been used in the study of molecular spectra, chemical reactions, conformational equilibria, etc.

In the mean field aPProximation<sup>58</sup> the average value of any mechanical property of the system is replaced by the value that this property takes in the average configuration. This permits the reduction in the number of quantum calculation from several thousands to only a few tens and, as a result, the quantum calculation level used in the dipeptide description can be similar to that used in gas phase, making easier a direct comparison between gas phase and in solution results.

In ASEP/MD, as in other QM/MM methods<sup>53-55</sup>, the solute wave function is obtained by solving the following effective Schrödinger equation

$$(\hat{H}_{QM} + \hat{H}_{QM/MM}^{elect} + \hat{H}_{QM/MM}^{vdw})|\Psi \rangle = E|\Psi \rangle \quad (1)$$

where  $\hat{H}_{QM}$  is the in vacuo solute molecular Hamiltonian. ASEP/MD differs from other QM/MM methods in the fact that the electrostatic solute-solvent interaction term ( $\hat{H}_{QM/MM}^{elect}$ ) is calculated using the mean field aPProximation:

$$\hat{H}_{QM/MM}^{elect} = \int dr \cdot \hat{\rho} \cdot V_{ASEP}(r) \quad (2)$$

Here,  $V_{ASEP}(r)$  is the average electrostatic potential generated by the solvent in the volume occupied by the solute. It is calculated as a statistical average over the configuration space obtained in molecular dynamics (MD) simulations. The potential thus obtained is numerical; it is convenient from a practical point of view to represent it through a set of point charges. Technical details about the determination of the number, position, and values of the charges have been reported previously.<sup>50</sup> Finally,  $\hat{\rho}$  is the solute charge density operator and  $\hat{H}_{QM/MM}^{vdw}$  is the Hamiltonian for the van der Waals interaction. When represented by a Lennard-Jones potential this term is a function only of the nuclear coordinates and it is a constant added to the energy, so it has no direct effect on the dipeptide wave function.

In ASEP/MD, the solute charge distribution polarizes in response to the solvent electrostatic potential. Obviously, the solvent structure and hence the solvent potential is a function of the solute charge distribution, i.e.,  $V_{ASEP}(r)$  and  $\hat{\rho}$  depend mutually on each other and Eqs. 1 and 2 must be solved iteratively. The process finishes when the dipeptide charge distribution and the solvent structure around it become mutually equilibrated. The basic scheme is straightforward. The procedure begins by performing one quantum calculation for the solute molecule in gas phase. The solute charge distribution obtained from this calculation is then used as input in a complete molecular dynamics (MD) simulation. Once the structure of the solvent around the solute molecule has been obtained from a representative set of solvent configurations, the ASEP (averaged solvent electrostatic potential) is calculated and introduced into the molecular Hamiltonian of the solute. The electronic wavefunction of the solute, now in solution, can be obtained by solving the associated effective Schrödinger equation, Eq. 1. A new solute charge distribution can be calculated from the solute

wavefunction and used again as input in a new molecular dynamics simulation. This process is repeated until convergence in the solute charges is achieved

The dipeptide geometry is optimized in the presence of the solvent perturbation using a technique described in a previous paper<sup>59</sup> and based on the joint use of the free-energy gradient method<sup>60-62</sup> (FEG) and the mean field approximation<sup>58</sup>. The technique has been successfully applied to the geometry optimization of ground and excited states of molecules in solution. It is worthy of note that in the FEG method a rigid body approximation is used for the solute, i.e., the solute geometry is kept fixed during the MD simulations. Next, the solvent contribution to the mean values of the total force,  $F$ , and the Hessian,  $G$ , are calculated from a representative set of solvent configurations and added to the internal solute contribution. The force and the Hessian are approximated as:

$$F(R) = -\frac{\partial A(R)}{\partial R} = -\left\langle \frac{\partial E}{\partial R} \right\rangle = -\left\langle \frac{\partial E_{QM}}{\partial R} \right\rangle - \left\langle \frac{\partial E_{int}}{\partial R} \right\rangle \approx \frac{\partial \langle E \rangle}{\partial R}$$

$$G(R, R') = -\frac{\partial^2 A(R)}{\partial R^2} = -\left\langle \frac{\partial^2 E}{\partial R \partial R'} \right\rangle - \beta \left\langle \frac{\partial E}{\partial R} \left( \frac{\partial E}{\partial R'} \right)^T \right\rangle + \left\langle \frac{\partial E}{\partial R} \right\rangle \left\langle \left( \frac{\partial E}{\partial R'} \right) \right\rangle^T \approx \frac{\partial^2 \langle E \rangle}{\partial R \partial R'}$$

where the brackets indicate a statistical average. Afterward, the force and Hessian values are used to obtain a new geometry with the Rational Function Optimization method<sup>63</sup>.

The change of the dipeptide energy when it passes from gas phase to solution can be written as:

$$\Delta E = \langle \Psi | \hat{H}_{QM} + \hat{H}_{QM/MM}^{elect} + \hat{H}_{QM/MM}^{vdw} | \Psi \rangle - \langle \Psi^0 | \hat{H}_{QM} | \Psi^0 \rangle \quad (3)$$

$$\Delta E = \langle \Psi | \hat{H}_{QM} | \Psi \rangle - \langle \Psi^0 | \hat{H}_{QM} | \Psi^0 \rangle + \langle \Psi | \hat{H}_{QM/MM}^{elect} | \Psi \rangle + E_{QM/MM}^{vdw} \quad (4)$$

$$\Delta E = E_{QM} - E_{QM}^0 + E_{QM/MM}^{elect} + E_{QM/MM}^{vdw} \quad (5)$$

$$\Delta E = E_{dist} + E_{int} = E_{dist} + E_{int}^{elect} + E_{int}^{vdw} \quad (6)$$

The first term in this expression,  $E_{dist} = \langle \Psi | \hat{H}_{QM} | \Psi \rangle - \langle \Psi^0 | \hat{H}_{QM} | \Psi^0 \rangle$ , defines the distortion energy of the dipeptide, i.e., the energy spent in its polarization; the second term,  $E_{int}$ , is the solute solvent interaction energy that in turn can be split in the electrostatic and van der Waals contributions.  $\Psi$  is the dipeptide wavefunction polarized by the solvent and  $\Psi^0$  is the in gas phase wavefunction.

In order to determine the relative stability of the dipeptide conformers, it is needed to calculate free energy differences. Here, we follow a dual level strategy where the dipeptide contributions are quantum-mechanically calculated while the solute-solvent interaction contribution is classically obtained by using the free energy perturbation method. The final expression for the free energy change associated to the  $A \rightarrow B$  transformation (A and B being two different conformers) is:

$$\Delta A = \Delta E_{\text{QM}} + \Delta A(\text{rot+vib}) + \Delta A_{\text{int}} \quad (7)$$

here  $\Delta E_{\text{QM}} = E_{\text{QM}}^{\text{B}} - E_{\text{QM}}^{\text{A}}$  is the internal energy variation and  $\Delta A(\text{rot+vib})$  accounts for the change in the zero point energy and rotational and vibrational contributions during the  $A \rightarrow B$  transformation. This term is included because during the simulation the dipeptide molecule was approximated to a rigid body. These rotational and vibrational components are calculated applying the rigid rotor and harmonic oscillator models and using the in solution geometry and frequencies. Recently, it has been shown<sup>64</sup> that this approximation permits the reduction, in about one order of magnitude, of the free energy errors due to the use of constrained geometries. Finally, the  $\Delta A_{\text{int}}$  term was calculated with the free energy perturbation method using the in solution charges and geometries obtained with ASEP/MD.

### Computational details

In order to study the conformational equilibrium of the AD molecule in water, ethanol and water-ethanol, MP2/aug-cc-pVDZ calculations were performed. This calculation level provides a correct description of the AD conformational space at a reasonable computational cost and it has been previously<sup>12,45</sup> used in gas phase studies, so it permits to directly analyze solvent effects on the geometry and charge distributions of AD. The use of a basis set with diffuse functions in solvent effect studies can be problematic since, in some situations, they can produce an overpolarization of the solute charge distribution.<sup>65</sup> In order to check this effect we also ran some calculation with the cc-pVDZ basis set.

All the in solution calculations were carried out with the ASEP/MD method described in the previous section. As it was indicated before, ASEP/MD allows, through the combination of molecular dynamics simulations and ab initio quantum calculations, a correct description of the solute-solvent interactions and of the effect that the solvent causes on the structural features of the solute. It is worth remembering that this result can be reached with a reduced number of high-level quantum calculations, overcoming the main drawback of conventional QM/MM methods. Ab initio calculations were carried out with the Gaussian09<sup>66</sup> package and Gromacs<sup>67,68</sup> was employed for the molecular dynamics simulations with flexible solvent molecules. MD simulations with rigid solvent molecules were performed with Moldy<sup>69</sup>. The simulations contain one rigid AD molecule and a number of flexible solvent molecules depending on the system considered. We used 1496 TIP3P water molecules, 657 OPLS-AA ethanol molecules and finally in the case of the water-ethanol mixture 606 water molecules and 280 ethanol molecules. The Lennard-Jones parameters for the alanine

dipeptide were taken from the OPLS-AA force field<sup>70</sup> and the atomic charges were updated at each ASEP/MD cycle using the CHELPG<sup>71,72</sup> method. It has been shown that the OPLS-AA/TIP3P combination reproduces adequately the conformational space of alanine dipeptide<sup>45</sup> in water solution and we hope that it could have the same behavior in the water-ethanol mixture.

Periodic boundary conditions were applied in all directions. Short-range electrostatic interactions were cutoff at 1.3 nm and long-range interactions were calculated with the Particle-Mesh Ewald (PME) method<sup>73</sup>. The temperature was fixed at 298 K with the Nosé-Hoover thermostat<sup>74</sup>. Each simulation was run in the NVT ensemble for 500 ps, with a time step of 1 fs, where the first 200 ps were used for equilibration and the last 300 ps for production. In solution final results were obtained by averaging the last five ASEP/MD cycles, and therefore they represent a 1.5 ns average.

Finally, for each free energy perturbation calculation that connects two conformers, A and B, 100 MD simulations of 500 ps each were performed. In all cases the free energy for the direct,  $A \rightarrow B$ , transformation differs by less than 0.2 kcal/mol from the free energy of the reverse,  $B \rightarrow A$ , transformation. The final result is the mean value of the direct and reverse transformation.

The atom numbering and representative dihedral angles of the AD molecule are displayed in Figure 1. As usual, the backbone disposition is referred according to the values of the central dihedral angles  $\phi$  and  $\psi$  ( $C_5N_7C_9C_{11}$  and  $N_7C_9C_{11}N_{17}$ , respectively).

## RESULTS AND DISCUSSION

Although alcohol co-solvents have been profusely used as protein stabilizers due to their ability to modify the energy landscape, the underlying mechanism is still under debate. Hydrogen bonding<sup>4</sup>, steric interactions<sup>41</sup> or preferential solvation<sup>75</sup> have been proposed as the determinant effects in the protein stabilization. Studies carried out in small peptides conclude that the presence of small quantities of alcohol shift the conformational equilibrium favoring the PPII conformer. Other studies argue about the exclusion of alcohol molecules from the first solvation shell. In this regard, recent vibrational spectra of AT<sup>38</sup> have revealed interactions between aliphatic side chains of AT and alcohol molecules in binary mixtures. In order to clarify these points, the alanine dipeptide was chosen as a model system. Our main aim is to determine the factors, including solvent effects, that control the conformational equilibrium in AD. We consider three solvents: water, ethanol and a 40-60 % in volume water-ethanol mixture, which corresponds to a water molar fraction of 0.68.

Let's start by clarifying some technical details. As it was indicated in previous sections, the use of diffuse functions in describing molecules in solution through QM/MM methods could be problematic due to a possible solute overpolarization effect.<sup>65</sup> In most QM/MM methods (see reference 76 for an exception) only the electrostatic component of the

solute-solvent interaction energy is included into the solute Hamiltonian. The effect of the solute-solvent electron exchange on the solute wavefunction is completely neglected. As a consequence, the overlapping of the solute wavefunction with the solvent molecules is not penalized and the attractive interaction between the solute's electrons by positive charges in the solvent could be overestimated. This unphysical effect can be especially important in the case of diffuse functions as they extend to large distances into the solvent. Table 1 compares in solution results using *cc*-pVDZ and *aug-cc*-pVDZ Dunning basis sets. The free energy change associated with the transformation  $\alpha_R \rightarrow$  PPII in water is similar with the two basis sets: 0.5 and 0.7 kcal/mol with *cc*-pVDZ and *aug-cc*-pVDZ, respectively. The induced dipoles are somewhat larger with the more extended basis set as consequence of a better description of the AD polarizability, but the values are within the expected range. In sum, in the system under study diffuse functions seem not to produce artifacts and solute overpolarization can be ruled out. As the *aug-cc*-pVDZ basis set has been profusely used in theoretical calculation of AD, and in order to facilitate the comparison with previous results, the rest of calculations were performed using this basis set.

Table 2 displays the free energy difference and its components (see Eq. (7)) for the  $\alpha_R \rightarrow$  PPII transformation in the three solvents considered. The free energies were calculated using the dual-level method indicated in the previous section (quantum estimation for the solute contribution, classical for the solute-solvent interaction). A positive sign means a larger stability of  $\alpha_R$  with respect to the PPII conformer. In the three solvents the AD conformational equilibrium is dominated by  $\alpha_R$  and PPII forms. The population of any other conformer can be considered as negligible and they have not been included in the table. This contrasts with alanine tripeptide and longer alanine polypeptides where C5 population can make a substantial contribution to the conformational equilibrium. We tried to optimize the C5 conformer starting from several structures, however in all cases the optimization evolves toward the  $\alpha_R$  structure. In a previous paper using a rigid body approximation for the solvent and a gradient-based optimization method it was possible to obtain a C5 optimized geometry. Using this geometry, the free energy change has been re-calculated assuming flexible solvent, the vibrational correction was not included as it is not a minimum under the new computational conditions. In water, our calculations place the C5 conformer 3.3 and 4.0 kcal/mol higher in energy than PPII and  $\alpha_R$  respectively, confirming the low population of this conformer at room temperature.

At the MP2/*aug-cc*-pVDZ level, the  $\alpha_R$  conformer is slightly more stable than PPII in water and in the mixture; on the contrary, in ethanol PPII becomes the most stable conformer. The result in water differs from the results of a previous paper<sup>45</sup> where the equilibrium was slightly dominated by the PPII conformer. The difference probably comes from the different models used in describing the solvent (rigid molecules in the previous one and flexible here). Furthermore, due to the flatness of the surface around the global minimum the final result is sensitive to the used optimization method (gradient-based before and Hessian-based now). In any case, the free energy differences are very small and the final result turns to be very sensitive to factors such as the solvent model or the values of the charges used during the

simulation (we will come back to this point later). Which conformer is the most stable is the result of the interplay between the internal energy and the solute-solvent interaction energy. As a general conclusion the internal energy,  $\Delta E_{QM}$ , favors PPII conformer; on the contrary, the interaction energy,  $\Delta A_{int}$ , favors  $\alpha_R$ . Rotational and vibrational contributions are completely negligible for water and water-ethanol mixture but they are somewhat more important for ethanol. Small changes in the experimental conditions (temperature or solvent) or in the potential parameters (geometry, charges, etc.) could shift the equilibrium toward one or other conformer. In general any factor decreasing the solute-solvent interaction energy increases the relative stability of the PPII conformer.

In agreement with the experiment, the population of PPII increases with the addition of ethanol, even though the free energy variations are small, about 1 kcal/mol, passing from 0.6 in water to -0.3 kcal/mol in ethanol if a flexible solvent model is used and from 0.2 to -0.8 kcal/mol with a rigid solvent model. According to our data and using the flexible model, the PPII population increases from 27% to 62% when AD passes from water to ethanol. In order to understand this behavior it is opportune to perform a deeper analysis of the AD-solvent interactions.

We start by analyzing the solvent effects on the AD structure. Tables 3 and 4 display the main geometrical parameters of PPII and  $\alpha_R$  optimized at MP2/aug-cc-pVDZ level in water, ethanol and water-ethanol solution. Direct comparison with gas phase values is not possible as PPII is not stable in vacuum and the  $\alpha_R$  conformer appears in solution from the structural modification of  $\delta_R$ . The most significant changes introduced by the solvent are related with the peptidic bond lengths. As a general trend C—O bond lengths increase with the solvent polarity whereas C—N distances decrease. These changes are compatible with the stabilization by the solvent of the charge-separated resonant form of the peptide bond. Furthermore, N—H bond lengths also increase. The increase of N—H and C—O bond lengths points to the formation of stronger intermolecular hydrogen bonds with water than with ethanol molecules. The bond lengths in the water-ethanol mixture are similar to those found in water, which suggests the presence of preferential solvation. Tables 3 and 4 also show  $\phi$  and  $\psi$  dihedral angles. The effect of the solvent on these angles is also small. In sum, despite the high values of the interaction energies (see below), the increase of the solvent polarity produces only small changes in the geometry of the two conformers.

Table 5 collects distortion energies,  $E_{dist}$ . According to equation 4, the distortion term is calculated as the difference between the internal energy in solution  $E_{QM}$  and the corresponding energy in gas phase  $E_{QM}^0$ . Given that, as indicated above, neither PPII nor  $\alpha_R$  have defined geometries in gas phase,  $E_{QM}^0$  was calculated using the in solution geometry. The gas phase internal energies are almost the same in the three solvents. On the contrary,  $E_{dist}$  strongly depends on the solvent polarity. The high value of this component evidences the great polarization suffered by the system during the solvation process, a fact confirmed by the increase in the AD dipole moment values. The differences in  $E_{dist}$  between the two conformers vary between -2.7 and -3.2 kcal/mol depending of the solvent considered, being  $E_{dist}$

systematically larger in the  $\alpha_R$  conformer than in PPII. As expected, the larger the interaction energy, the larger the distortion energy, but with the opposite sign.

As for the solute-solvent interaction energies,  $E_{\text{int}}$  in Table 5, they increase in about a 60 % when we pass from ethanol to the water-ethanol mixture but only 7% from the mixture to water. In the three solvents the  $E_{\text{int}}$  term is larger in the  $\alpha_R$  conformer than in PPII. The solute-solvent interaction energies clearly depart from the ideal (linear) behavior. In fact, it is possible to define an “effective molar fraction”  $x_W = \frac{E_{\text{int}}^{\text{W/E}} - E_{\text{int}}^{\text{E}}}{E_{\text{int}}^{\text{W}} - E_{\text{int}}^{\text{E}}}$ , where the superscripts W, E and W/E refer to water, ethanol and water-ethanol, respectively. Using the  $E_{\text{int}}$  values of table 5 for  $\alpha_R$  and PPII conformers one gets values of 0.85 and 0.83, respectively, clearly greater than the actual value 0.68, confirming the presence of preferential solvation. A preferential solvation is said to exist when the local composition of the mixture in the neighborhood of the solute molecule is different from the bulk composition. It is, in general, originated by specific interactions; however, even in the absence of such specific interactions, a polar solute can induce a change in the composition of its environment in a process called “dielectric enrichment”. As it is showed below, in the water-ethanol mixture the ethanol molecules are almost expelled from the first solvation shell, being the solute mainly solvated by water molecules.

Figure 2 permits to analyze the composition of the solvation shell around the AD molecule in the water-ethanol mixture. This figure displays the local molar fraction of water at several distances for the water-ethanol system. In the system configurations extracted from the molecular dynamics simulations, we counted the number of molecules of each solvent species located at less than a given distance of the solute. We define this “distance” as the distance between any solute atom and the oxygen atom of the solvent molecule. The local molar fraction of water in the first solvation shell for  $\alpha_R$  is 0.93 at 2.1 Å and 0.84 at 2.3 Å. The local concentration near the solute molecule is significantly larger than the bulk concentration ( $x_w=0.68$ ), up to a distance of around 5 Å. This proves that there is actually preferential solvation in this system.

Next, we will deal with the differences in the solvent structure around the two isomers. Radial distribution functions (rdf), coordination numbers and tridimensional density maps of the solvent molecules around AD were analyzed. Figure 3 displays radial distribution functions around the carbonyl oxygen and amide hydrogen atoms. It is worth noting that the height of the first peak of the oxygen(AD)-hydrogen(ethanol) and hydrogen(AD)-oxygen(ethanol) rdfs decreases with respect to the value in neat ethanol, see Figure 3a. The opposite is found in those rdfs involving water molecules, see Figure 3b. This fact confirms again the presence of preferential solvation. Figure 3c shows as the water molecule distribution in the first solvation shell is very similar in the two conformers, showing high water molecule densities around the amine and carbonyl groups. There are more differences in the ethanol distribution, while in  $\alpha_R$  carbonyl and amine groups interact with the ethanol

molecules, in PPII just O12 (the carbonyl oxygen of the N terminal end) shows significant concentrations of ethanol molecules around it.

Table 6 shows coordination numbers around some AD atoms. Several conclusions can be extracted from this table: 1) The total coordination number decreases with increasing ethanol concentration, 2) In the water-ethanol mixture about the 84% of the molecules in the first solvation shell are water molecules. On average, only one ethanol molecule is close to AD (in ethanol, about 6 molecules are in the first solvation shell). 3) The number of solvent molecules in the first solvation shell is slightly larger in  $\alpha_R$  than in PPII, although the differences are small, about 0.3-0.4 molecules. These differences come mainly, again, from those solvent molecules close to the carbonyl groups. But while the carbonyl group of the C-terminal end favors PPII, the N-terminal carbonyl group favors  $\alpha_R$ .

Figures 4 and 5 show density maps of solvent molecules around  $\alpha_R$  and PPII for the water-ethanol mixture. These figures have been generated using different atom density values in order to highlight different solvent structural features. In both conformers, and in agreement with rdf results, the first solvation shell is mainly populated by water molecules, only around the carbonyl group of the N-terminal end it is possible to find appreciable densities of ethanol molecules (see red and grey elements), the ethanol densities being somewhat larger in PPII than in the  $\alpha_R$  conformer (see also Fig. 3). In the  $\alpha_R$  conformer (but not in PPII) it is possible to note the presence of ethanol molecules close to the methyl side groups, a fact that has been experimentally highlighted<sup>38</sup>. An interesting point to clarify is whether a water molecule can form a bridge between the carbonyl and amine groups of AD or, on the contrary, the hydrogen bond involves two independent water molecules. This water bridge has been proposed as the origin of the greater stability of PPII. Figure 5 shows that carbonyl and amine groups have completely separated solvation shells, so our study does not find evidence of the presence of a water bridge between these two functional groups.

Another important point to clarify is the effect of preferential solvation on the  $\alpha_R$ -PPII differential stability. Figure 6 displays the interaction energy as a function of the distance between solvent atoms and any AD atom. Only the water-ethanol system is shown, as the behavior is similar in the other solvents. The  $E_{\text{int}}$  difference between  $\alpha_R$  and PPII strongly depends on the distance. It reaches its maximum value between 2.1 and 2.2 Å depending on the solvent and coinciding with the first solvation shell. Then the difference decreases to zero at 2.9-3.1 Å, close to the second solvation shell, finally, the difference slowly increases until it reaches the final value, which always favors the  $\alpha_R$  conformer. This difference decreases when the solvent polarity increases. In sum, although the first solvation shell favors clearly the  $\alpha_R$  conformer, this effect is canceled by the contribution of the second solvation shell; consequently, the larger  $E_{\text{int}}$  value in  $\alpha_R$  is a long-range effect associated to the larger polar character of this conformer.

Complementary and correlated to the solvent disposition around the different functional groups, it is interesting to analyze the contribution of those groups to the classical

solute-solvent interaction energy, see Table 7. The AD molecule was split in several parts according to the functional groups: CH<sub>3</sub>, CO, NH and CHCH<sub>3</sub>, then the interaction of each group with the solvent was calculated. Carbonyl groups account for about 70% of the total interaction energy in  $\alpha_R$  and 78% in PPII. NH groups contribute in about a 20%. This latter contribution is similar in both conformers, their contribution to the differential stability being practically negligible. So, the larger solute-solvent interaction energy in  $\alpha_R$  when compared to the PPII conformer is mainly due to the solvation of the two carbonyl groups. In order to analyze the origin of this behavior the interaction energy was recalculated using the same set of charges for the two conformers (note that in our model the charges are calculated quantum-mechanically and consequently they are configuration-dependent). For instance, if one uses the set of charges of PPII to re-calculate the water- $\alpha_R$  interaction energy, this decreases from -136.5 to -133.7 kcal/mol. Since the  $\Delta E_{\text{int}}$  value is 5.3 kcal/mol in water that means that 2.8 kcal/mol correspond to the use of different charge sets for the two conformers and 2.5 are related with the differences in the solvent structure around them.

Finally, Table 8 displays the interaction energies using gas phase charges for AD. This table permits the analysis of the influence that the solute polarization has on the charge distribution and conformational equilibrium. The increase in the interaction energy with the solvation is mainly related to the polarization of the carbonyl groups. In particular, the charge on the oxygen increases in about a 40% while that on the carbon atom increases in about 25%. The amide hydrogen atoms show an important charge increase as well (see Tables S1 and S2 in the supporting information). In general, the use of gas phase charges decreases significantly the interaction energy. As this decrease is more evident for  $\alpha_R$ , the difference between the two conformers,  $\Delta E_{\text{int}}(\alpha_R\text{-PPII})$ , decreases as well. For instance, in water the difference passes from 5.3 kcal/mol to 0.6 kcal/mol when in solution charges are changed by the in gas phase ones. By analyzing the group contribution it is found that while carbonyl groups favor the  $\alpha_R$  conformer N—H groups favor PPII. In sum, the AD polarization increases the interaction energy favoring the  $\alpha_R$  conformer.

## Summary

The comparison of the conformational equilibrium of AD in water, ethanol and a water-ethanol mixture permits to shed light on the ability of alcohols to modify the energy landscape of peptides and proteins. Our results confirm the experiments as they show that ethanol affects the energy landscape of alanine dipeptide favoring the formation of PPII conformer. In an attempt to understand the origin of this behavior the main contributions to the free energy value have been analyzed. Internal and solute-solvent interaction energies work in opposite directions, so, while the internal energy is lower in PPII than in  $\alpha_R$  and, consequently, it favors the PPII conformer, the AD-solvent interaction works in the opposite direction and favors the  $\alpha_R$  conformer. Remarkably, the gas phase internal energies are very

similar in the two conformers, consequently, the differences in the in solution values are associated with the polarization of the charge distribution during the solvation process, what is larger in  $\alpha_R$  than in PPII. We find that any factor decreasing the AD-solvent interaction energy decreases the relative stability of the  $\alpha_R$  conformer. This fact explains the effect of the basis set (*cc*-pVDZ versus *aug-cc*-pVDZ), solvent model (flexible or rigid) and dipeptide charge distribution (polarized and non-polarized). The use of diffuse functions, flexible solvent models and/or polarized charge distributions increases the  $\alpha_R$  population.

The analysis of the solvent structure around the AD molecule permits to conclude that, in the water-ethanol mixture, the first solvation shell is mainly constituted by water molecules; consequently, solute-solvent interaction energies are very similar in water and in the mixture. In spite of this, and in agreement with IR results, we found that there are direct interactions between ethanol and the dipeptide, the carbonyl group of the N-terminal end and the methyl side chains are the most probable areas of interaction.

Another remarkable fact is that the presence of preferential solvation in this system is not relevant for the differential stability of the two conformers. The reason is that the AD-solvent interaction energy value is the same in both conformers when all the molecules up to the second solvation shell are included. Consequently, the  $\alpha_R$ -PPII interaction energy difference is a bulk effect controlled by long-range interactions. However, the interaction energy difference when only the first solvation shell is considered is very large and it clearly favors the  $\alpha_R$  conformer. This fact points to a change in the conformational equilibrium of AD when solvated by small clusters.

### **Associated content**

### **SuPPorting Information**

Atomic charges for the alanine dipeptide in gas phase, water, water-ethanol mixture and ethanol solution. Note that in vacuum charges are obtained from the in solution optimized geometries

This material is available free of charge via the Internet at <http://pubs.acs.org>

### **Acknowledgement**

This work was suPPorted by the GR10048 project from the Consejería de Economía, Comercio e Innovación of the Gobierno de Extremadura. G.A. acknowledges a fellowship from program PDSE-CAPES, (n° 99999.014630 / 2013-03)

## References

- (1) Feig, M. *J. Comp Theor. Chem.* **2008**, *4*, 1555-1564.
- (2) Scarsdale, J. N.; Can Alsenoy, C.; Klimkowski, V. J.; Schäfer, L.; Momany, F. A. *J. Am. Chem. Soc.* **1983**, *105*, 3438.
- (3) Böhm, H.-J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129.
- (4) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 9255.
- (5) Jalkanen, K. J.; Suhai, S. *Chem. Phys.* **1996**, *208*, 81.
- (6) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908.
- (7) Philippi, D. M.; Friesner, R. A. *J. Comp. Chem.* **1999**, *20*, 1468.
- (8) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2239.
- (9) Anderson, A.; Hermans, J. *Proteins* **1988**, *3*, 262.
- (10) Pettitt, B. M.; Karplus, M. *Chem. Phys. Lett.* **1985**, *12Z*, 194.
- (11) Pettitt, B. M.; Karplus, M. *J. Phys. Chem.* **1988**, *92*, 3994.
- (12) Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 3213.
- (13) Grdadolnik, J.; Grdadolnik, S. G.; Avbelj, F. *J. Phys. Chem. B* **2008**, *112*, 2712.
- (14) Madison, V.; KoPple, K. D. *J. Am. Chem. Soc.* **1980**, *102*, 4855.
- (15) Mehta, M. A.; Fry, E. A.; Eddy, M. T.; Dedeo, M. T.; Anagnost, A. E.; Long, J. R. *J. Phys. Chem. B* **2004**, *108*, 2777.
- (16) Poon, C.-D.; Samulski, E. T.; Weise, C. F.; Weisshaar, J. C. *J. Am. Chem. Soc.* **2000**, *122*, 5642.
- (17) Takekiyo, T.; Imai, T.; Kato, M.; Taniguchi, Y. *Biopolymers* **2004**, *73*, 283.
- (18) Weise, C. F.; Weisshaar, J. C. *J. Phys. Chem. B* **2003**, *107*, 3265.
- (19) Seabra, G. M.; Walker, R. C.; Roitberg, A. E. in *Solvation effects on molecules and biomolecules*; Canuto, S., Ed.; Springer Science: Berlin **2008**; p 507.
- (20) Seabra, G. M.; Walker, R. C.; Roitberg, A. E. *J. Phys. Chem. A* **2009**, *113*, 11938.
- (21) Kwac, K.; Lee, K.-K.; Han, J. B.; Oh, K.-I.; Cho, M. *J. Chem. Phys.* **2008**, *128*, 105106.
- (22) Hu, H.; Eltsner, M.; Hermans, J. *Proteins* **2003**, *50*, 451.
- (23) Mu, Y.; Stock, G. *J. Phys. Chem. B* **2002**, *106*, 5294.
- (24) Mu, Y.; Kosov, D. S.; Stock, G. *J. Phys. Chem. B* **2003**, *107*, 5064.
- (25) Drozdov, A. N.; Grossfield, A.; PaPpu, R. V. *J. Am. Chem. Soc.* **2004**, *126*, 2574.
- (26) Smith, P. E. *J. Chem. Phys.* **1999**, *111*, 5568.
- (27) Jono, R.; Watanabe, Y.; Shimizu, K.; Terada, T. *J. Comp. Chem.* **2010**, *31*, 1168.
- (28) Hudáky, I.; Hudáky, P.; Perczel, A. *J. Comp. Chem.* **2004**, *25*, 1522.
- (29) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250.
- (30) Wang, A.-X.; Duan, Y. *J. Comp. Chem.* **2004**, *25*, 1699.
- (31) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439.
- (32) Kang, Y.K. *J. Phys. Chem. B* **2006**, *110*, 21338.
- (33) Kim, Y. S.; Wang, J.; Hochstrasser, R. M. *Phys. Chem. B* **2005**, *109*, 7511.
- (34) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179.
- (35) Tobias, D. J.; Brooks, C. L. *J. Chem. Phys.* **1992**, *96*, 3864.
- (36) Kalko, S. G.; Guardia, E.; Padró, J. A. *J. Phys. Chem. B* **1999**, *103*, 3935.
- (37) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
- (38) Toal, S.; Amidi, O.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.*, **2011**, *133*, 12728-12739.

- (39) Vagenende, V.; Yap, M. G. S.; Trout, B. L. *Biochemistry*, **2009**, *48*, 11084-11096.
- (40) Han, W.-G.; Jakanen, K. L.; Elstner, M.; Suhai, S. *J. Phys. Chem. B* **1998**, *102*, 2587.
- (41) Drozdov, A. N.; Grossfield, A.; PaPPu, R. V. *J. Am. Chem. Soc.*, **2004**, *55*, 493.
- (42) Sohn, W. Y.; Lee, J. S. *J. Phys. Chem A*, **2010**, *114*, 7537-7543.
- (43) Nerenberg, P.; Head-Gordon, T. *J. Chem. Theory Comput* **2011**, *7*, 1220.
- (44) Zagrovic, B.; Lipfert, J.; Sorin, E. J.; Millett, I. S.; van Gunsteren, W. F.; Doniach, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, *102*, 11698.
- (45) García-Prieto, F. F.; Fdez. Galván, I.; Aguilar, M. A.; Martín, M. E. *J. Chem. Phys.*, **2011**, *135*, 194502.
- (46) Cordeiro, J. M. M.; Soper, A. K. *J. Phys. Chem. B*, **2009**, *113*, 6819.
- (47) Almeida, G. G.; Cordeiro, J. M. M. *J. Braz. Chem. Soc.*, **2011**, *22*, 2178.
- (48) Cordeiro, J. M. M.; Cordeiro, M. A. M.; Bosso, A. R. S. A.; Politi, J.R.S., *Chem. Phys. Letters* **2006**, *423*, 67-70.
- (49) Sánchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. *J Comput Chem*, **1997**, *18*, 313.
- (50) Fdez. Galván, I.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *Comput. Phys. Commun.* **2003**, *155*, 244.
- (51) Martín, M. E.; Sanchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. *J of Mol. struct. (Theochem)* **2001**, *537*, 213-222.
- (52) Fdez. Galván, I.; Martín, M. E.; Aguilar, M. A. *J. Chem. Theory Comput.*, **2010**, *6*, 2445-2454.
- (53) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
- (54) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
- (55) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199.
- (56) Nakano, H.; Yamamoto, T. *J. Chem. Phys.*, **2012**, *136*, 134107.
- (57) Coutinho, K.; Georg, H. C.; Fonseca, T. L.; Ludwig, V.; Canuto, S. *Chem. Phys. Lett.* **2007**, *437*, 148-152.
- (58) Sánchez, M. L.; Martín, M. E.; Fdez. Galván, I.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Phys. Chem. B* **2002**, *106*, 4813.
- (59) Fdez. Galván, I.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2003**, *118*, 255.
- (60) Okuyama-Yoshida, N.; Nagaoka, M.; Yamabe, T. *Int. J. Quantum Chem.* **1998**, *70*, 95.
- (61) Okuyama-Yoshida, N.; Kataoka, K.; Nagaoka, M.; Yamabe, T. *J. Chem. Phys.* **2000**, *113*, 3519.
- (62) Hirao, H.; Nagae, Y.; Nagaoka, M. *Chem. Phys. Lett.* **2001**, *348*, 350.
- (63) a) Simons, J.; Jørgensen, P.; Taylor, H.; Ozment, J., *J. Phys. Chem.*, **1983**, *87*, 2745. b) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R., *J. Phys. Chem.*, **1985**, *89*, 82.
- (64) König, B. G.; Brooks, B. R. *Biochimica and Biophysic Acta*, **2015**, *1850*, 932-943.
- (65) Sánchez, M. L.; Aguilar, M. A.; Olivares del Valle, F. J. *J. of Molecular Structure (Theochem)*, **1998**, *426*, 181-190.
- (66) Gaussian 09, Revision A.1, 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.;

- Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; DaPPrich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.
- (67) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Comp. Phys. Comm.* **1995**, *91*, 43-56.
- (68) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.* **2001**, *7*, 306-317.
- (69) Refson, K. *Comput. Phys. Commun.* **2000**, *126*, 310.
- (70) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (71) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.
- (72) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 316.
- (73) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (74) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (75) Gekko, K.; Timasheff, S. N. *Biochemistry* **1981**, *20*, 4667.
- (76) Zhou, X.; Kaminski, J. W.; Wesolowski, T. A. *Phys. Chem. Chem. Phys.*, **2011**, *13*, 10565.

Table 1. Comparison of free energy (kcal/mol) and dipole moment (D) values for  $\alpha_R$  and PPII conformers using cc-pVDZ and aug-cc-pVDZ basis sets in water solution

	$\alpha_R \rightarrow$ PPII			
	$\mu$	$\Delta E_{QM}$	$A_{int}$	A
cc-pVDZ	3.8	1.5	.0	.5
aug-cc-pVDZ	4.0	2.4	.1	.7

Table 2. Free energy differences and their components for the transformation  $\alpha_R \rightarrow$  PPII in different solvents. All the energies are in kcal/mol. In parenthesis values excluding the  $\Delta A(\text{rot+vibr})$  contribution.

	Water	Water-Ethanol	Ethanol
$\Delta A_{int}$	3.1	3.1	3.4
$\Delta E_{QM}$	-2.4	-2.9	-3.3
$\Delta A(\text{rot+vibr})$	-0.08	+0.09	-0.42
$\Delta A$	0.6 (0.7)	0.2 (0.3)	-0.3 (0.1)
$\Delta A$ (rigid solvent)	0.2	-	-0.8

Table 3. Selected geometrical parameters of the  $\alpha_R$  conformer in several solvents. Distances in angstroms and angles in degrees.

	Water	Water-Ethanol	Ethanol
$C_5O_6$	1.26	1.26	1.25
$C_5N_7$	1.34	1.35	1.35
$N_7H_8$	1.04	1.04	1.03
$C_{11}O_{12}$	1.26	1.26	1.25
$C_{11}N_{17}$	1.33	1.33	1.34
$N_{17}H_{18}$	1.04	1.04	1.03
$\phi$	- 62.2	-61.3	-58.9
$\psi$	- 32.3	-32.1	-34.8

Table 4. Selected geometrical parameters of the PPII conformer in several solvents. Distances in angstroms and angles in degrees.

	Water	Water-Ethanol	Ethanol
	r		l
C <sub>5</sub> O <sub>6</sub>	1.26	1.26	1.25
C <sub>5</sub> N <sub>7</sub>	1.34	1.34	1.35
N <sub>7</sub> H <sub>8</sub>	1.04	1.04	1.03
C <sub>11</sub> O <sub>12</sub>	1.26	1.26	1.25
C <sub>11</sub> N <sub>17</sub>	1.33	1.33	1.34
N <sub>17</sub> H <sub>18</sub>	1.04	1.04	1.03
φ	-55.9	-54.2	-54.5
ψ	152.5	150.9	150.5

Table 5. Interaction energies, E<sub>int</sub>, (kcal/mol), internal energies (a.u.), distortion energies, E<sub>dist</sub> (kcal/mol) and dipole moments (D) in solution (μ) and in gas phase (μ<sub>0</sub>). \* Note that μ<sub>0</sub> refers to the dipole moment in gas phase of the in solution optimized geometries.

	Water		Water-Ethanol		Ethanol	
	α <sub>R</sub>	P <sub>PII</sub>	α <sub>R</sub>	PP <sub>I</sub>	α <sub>R</sub>	P <sub>PII</sub>
E <sub>int</sub>	-136.5	-131.2	-128.7	-122.0	-83.4	-76.0
E (a.u.)	494.52099	494.52037	494.521217	494.52082	494.52325	494.52373
E (a.u.)	494.46870	494.47246	494.47260	494.47727	494.49361	494.49889
E <sub>dist</sub>	3.2	3.0	3.0	2.7	3.1	3.1
μ	1.2	1.8	1.2	1.4	1.1	1.7
μ <sub>0</sub> *	1.7	1.5	1.7	1.5	1.7	1.4

Table 6. Coordination numbers around selected functional groups.

	Water		Water-Ethanol		Ethanol	
	α <sub>R</sub>	P <sub>PII</sub>	α <sub>R</sub>	PP <sub>I</sub>	α <sub>R</sub>	P <sub>PII</sub>
H <sub>w</sub> O <sub>6</sub>	2.70	2.75	2.20	2.50		
O <sub>12</sub>	3	2	2.7	2.2		

H <sub>w</sub>		.00	.80	5	5		
H <sub>8</sub>	1	1		0.7	1.0		
O <sub>w</sub>		.00	.00	5	0		
H <sub>18</sub>	1	1		0.6	0.9		
O <sub>w</sub>		.00	.00	0	0		
O <sub>6</sub>	-	-		0.2	0.0	1	1
H <sub>a</sub>				5	0	.75	.80
O <sub>12</sub>	-	-		0.2	0.5	2	1
H <sub>a</sub>				5	0	.00	.80
H <sub>8</sub>	-	-		0.2	0.0	1	1
O <sub>a</sub>				5	0	.00	.00
H <sub>18</sub>	-	-		0.4	0.1	1	1
O <sub>a</sub>				0	0	.00	.00

Table 7. Group contribution to the classical interaction energies (kcal/mol) for  $\alpha_R$  and PPII conformers.

	Water		Water-Ethanol		Ethanol	
	$\alpha_R$	PPII	$\alpha_R$	PPII	$\alpha_R$	PPII
C	-	-	-	-	0	-
H <sub>3</sub>	0.5	2.3	0.2	2.0	.4	0.8
C <sub>5</sub>	-	-	-	-	-	-
O <sub>6</sub>	54.2	47.2	50.1	43.4	34.0	26.6
N <sub>7</sub>	-	-	-	-	-	-
H <sub>8</sub>	12.1	15.2	12.7	17.2	8.2	11.1
C	-	-	0.	-	0	-
HCH <sub>3</sub>	0.1	2.8	4	1.9	.1	2.4
C <sub>1</sub>	-	-	-	-	-	-
O <sub>12</sub>	53.6	47.8	50.5	42.1	30.2	24.3
N <sub>1</sub>	-	-	-	-	-	-
H <sub>18</sub>	14.9	15.4	15.5	15.2	11.7	11.1
C	-	-	-	-	0	0
H <sub>3</sub>	1.1	0.5	0.04	0.2	.1	.3
T	-	-	-	-	-	-
total	136.5	131.2	128.7	122.0	83.4	76.0

Table 8. Group contribution to the classical interaction energies (kcal/mol) using gas phase charge distribution of the in solution optimized geometries.

		Water		Water-Ethanol		Ethanol	
		$\alpha$	P	$\alpha_R$	PP	$\alpha$	P
		R	PII		II	R	PII
<sub>3</sub>	CH	0	-	0.	-	1	-
		.5	0.9	6	0.9	.0	0.3
O <sub>6</sub>	C <sub>5</sub>	-	-	-	-	-	-
		37.9	31.1	35.4	28.9	26.5	20.3
H <sub>8</sub>	N <sub>7</sub>	-	-	-	-	-	-
		6.0	11.7	6.3	12.7	3.3	8.0
	CH	1	-	1.	-	0	-
CH <sub>3</sub>		.3	1.6	5	1.3	.6	1.7
O <sub>12</sub>	C <sub>11</sub>	-	-	-	-	-	-
		34.6	29.5	33.1	26.7	22.4	18.2
<sub>7</sub> H <sub>18</sub>	N <sub>1</sub>	-	-	-	-	-	-
		9.6	11.1	10.2	11.2	7.6	7.6
<sub>3</sub>	CH	-	-	0.	0.	0	0
		0.3	0.1	4	2	.4	.4
tal	To	-	-	-	-	-	-
		86.6	85.9	82.6	81.6	57.7	55.7

## Figure Captions

Figure 1: Atom numbering and representative dihedral angles of the AD molecule

Figure 2: Local molar fraction of water as a function of the distance (in Å) for the AD-water-ethanol system (full line). The dashed line corresponds to the value  $x_w=0.68$ .

Figure 3: Selected radial distribution functions for AD in different solvents. a)  $\alpha_R$  conformer in pure ethanol (full line) and in water-ethanol mixture (dashed line). b)  $\alpha_R$  conformer in pure water (full line) and in water-ethanol mixture (dashed line). c)  $\alpha_R$  (full line) and PPII (dashed line) conformers in water-ethanol mixture.

Figure 4: Three-dimensional probability distribution around  $\alpha_R$  dissolved in water-ethanol mixture. Color code: Magenta: water oxygen atoms. Orange: water hydrogen atoms. Red: ethanol oxygen atoms. Grey: ethanol hydrogen atoms

Figure 5: Three-dimensional probability distribution around PPII dissolved in water-ethanol mixture. Color code: Magenta: water oxygen atoms. Orange: water hydrogen atoms. Red: ethanol oxygen atoms. Grey: ethanol hydrogen atoms

Figure 6: AD-solvent interaction energy (in kcal/mol) as a function of the distance (in Å) for AD in water-ethanol mixture. Full line:  $\alpha_R$ . Dashed line: PPII.

Figure 1: Atom numbering and representative dihedral angles of the AD molecule.

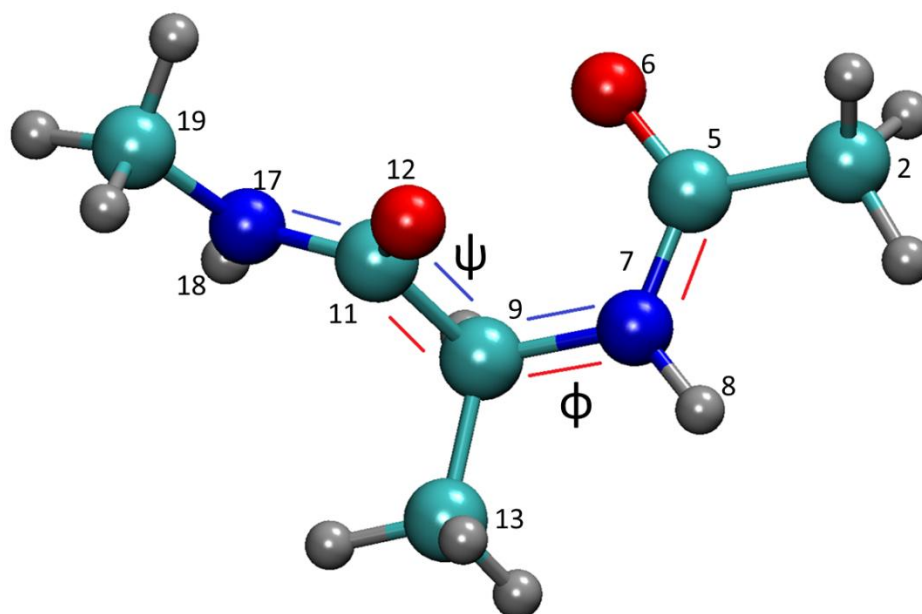


Figure 2: Local molar fraction of water as a function of the distance (in Å) for the AD-water-ethanol system (full line). The dashed line corresponds to the value  $x_w=0.68$ .

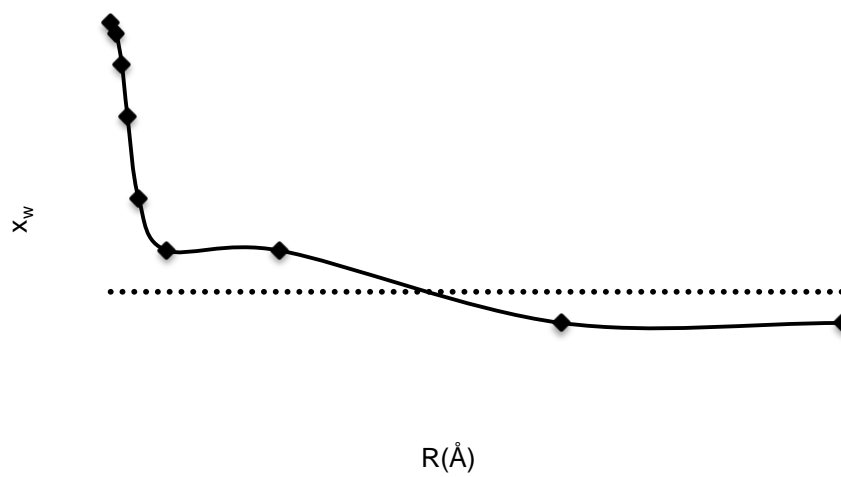
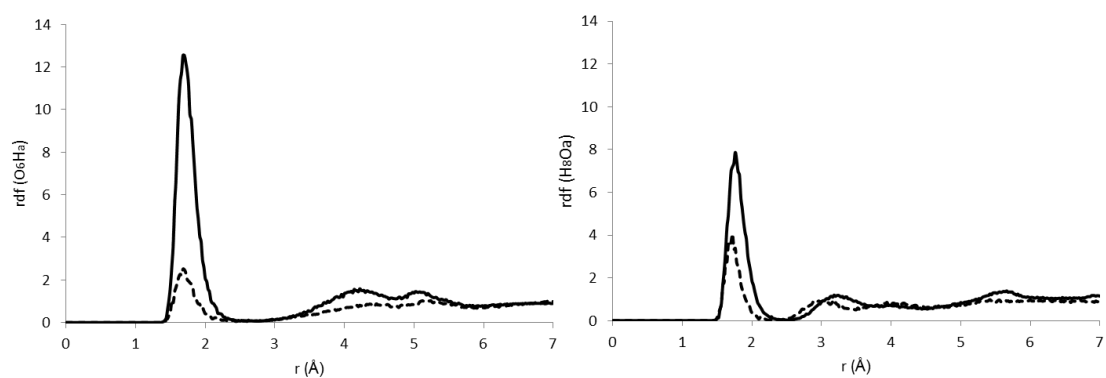
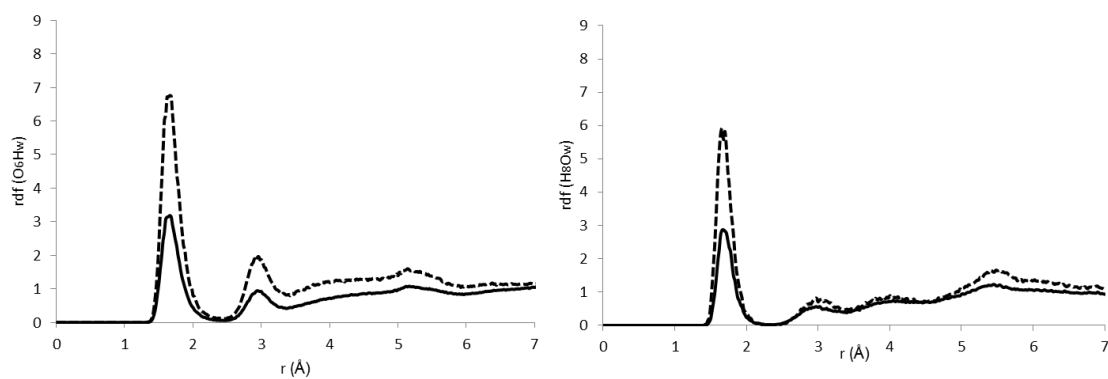


Figure 3: Selected radial distribution functions for AD in different solvents. a)  $\alpha_R$  conformer in pure ethanol (full line) and in water-ethanol mixture (dashed line). b)  $\alpha_R$  conformer in pure water (full line) and in water-ethanol mixture (dashed line). c)  $\alpha_R$  (full line) and PPII (dashed line) conformers in water-ethanol mixture.

a)



b)



c)

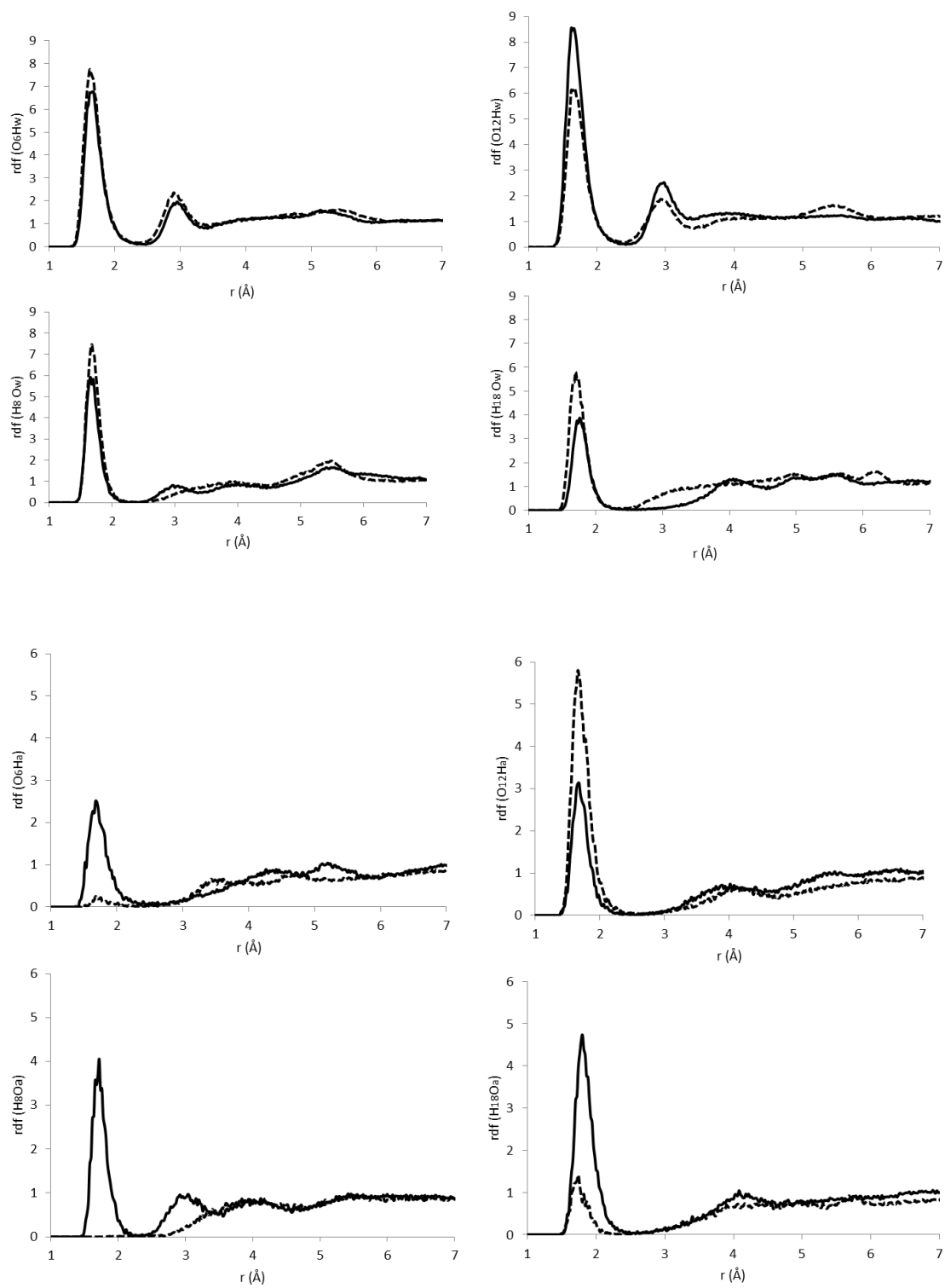


Figure 4: Three-dimensional probability distribution around  $\alpha_R$  dissolved in water-ethanol. Color code: Magenta: water oxygen atoms. Orange: water hydrogen atoms. Red: ethanol oxygen atoms. Grey: ethanol hydrogen atoms.

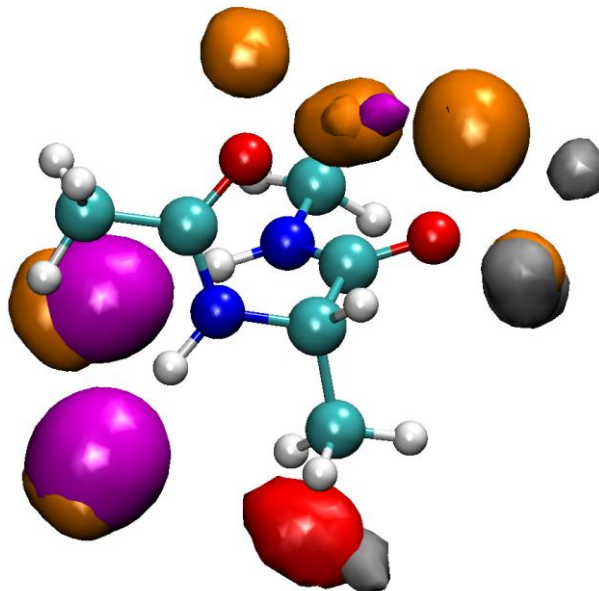


Figure 5: Three-dimensional probability distribution around PPII dissolved in water-ethanol. Color code: Magenta: water oxygen atoms. Orange: water hydrogen atoms. Red: ethanol oxygen atoms. Grey: ethanol hydrogen atoms.

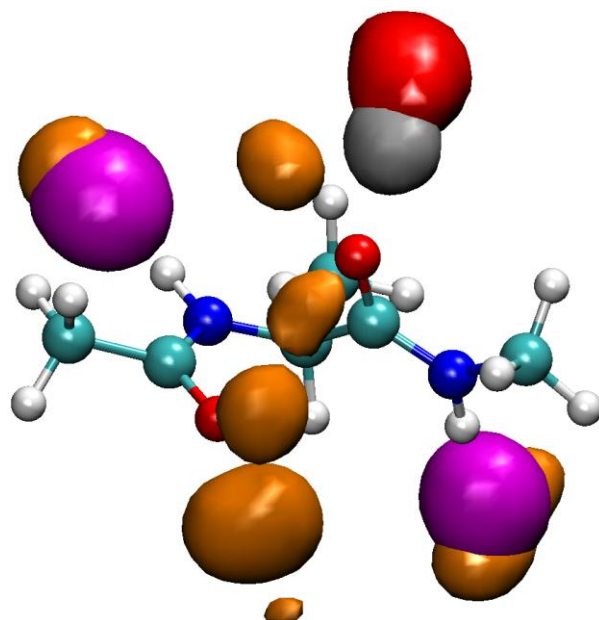


Figure 6: AD-solvent interaction energy (in kcal/mol) as a function of the distance (in Å) for AD in water-ethanol mixture. Full line:  $\alpha_R$ . Dashed line: PPII.

