

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO  
CENTRO DE INFORMÁTICA

TRABALHO DE GRADUAÇÃO

DESENVOLVIMENTO DO PROGRAMA AGOA, PARA  
INCLUSÃO DOS EFEITOS DO SOLVENTE EM  
PROCEDIMENTOS DE MODELAGEM MOLECULAR  
ATRAVÉS DA GERAÇÃO DE AGLOMERADOS OU  
CLUSTERS DE HIDRATAÇÃO

Klaus Ribeiro Cavalcante

Orientadora: Kátia Silva Guimarães  
Co-orientador: Marcelo Zaldini Hernandes, DCFar – UFPE

Recife  
Março de 2005

*Quando tudo está perdido  
Sempre existe um caminho  
Quando tudo está perdido  
Sempre existe uma luz  
Mas não me diga isso*

*Hoje a tristeza não é passageira  
Hoje fiquei com febre a tarde inteira  
E quando chegar a noite  
Cada estrela parecerá uma lágrima*

*Queria ser como os outros  
E rir das desgraças da vida  
Ou fingir estar sempre bem  
Ver a leveza das coisas com humor*

*Mais não me diga isso!  
É só hoje e isso passa...  
Só me deixe aqui quieto  
Isso passa.*

*Amanhã é outro dia  
Não é?*

*Eu nem sei por que me sinto assim  
Vem de repente um anjo triste perto de mim  
E essa febre que não passa  
E meu sorriso sem graça  
Não me dê atenção  
Mais obrigado por pensar em mim.*

*Quando tudo está perdido  
Sempre existe uma luz  
Quando tudo está perdido  
Sempre existe um caminho  
Quando tudo está perdido  
Eu me sinto tão sozinho  
Quando tudo está perdido  
Não quero mais ser quem sou.*

*Mais não me diga isso!  
Não me dê atenção!  
E obrigado por pensar em mim...*

— RENATO RUSSO (A Via Láctea)

*À minha esposa Manoela  
Ao meu filho amado Vinícius*

## AGRADECIMENTOS

*Aos meus pais pela educação que recebi assim como todos os ensinamentos de vida passados.*

*À minha esposa Manoela pela compreensão, paciência e carinho, e ao meu filho Vinícius que aos onze meses de idade, traz intensa alegria em nosso lar.*

*Aos meus irmãos Peterson e Jansen por de uma forma ou outra estarem sempre presentes em minha vida.*

*Aos meus amigos de curso do Centro de Informática – UFPE pelas horas de companheirismo que passamos durante esta fase. Como são muitos e, portanto não caberiam nesta seção, um abraço a todos vocês!*

*Em especial a Diogo Pedrosa e Tarcísio Falcão, dois grandes amigos que me acompanharam durante o curso sempre estendendo a mão nas horas mais difíceis.*

*A Marcelo Zaldini Hernandez, que demonstrou durante toda a iniciação científica não ser apenas meu professor orientador, mas sim um verdadeiro amigo me incentivando e apoiando em todas as atividades em que participei. Também gostaria de agradecer por todas as lições de vida e conselhos que tive a oportunidade de receber. Entre estes, não posso deixar de citar o famoso conselho: "Juízo, viu?", não é mesmo Marcelo?*

*A tantos outros que não foram citados nesta página, muito obrigado a todos!*

*Finalmente ao nosso Senhor Jesus Cristo por Sua imensa misericórdia e amor incondicional.*

## RESUMO

No campo das previsões teóricas e da modelagem molecular, os efeitos provocados pela presença do solvente na descrição dos modelos vêm sendo estudados em diversas áreas, desde a termodinâmica e a cinética de processos e reações químicas, até em espectroscopia de absorção e emissão, na ciência dos materiais, em estudos de relação quantitativa entre estrutura e atividade (QSAR) e na compreensão das mudanças estruturais causadas pelo meio condensado. O principal foco deste projeto diz respeito à implementação de um programa para a automação da metodologia AGOA, que representaria uma ferramenta computacional de grande utilidade na abordagem metodológica que permite a investigação dos efeitos do solvente (água) nos sistemas moleculares que apresentam potencial biológico ou farmacológico, por exemplo.

**Palavras-chave:** AGOA, Software, Hidratação, Modelagem Molecular

## **ABSTRACT**

In the field of theoretical predictions and molecular modelling, the effect caused by solvent presence in model descriptions are being studied in several areas, since thermodynamics and kinetics of processes and chemical reactions, until absorption and emission spectroscopy, material science, in quantitative structure activity relationship (QSAR) and about the comprehension of structural changes caused by condensed environment. The main concern of this project is the development of a software to turn the AGOA methodology automated, that would represent a computational tool very useful in investigation of solvent effects (water) in molecular systems that contain, for instance, biological or pharmacological potential.

**Keywords:** AGOA, Software, Hydration Clusters, Molecular Modelling

# SUMÁRIO

<b>CAPÍTULO 1 – INTRODUÇÃO .....</b>	<b>1</b>
1.1 MOTIVAÇÃO .....	1
1.2 OBJETIVOS .....	4
1.3 VISÃO GERAL .....	5
<b>CAPÍTULO 2 – A METODOLOGIA AGOA .....</b>	<b>6</b>
2.1 O ARQUIVO CUBE.....	8
2.2 O MODELO DE REPRESENTAÇÃO DA ÁGUA.....	9
2.3 O ALGORITMO DO PROGRAMA .....	9
<b>CAPÍTULO 3 – DESENVOLVIMENTO DA APLICAÇÃO .....</b>	<b>12</b>
3.1 ÁREA DE ACESSIBILIDADE AO SOLVENTE.....	14
3.2 RAIOS DE CORTE .....	15
3.3 GERAÇÃO DAS MOLÉCULAS DE ÁGUA.....	17
3.4 ARQUIVOS DE SAÍDA .....	19
<b>CAPÍTULO 4 – EXEMPLO DE USO: GERANDO ESTRUTURAS DE HIDRATAÇÃO .....</b>	<b>21</b>
4.1 PARÂMETROS UTILIZADOS NA MALHA TRIDIMENSIONAL .....	21
4.2 GERANDO AS ESTRUTURAS DE HIDRATAÇÃO COM O PROGRAMA AGOA.....	22
4.3 RESULTADOS OBTIDOS.....	24
<b>CAPÍTULO 5 – CONCLUSÕES E TRABALHOS FUTUROS .....</b>	<b>26</b>
5.1 PERSPECTIVAS.....	27
<b>APÊNDICE A – FORMATO DOS ARQUIVOS.....</b>	<b>28</b>
A.1 O FORMATO XYZ.....	28
A.2 O FORMATO PDB .....	29
A.3 O FORMATO CUBE .....	30
A.4 O FORMATO INP .....	32
<b>APÊNDICE B – GUIA DO USUÁRIO .....</b>	<b>33</b>
B.1 VISÃO GERAL .....	34
B.2 MENUS E OPÇÕES .....	35
<i>B.2.1 Menu File .....</i>	<i>35</i>
<i>B.2.2 Menu Actions.....</i>	<i>35</i>
<i>B.2.3 Menu Options .....</i>	<i>36</i>
<i>B.2.4 Menu View.....</i>	<i>39</i>

B.3 ARQUIVOS DE SAÍDA .....	39
<i>B.3.1 O arquivo AllPoints</i> .....	40
<i>B.3.2 O arquivo SelectedPoints</i> .....	40
<i>B.3.3 O arquivo SelectedPointsMultiple</i> .....	41
<i>B.3.4 O arquivo WaterMolecules</i> .....	42
<i>B.3.5 O arquivo WaterMoleculesMultiple</i> .....	42

## LISTA DE FIGURAS

Figura 1.1 – Aglomerado (cluster) de hidratação para uma molécula de interesse biológico.....	3
Figura 2.1 – Momento de dipolo ( $\mu$ ) da água e vetor campo elétrico (E) com orientação paralela e antiparalela, respectivamente.....	7
Figura 2.2 – Ângulo de rotação para o posicionamento dos átomos de hidrogênio..	7
Figura 2.3 – Malha tridimensional do MEP calculada com o programa Gaussian para uma molécula de interesse biológico.....	8
Figura 2.4 – Modelo TIP4P para a molécula de água.....	9
Figura 2.5 – Orientação paralela do momento dipolar da água com o campo elétrico estimado para (a) ponto de mínimo e (b) ponto de máximo... 10	
Figura 3.1 – Diagrama de casos de uso.....	12
Figura 3.2 – Diagrama de classes.....	14
Figura 3.3 – Área superficial de acessibilidade ao solvente (SASA).....	15
Figura 3.4 – Estruturas de hidratação obtidas com o AGOA (a) sem raio de corte para o solvente e (b) com raio de corte para o solvente.....	16
Figura 3.5 – Esquema de posicionamento dos átomos de hidrogênio.....	18
Figura 4.1 – Janela principal do programa AGOA com a molécula de metanol escolhida.....	22
Figura 4.2 – Janela de diálogo do cálculo do SASA.....	23
Figura 4.3 – Janela de diálogo Parameters contendo opções de raio de corte.....	23
Figura 4.4 – Janela de diálogo Preferences contendo opções da saída dos arquivos.....	24
Figura 4.5 – Janela principal após geração das estruturas de hidratação e geração dos arquivos de saída.....	24
Figura 4.6 – Visualização do arquivo <i>SelectedPointsMultiple</i> .....	25
Figura B.1 – Janela principal do programa AGOA.....	33
Figura B.2 – O menu <i>File</i> do programa AGOA.....	35
Figura B.3 – O menu <i>Actions</i> do programa AGOA.....	36
Figura B.4 – O menu <i>Options</i> do programa AGOA.....	37
Figura B.5 – A janela de diálogo <i>Parameters</i> do menu <i>Options</i> .....	37
Figura B.6 – A aba <i>Output</i> pertencente à janela de diálogo <i>Preferences</i> do menu <i>Options</i> .....	38
Figura B.7 – A aba <i>View</i> pertencente à janela de diálogo <i>Preferences</i> do menu <i>Options</i> .....	38
Figura B.8 – O menu <i>View</i> do programa AGOA.....	39

## LISTA DE TABELAS

Tabela 3.1 – Descrição dos casos de uso.....	13
Tabela 3.2 – Raios de corte padrão para os tipos de átomos do soluto.....	16
Tabela 3.3 – Descrição dos arquivos de saída.....	19
Tabela A.1 – Cabeçalhos do formato PDB.....	29



# CAPÍTULO 1

## INTRODUÇÃO

O recente avanço da computação tem trazido grande contribuição em diversas áreas do meio científico contemporâneo. A capacidade de realizar simulações, cálculos, e outras atividades onde o homem não o faria de forma semelhante ou em forma alguma, tornou o computador uma ferramenta indispensável em qualquer laboratório de pesquisa.

Softwares produzidos para estes propósitos muitas vezes, por serem específicos demais, acabam sendo difíceis de serem obtidos seja por razões financeiras, ou mesmo por questões de propriedade intelectual. A produção de software próprio, na maioria destes casos, acaba sendo uma alternativa viável e possui vantagens diretas como possibilidade de modificação do código-fonte para atualização ou inclusão de novas funcionalidades além de ausência de custos com suporte e treinamento.

Quando o software ainda não existe para uma dada aplicação, a necessidade pelo seu desenvolvimento torna-se ainda mais justificável, uma vez que não é possível obtê-lo de outra forma. Este trabalho também parte da necessidade de produção de software próprio dado que a metodologia teórica envolvida propõe o desenvolvimento de uma ferramenta computacional para sua utilização.

### 1.1 MOTIVAÇÃO

A química e bioquímica ocorre em sua grande parte em meio condensado, particularmente em soluções aquosas. Para a realização de simulações adequadas destes processos, os efeitos do solvente devem ser considerados.

Como foi ressaltado por Cramer & Thrular [CT95], estes efeitos originados pela presença do solvente têm se mostrado de fundamental

importância para vários processos químicos e biológicos e vêm recebendo considerável atenção.

No campo das previsões teóricas e da modelagem molecular, os efeitos provocados pelo solvente na descrição dos modelos vêm sendo estudados em diversas áreas, desde a termodinâmica e a cinética de processos e reações químicas [MJ86, CSJ85], até em espectroscopia [CC97] de absorção e emissão, na ciência dos materiais, em estudos de relação quantitativa entre estrutura e atividade (QSAR) e na compreensão das mudanças estruturais causadas pelo meio condensado [Rei90].

O modelo contínuo de solvatação descreve o solvente como um meio dielétrico sem estrutura, onde o soluto é inserido dentro de uma cavidade. Várias abordagens para estes métodos [TCM99, TG75] encontram-se atualmente implementados em muitos programas de química quântica. Apesar disto, independente de implementações, os modelos contínuos não são capazes de descrever interações específicas entre soluto e solvente, em particular, ligações de hidrogênio.

O modelo discreto de solvatação, por sua vez, trata o solvente como moléculas individuais, que interagem com o soluto via métodos clássicos [All87] ou quânticos. Este modelo soluciona, pelo menos parcialmente, os problemas com o modelo contínuo, uma vez que descreve apropriadamente as interações específicas entre soluto e solvente.

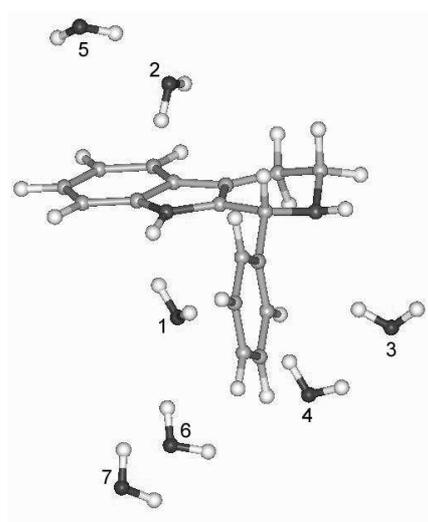
Além disto, a compreensão e representação correta destas interações, como por exemplo as ligações de hidrogênio, que ocorrem com frequência em moléculas de interesse biológico, continua sendo um desafio, principalmente pela falta de dados obtidos experimentalmente para diversos sistemas.

Os modelos discretos, contudo, apresentam uma demanda computacional muito maior que os modelos contínuos, e são geralmente dependentes da posição inicial das moléculas de solvente ao redor do soluto. A obtenção deste posicionamento pode ser feita pela utilização de simulações de mecânica estatística [All87], que também apresenta alta demanda computacional, além de requerer o conhecimento prévio dos potenciais de interação soluto-solvente e solvente-solvente que em geral exigem muito trabalho para serem obtidos.

Outra alternativa consiste no posicionamento aleatório das moléculas do solvente [FLS92, LNB01] em torno do soluto seguido de um procedimento de otimização de geometria para a obtenção de estruturas correspondentes aos mínimos de energia na superfície de potencial. Este procedimento também exige alta demanda computacional e tem elevada dependência em relação ao ponto de partida utilizado para as moléculas do solvente.

Na busca por novas metodologias que não necessitem nem do uso de técnicas estatísticas de amostragem (Mecânica Estatística ou Simulação Computacional) e nem do conhecimento prévio dos potenciais de interação intermolecular (campos de força), como no caso dos métodos de Monte Carlo e de Dinâmica Molecular, a proposta da metodologia para um modelo discreto de solvatação, denominada AGOA mostra-se uma alternativa viável e adequada para gerar estruturas ou aglomerados (clusters) de solvatação, particularmente de hidratação, como será visto ao longo deste texto.

Como o principal interesse está voltado para sistemas aquosos, a metodologia AGOA foi testada primeiramente para moléculas de água, como pode ser visto na figura 1.1, principalmente por ser uma molécula pequena, com um momento de dipolo elevado, que se reflete na sua elevada constante dielétrica. Além disso, também devido à sua reconhecida importância como solvente de praticamente todos os sistemas biológicos e de grande parte de química em solução.



**Figura 1.1 – Aglomerado (cluster) de hidratação para uma molécula de interesse biológico**

Em geral, nota-se que o aprimoramento de metodologias teóricas para o cálculo de descritores moleculares usados, por exemplo, em QSAR, através do emprego de métodos de química quântica e mecânica molecular, particularmente levando-se em conta alguns efeitos do solvente (AGOA), continua sendo um dos desafios da área de química teórica medicinal.

Além disto, a aplicação da metodologia AGOA em outras áreas da ciência, a exemplo da ciência dos materiais, dentre outras que utilizam Modelagem Molecular, pode ser útil para a compreensão dos fenômenos envolvidos com os efeitos do solvente ou do meio condensado nos mais diversos sistemas moleculares.

Portanto, o desenvolvimento de um software robusto, preciso e amigável para a realização de tais procedimentos de hidratação (geração de clusters) torna-se uma questão imprescindível para a continuação dos trabalhos de modelagem molecular que se preocupam com os efeitos do solvente nos sistemas moleculares, principalmente nos sistemas de interesse biológico.

## 1.2 OBJETIVOS

O principal objetivo deste trabalho diz respeito à implementação de um software para a automação da metodologia AGOA, que possibilite a investigação dos efeitos do solvente (água) nos sistemas moleculares que apresentam potencial biológico ou farmacológico, por exemplo.

O programa deverá contar ainda com um conjunto de opções e funcionalidades propostas para esta implementação, como por exemplo, a escolha automática do número de moléculas de água necessárias para gerar as estruturas dos aglomerados de hidratação, através do cálculo da área de acessibilidade ao solvente (SASA) [GT92]; e também a geração de arquivos de saída contendo todas as geometrias escritas em coordenadas cartesianas com opções para formato XYZ e PDB (Protein Data Bank) [Pro05], que são formatos bastante utilizados pela comunidade científica envolvida com Modelagem Molecular.

Mais detalhes sobre as funcionalidades do programa serão apresentados no Capítulo 3.

### 1.3 VISÃO GERAL

Neste Capítulo introdutório, foi apresentado ao leitor o contexto geral no qual este trabalho está inserido. Foi citada a metodologia teórica, denominada AGOA, pela qual propõe-se o desenvolvimento de um software que permita sua utilização prática, além da contribuição em outras áreas de interesse.

O Capítulo 2 detalhará a metodologia AGOA já introduzida neste Capítulo, incluindo uma breve descrição sobre o arquivo *cube* gerado pelo programa Gaussian [Gau05] que servirá como ponto de partida para a posterior utilização do programa. Neste Capítulo ainda constará detalhes sobre o modelo de representação da molécula de água e uma visão geral sobre o algoritmo principal do programa.

Questões relacionadas ao desenvolvimento da aplicação assim como as descrições de todas as funcionalidades propostas serão abordadas no Capítulo 3. Os conteúdos dos arquivos gerados pelo AGOA também serão brevemente discutidos.

O Capítulo 4 apresentará um exemplo típico de uso do software desenvolvido utilizando-se a molécula de metanol como soluto, que incluirá a geração das suas estruturas ou aglomerados de hidratação assim como a visualização dos arquivos de saída produzidos pelo programa.

No Capítulo 5 serão discutidos os resultados do trabalho, críticas, sugestões de modificação e trabalhos futuros.

## CAPÍTULO 2

### A METODOLOGIA AGOA

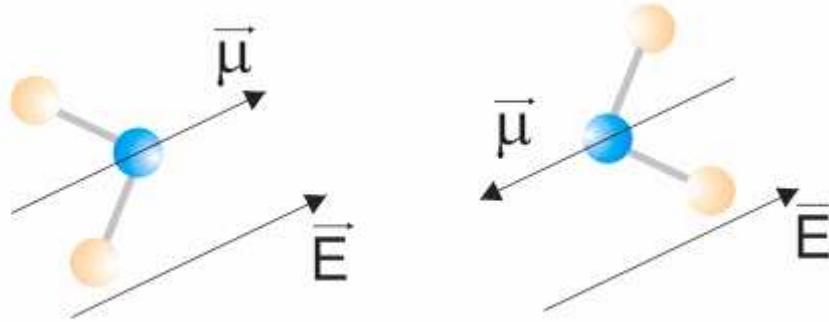
A metodologia AGOA descrita por Hernandez *et.al.* [HLS02] baseia-se na superfície de potencial eletrostático molecular (MEP), ou seja, no princípio de que as interações mais importantes entre o soluto e o solvente são de natureza eletrostática, portanto, as posições das moléculas de água em sua maioria são definidas pelo potencial eletrostático do soluto.

A distribuição eletrônica e nuclear de um sistema molecular seja ele uma molécula simples ou uma supermolécula, gera um potencial eletrostático nas vizinhanças.

Sabendo-se que este potencial decresce em módulo, à medida que se afasta da molécula, seus pontos de máximo e mínimo encontram-se invariavelmente em regiões próximas à molécula. Estes pontos são fortemente influenciados pela diferença de eletronegatividade entre os átomos da molécula, e portanto são importantes para explicar interações intermoleculares, como por exemplo, ligações de hidrogênio.

Portanto, estes pontos de máximo e mínimo serão utilizados para definir as posições das moléculas de água, ou seja, as coordenadas do átomo de oxigênio da água.

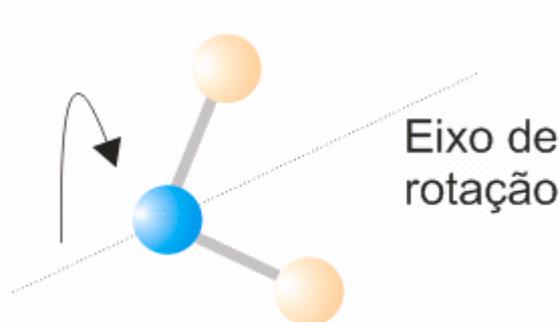
Para definir a orientação das moléculas de água, e uma vez que ela é uma molécula polar, ou seja, com momento de dipolo permanente, pode-se utilizar a energia de interação entre o dipolo elétrico da água e o campo eletrostático que será mínima quando o dipolo estiver paralelo ao campo e máxima quando forem antiparalelos. A Figura 2.1 ilustra estas duas situações.



**Figura 2.1 - Momento de dipolo ( $\mu$ ) da água e vetor campo elétrico ( $E$ ) com orientação paralela e antiparalela, respectivamente**

Uma vez que a orientação mais estável da molécula de água encontra-se quando seu momento de dipolo está paralelo ao campo elétrico, e este sendo expresso como o gradiente do potencial eletrostático, então, através do método de diferenças finitas, é possível estimar o campo elétrico a partir do potencial eletrostático calculado numa rede tridimensional de pontos, desde que esta malha seja suficientemente fina (menor distância entre os pontos que a compõem), ou seja, tenha resolução suficientemente alta para suportar esta aproximação.

Por fim, ainda é necessário escolher o posicionamento dos átomos de hidrogênio da água, através do ângulo definido pelo plano da molécula de água e o plano perpendicular a este, ou seja, aquele no qual é perpendicular ao momento de dipolo da água. Este ângulo é escolhido de forma aleatória, com o objetivo de simular as flutuações térmicas numa solução, como mostra a Figura 2.2.



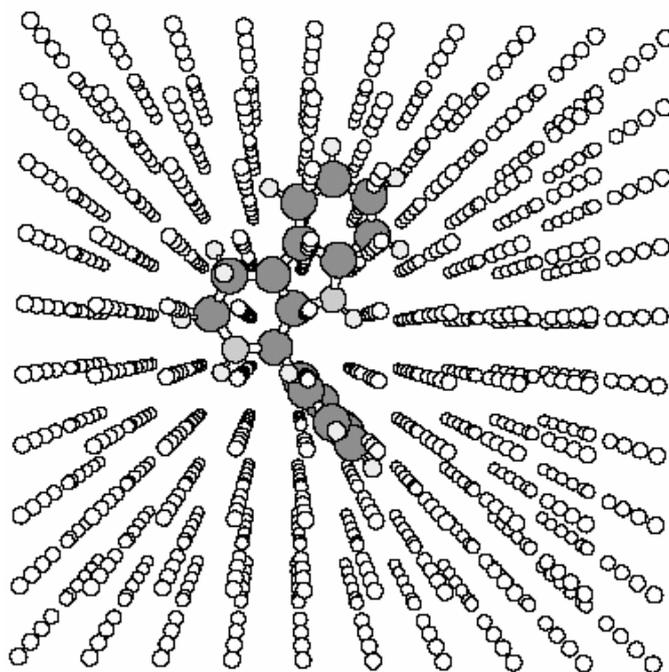
**Figura 2.2 - Ângulo de rotação para o posicionamento dos átomos de hidrogênio**

## 2.1 O ARQUIVO CUBE

Como descrito na seção anterior, o campo elétrico pode ser obtido de forma aproximada, utilizando-se para isso uma malha tridimensional contendo o valor de potencial eletrostático em cada ponto.

Para atender a este requisito, o programa AGOA utiliza um arquivo gerado pelo programa Gaussian [Gau05], denominado *cube* que contém tal malha, cujos valores de potencial foram calculados utilizando-se métodos de química quântica implementados neste referido programa.

A malha deve conter o soluto totalmente imerso em seu interior, e possuir os pontos espaçados de forma a reduzir ao máximo seu número, mas deve ainda manter esta resolução adequada para se obter uma boa aproximação do gradiente numérico do MEP, através de diferenças finitas.



**Figura 2.3 - Malha tridimensional do MEP calculada com o programa Gaussian para uma molécula de interesse biológico**

Vale ressaltar que na geração da grade contendo a molécula de interesse, a escolha das dimensões e da resolução da grade é de inteira

responsabilidade do usuário. Um exemplo de malha tridimensional pode ser visto na Figura 2.3.

Uma breve descrição sobre o formato do arquivo *cube* pode ser encontrado no Apêndice A.

## 2.2 O MODELO DE REPRESENTAÇÃO DA ÁGUA

Neste trabalho, a geometria do modelo TIP4P [JM85] foi utilizada como modelo de representação da água por ser largamente empregado em simulações computacionais.

Este modelo contém quatro sítios de interação, sendo eles os três átomos da molécula de água mais um sítio fantasma “*dummy*” representado pelo símbolo “XX” ou “M”, distanciado de 0,15 Å do átomo de oxigênio em direção aos átomos de hidrogênio como mostra a Figura 2.4.

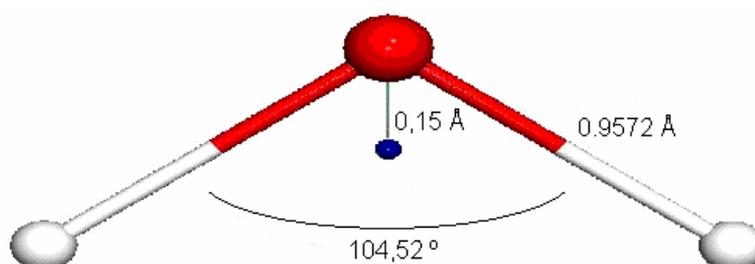


Figura 2.4 - Modelo TIP4P para a molécula de água

## 2.3 O ALGORITMO DO PROGRAMA

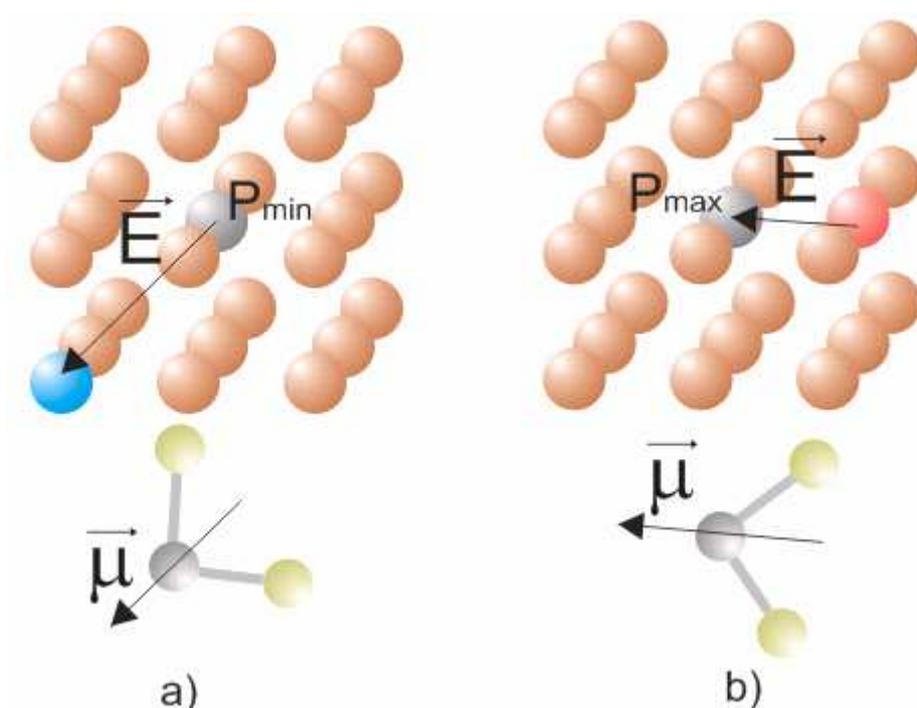
Nesta seção, serão apresentados de forma sucinta, os passos do algoritmo principal utilizado no programa.

Após a geração da malha tridimensional contendo os valores de potencial eletrostático calculados com o programa Gaussian para a molécula (soluto) de interesse, esta malha será lida, em forma de arquivo, pelo programa AGOA.

Serão excluídos desta grade, todos os pontos que estiverem inseridos no volume formado pela intersecção das esferas centradas nos núcleos e definidas pelos raios de corte dos tipos átomos do soluto.

O programa possui valores de raio de corte pré-definidos para os átomos de hidrogênio (H), carbono (C), nitrogênio (N), oxigênio (O), fósforo (P), enxofre (S), flúor (F), cloro (Cl) e bromo (Br), mas permite a inclusão dos raios de corte para toda a tabela periódica.

A seguir, o programa varre a malha tridimensional em busca dos pontos de máximo e mínimo em ordem decrescente em módulo. Nesta etapa, ao encontrar um ponto de máximo ou mínimo, os pontos diretamente vizinhos também são verificados e utilizados para se determinar o gradiente numérico do MEP, de tal forma que o momento de dipolo da água seja orientado paralelamente ao vetor do gradiente do MEP. Um exemplo deste processo pode ser visualizado através da Figura 2.5.



**Figura 2.5 - Orientação paralela do momento dipolar da água com o campo elétrico estimado para (a) ponto de mínimo e (b) ponto de máximo**

O posicionamento das moléculas de água é realizada nesta fase, onde são definidas as coordenadas dos sítios (O, M, XX) e acontece até que o número pré-definido de moléculas de solvente seja alcançado.

O programa é capaz ainda de utilizar raios de corte também para o solvente, a fim de se evitar estruturas ou aglomerados de hidratação muito correlacionadas, ou seja, com moléculas de água muito próximas entre si. Esta funcionalidade será vista com mais detalhes no próximo Capítulo.

Finalmente, o programa possibilita a geração de arquivos de saída para a visualização de propriedades eletrônicas como a superfície tridimensional de potencial eletrostático molecular e propriedades específicas das estruturas de hidratação previamente calculadas, utilizando-se para isso um programa de visualização.

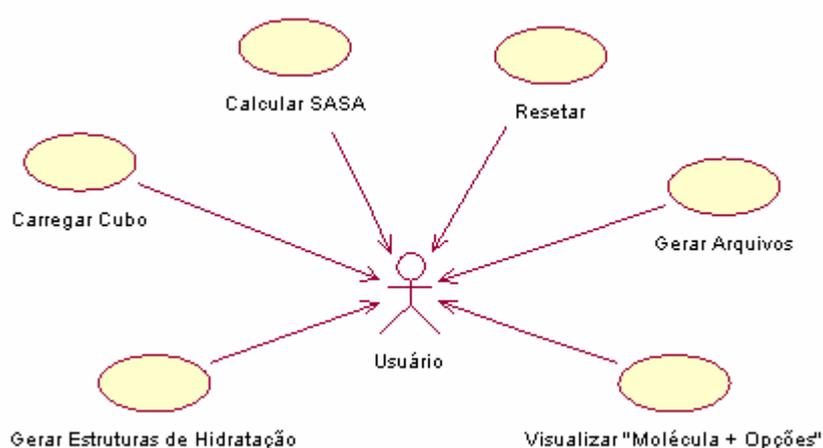
## CAPÍTULO 3

# DESENVOLVIMENTO DA APLICAÇÃO

Uma vez que a idéia principal do programa apresenta-se bem definida, seja através da descrição da metodologia AGOA, e pelo algoritmo mencionado no Capítulo anterior, foram realizados uma breve análise e projeto do software antes de sua implementação.

Esta etapa foi capaz de conduzir um modelo idealizado até o seu amadurecimento e, portanto, proporcionou a utilização de metodologias de engenharia de software que se mostraram essenciais para todo o processo de desenvolvimento.

Como o foco deste trabalho não se encontra na descrição deste processo, será mostrado a seguir, apenas dois diagramas UML [OMG03]: O diagrama de casos de uso, no qual mostra como o usuário pode interagir diretamente com o software e o diagrama de classes que apresenta uma visão geral das classes utilizadas na arquitetura do sistema, dado que este é um projeto orientado a objeto.



**Figura 3.1 - Diagrama de casos de uso**

A Tabela 3.1 descreve os casos de uso apresentados na Figura 3.1.

Caso de Uso	Descrição
<b>Carregar Cubo</b>	Carrega o arquivo <i>cube</i> que contém a molécula e a malha tridimensional necessária para a execução do programa.
<b>Calcular SASA</b>	Determina a Área Superficial de Acessibilidade ao Solvente (SASA) que é utilizada no cálculo do número de moléculas de água a serem geradas.
<b>Gerar Estruturas de Hidratação</b>	Define as coordenadas das moléculas de água e gera as estruturas de hidratação a partir da malha tridimensional previamente carregada.
<b>Gerar Arquivos</b>	Gera os arquivos de saída para posterior visualização de suas propriedades.
<b>Resetar</b>	Reinicializa o sistema descartando todos os dados existentes.
<b>Visualizar “Molécula + Opções”</b>	Visualiza, através de um programa externo, a molécula proveniente do arquivo <i>cube</i> . As opções existentes são: (i) os pontos de máximo e mínimo selecionados, juntamente com seus respectivos gradientes numéricos, (ii) os pontos da malha restantes após a exclusão daqueles sobrepostos à molécula e (iii) as estruturas de hidratação geradas pelo programa.

Tabela 3.1 – Descrição dos casos de uso

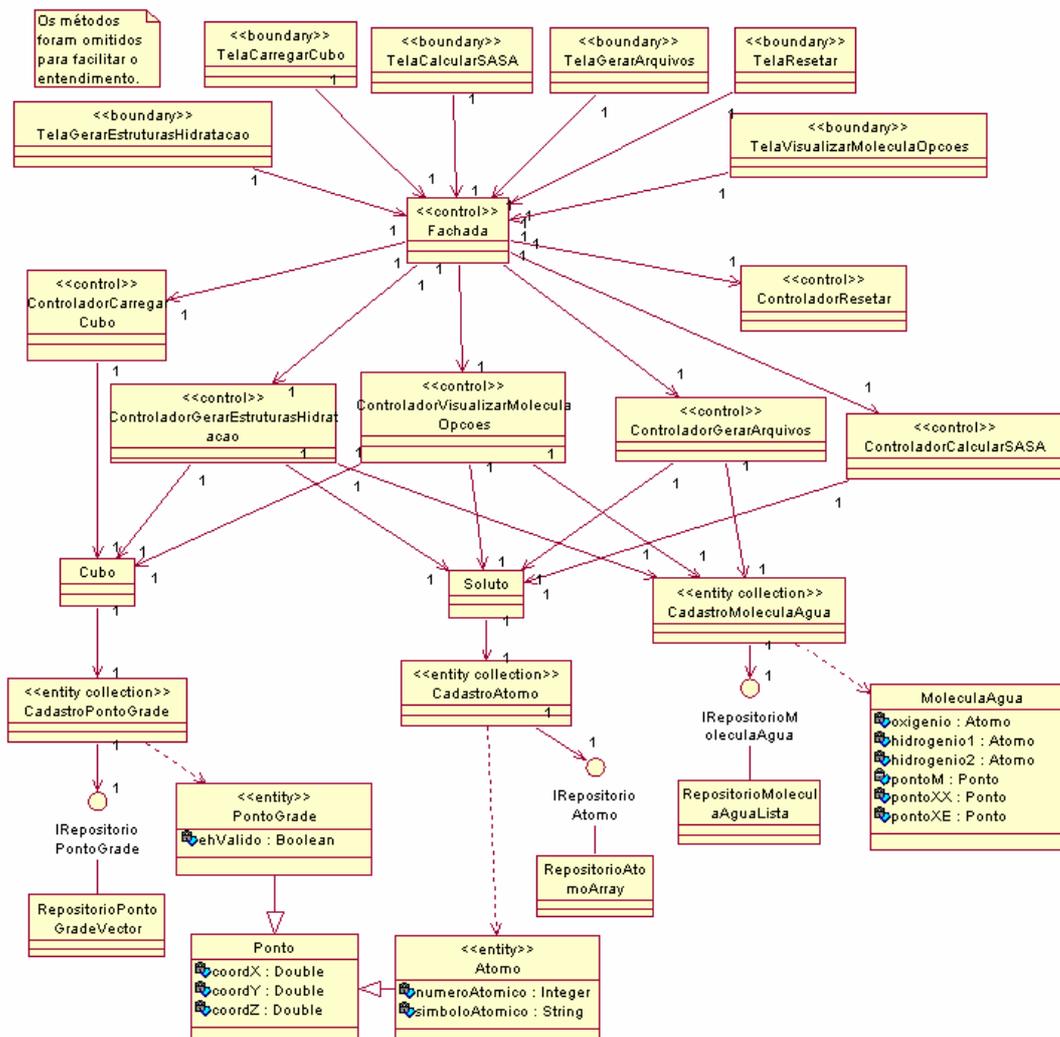


Figura 3.2 - Diagrama de classes

O programa AGOA foi implementado em linguagem de programação C++ por ser uma linguagem de alto desempenho, robusta e orientada a objeto. Foi utilizado o compilador *Microsoft Visual Studio* 6.0 [Mic05] com o auxílio da biblioteca *Qt* [Tro05] para o desenvolvimento da interface gráfica do programa.

Nas próximas seções serão apresentadas algumas funcionalidades e características que merecem um maior detalhamento.

### 3.1 ÁREA DE ACESSIBILIDADE AO SOLVENTE

Ao se utilizar o programa, a primeira questão com a qual o usuário se depara é a de determinar o número de moléculas de água para a geração das estruturas dos aglomerados de hidratação, principalmente pela dificuldade em certos casos em se estimar a área superficial da molécula em questão.

Para este propósito, o programa AGOA conta com o cálculo da área superficial de acessibilidade ao solvente (SASA) [GT92] que é utilizado na estimativa do número de moléculas de água que devem ser geradas ao redor do soluto. A SASA é definida como a área de uma superfície gerada pela passagem de uma esfera (sonda) com raio de 1,4 Angstrom, que percorre a superfície total do soluto, como pode ser visto na Figura 3.3.

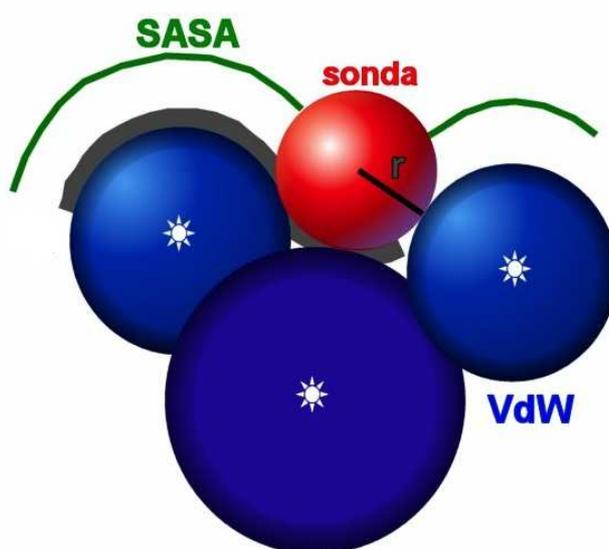


Figura 3.3 - Área superficial de acessibilidade ao solvente (SASA)

### 3.2 RAIOS DE CORTE

Como mencionado anteriormente, o programa AGOA utiliza raios de corte para cada átomo do soluto na exclusão dos pontos da malha tridimensional inscritos no volume definido pela intersecção das esferas formadas por estes átomos. Na Tabela 3.2, encontram-se os valores padrão dos raios de corte para alguns tipos de átomos mais comuns. O programa possibilita a inclusão de raios de corte para outros átomos, assim como a

mudança dos valores daqueles já existentes, através do arquivo *cutoff.inp* que acompanha o programa. Detalhes sobre o formato deste arquivo podem ser encontrados no Apêndice A.

Átomo	Raio de corte (em Angstroms)
H	1,3
C	2,0
N	2,0
O	2,0
F	1,8
P	2,0
S	2,0
Br	2,2
Cl	1,9

Tabela 3.2 – Raios de corte padrão para os tipos de átomos do soluto

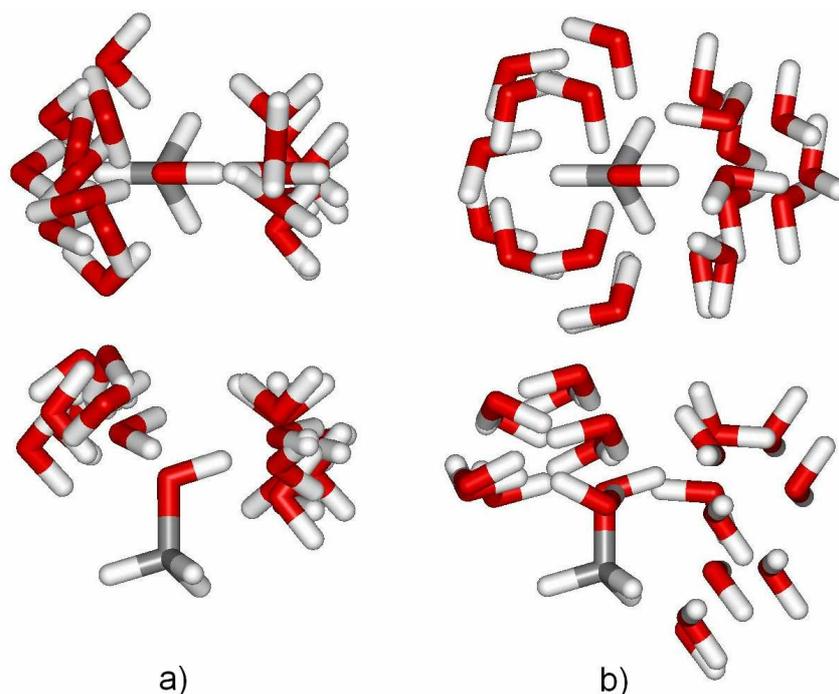


Figura 3.4 – Estruturas de hidratação obtidas com o AGOA (a) sem raio de corte para o solvente e (b) com raio de corte para o solvente

Para se evitar estruturas de hidratação altamente correlacionadas, foi proposta a inclusão de raios de corte para o solvente, como mostra a Figura 3.4. Neste caso, o posicionamento de uma nova molécula de água, durante o processo da geração das estruturas ou aglomerados de hidratação, só

poderá ser realizado em regiões externas a um raio de corte definido pelo usuário ou um valor padrão definido internamente no programa (1,4 Angstroms).

O valor de 1,4 Angstrom adotado como raio médio da molécula de água é utilizado como valor padrão, mas é possível a utilização de qualquer valor de raio no programa. Vale ressaltar que fica ao critério do usuário utilizar ou não este raio de corte, pois isto depende diretamente da finalidade da aplicação do programa.

### 3.3 GERAÇÃO DAS MOLÉCULAS DE ÁGUA

A geração da geometria da molécula de água necessita da seleção dos pontos de máximo ou mínimo e seus respectivos gradientes numéricos para o posicionamento inicial de seu átomo de oxigênio. Uma vez que a molécula de água encontra-se posicionada e orientada adequadamente, é necessário ainda definir a posição dos dois átomos de hidrogênio, além do sítio XX que a compõem. Esta geometria possui alguns aspectos importantes como, por exemplo, formar com o átomo de oxigênio (H-O-H) um ângulo de  $104,52^\circ$  e possuir distância entre cada átomo de hidrogênio e o de oxigênio de  $0,9572 \text{ \AA}$ . Outra característica fundamental é o ângulo definido pelo plano da molécula de água e o plano que passa perpendicular visualizado na Figura 2.2.

Para este trabalho, foi proposta a escolha aleatória para este ângulo, simulando-se desta forma as flutuações térmicas do meio condensado, como já comentado anteriormente. A solução encontrada para esta etapa envolveu grande esforço, dado que o problema mostrou ser não trivial.

A rotação destes átomos de hidrogênio em torno deste eixo cria uma superfície no espaço, mais precisamente um cone. Inicialmente a utilização de equações paramétricas aparentava ser uma solução viável, mas a complexidade crescia quando se pensava no cone fora dos eixos principais. Outra alternativa foi o uso de coordenadas esféricas e polares para a obtenção das posições dos átomos de hidrogênio referenciando cada situação possível entre o ponto proposto para o átomo de oxigênio e a

orientação da molécula. Ao todo, vinte e seis situações existentes (Figura 2.5) deveriam ser tratadas individualmente tendo estas dependência direta do ângulo aleatório escolhido. Novamente, o tratamento de todas as possibilidades se tornou complexo demais e, além disso, casos previamente não conhecidos, poderiam gerar inconsistências indesejadas nas geometrias das moléculas.

A solução final para este problema foi a utilização de operadores de rotação e translação (vide Figura 3.5) sobre os átomos de hidrogênio. A seguir, os passos utilizados na solução deste problema pode ser visualizado.

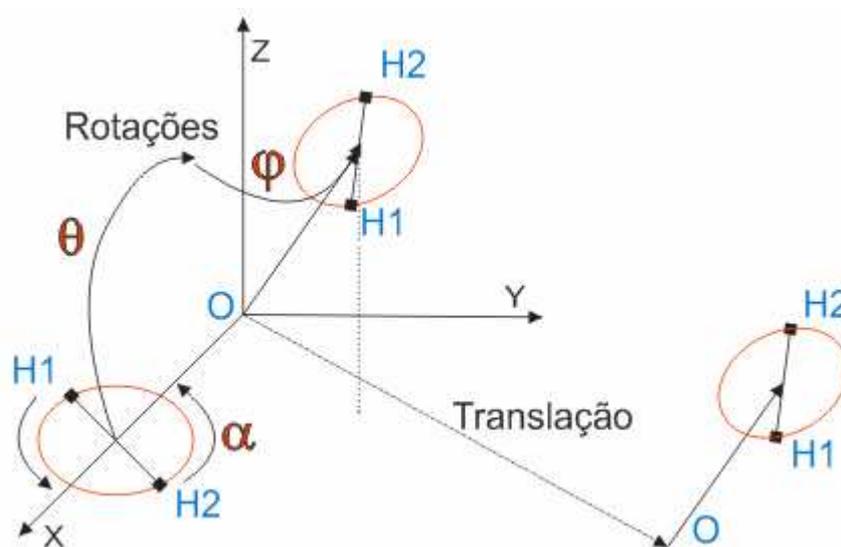


Figura 3.5 – Esquema de posicionamento dos átomos de hidrogênio

1. As coordenadas esféricas do vetor definido pelo ponto de máximo ou de mínimo que localiza o átomo de oxigênio e seu respectivo ponto do gradiente numérico são calculadas. Este vetor aponta para o centro da circunferência que contém os átomos de hidrogênio que podem deslizar sobre ela de acordo com o ângulo aleatório escolhido.
2. Uma circunferência equivalente é posicionada perpendicularmente sobre o eixo cartesiano X onde as coordenadas dos dois átomos de hidrogênio são determinadas com auxílio do ângulo aleatório de rotação.

3. Os átomos de hidrogênio são então rotacionados ao redor do eixo Y e depois ao redor do eixo Z, utilizando-se os valores dos ângulos obtidos das coordenadas esféricas do vetor citado anteriormente.
4. Finalmente os átomos de hidrogênio são transladados para a posição original do ponto de máximo ou mínimo (átomo de oxigênio).

### 3.4 ARQUIVOS DE SAÍDA

O programa AGOA possibilita a geração de arquivos de saída para a visualização das propriedades das estruturas de hidratação previamente calculadas. Cada um destes arquivos possui obrigatoriamente a molécula do soluto e as bordas da grade tridimensional. A Tabela 3.3 apresenta uma breve descrição sobre seus conteúdos.

Arquivo	Descrição
<b>AllPoints</b>	Contém todos os pontos da malha tridimensional de potencial eletrostático.
<b>SelectedPoints</b>	Contém os pontos de máximo e mínimo selecionados pelo programa juntamente com seus respectivos gradientes numéricos.
<b>SelectedPointsMultiple</b>	Possui o mesmo conteúdo do arquivo SelectedPoints, mas em formato múltiplo.
<b>WaterMolecules</b>	Contém as geometrias TIP4P das moléculas de água geradas pelo programa.
<b>WaterMoleculesMultiple</b>	Possui o mesmo conteúdo do arquivo WaterMolecules, mas em formato múltiplo.

Tabela 3.3 – Descrição dos arquivos de saída

O programa disponibiliza os formatos de arquivo XYZ [Xmo05] e PDB [Pro05] para a geração dos arquivos de saída. Estes formatos são bastante conhecidos no meio científico e, portanto, os arquivos nestes formatos são reconhecidos pela maioria dos programas de visualização tais como o *Rasmol* [Rog05] ou o *Chime* [Eri05], por exemplo. Maiores detalhes sobre estes formatos podem ser encontrados no Apêndice A.

## CAPÍTULO 4

### EXEMPLO DE USO: GERANDO ESTRUTURAS DE HIDRATAÇÃO

Neste Capítulo, será apresentado um exemplo de uso do programa AGOA que demonstrará a geração das estruturas de hidratação e dos posteriores arquivos de saída dos dados obtidos para a molécula de metanol (MeOH). Esta molécula, apesar de parecer pequena, representa um desafio robusto para a geração de estruturas de hidratação, pois apresenta dois sítios de ligação de hidrogênio bastante evidentes a saber: o átomo de oxigênio (como aceitador de Ligação de Hidrogênio) e o átomo de hidrogênio ligado à ele (como doador de Ligação de Hidrogênio).

#### 4.1 PARÂMETROS UTILIZADOS NA MALHA TRIDIMENSIONAL

Inicialmente, é necessário definir os parâmetros da malha tridimensional de pontos que irá conter o potencial eletrostático (MEP) para a molécula de metanol. Neste exemplo, as dimensões da caixa foram ajustadas para 10 Å x 10 Å x 10 Å. A resolução, ou seja, a distância entre pontos consecutivos da malha foi definido em 0,5 Å, e, portanto a grade contém 8000 pontos no total ( $20^3$ ). Vale a pena ressaltar que o aumento da resolução da grade, ou seja, a diminuição do valor citado anteriormente, tornará a malha mais precisa, mas em contrapartida, a quantidade de pontos da malha crescerá rapidamente, com uma dependência cúbica. O usuário deve levar em consideração para cada caso, estes dois aspectos no momento da escolha dos parâmetros da malha. Lembrando que um Angstrom (Å) é equivalente à  $10^{-10}$  metros (m).

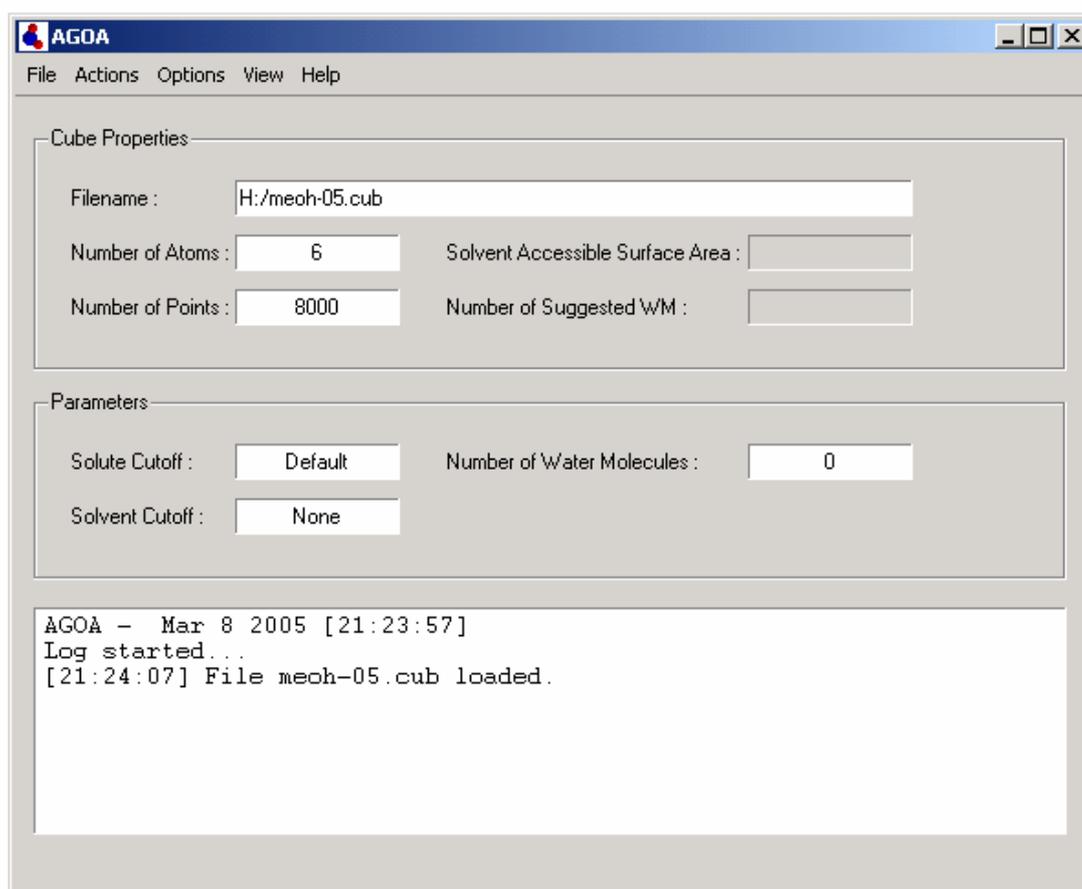
Em seguida, o programa Gaussian gera o arquivo *cube* contendo o potencial eletrostático em cada ponto da malha tridimensional. A descrição desta etapa não será comentada pois não está no foco deste trabalho.

O programa AGOA agora está apto para realizar a leitura deste

arquivo contendo os dados necessários para a geração dos *clusters* ou aglomerados de hidratação para a molécula de metanol.

## 4.2 GERANDO AS ESTRUTURAS DE HIDRATAÇÃO COM O PROGRAMA AGOA

Ao executar o programa AGOA, a janela principal do programa é apresentada ao usuário. O programa então deve ler o arquivo *cube* para carregar as informações referentes à molécula de metanol e sua malha de pontos. realizada esta etapa, esta janela deve estar semelhante à Figura 4.1



**Figura 4.1 - Janela principal do programa AGOA com a molécula de metanol escolhida.**

Para a definição do número de moléculas de água a serem geradas, o usuário pode modificar o valor diretamente em um dos menus do programa, ou fazer com que o programa estime aproximadamente o número de moléculas de água através do cálculo da área superficial de acessibilidade ao

solvente (SASA). A Figura 4.2 apresenta o diálogo de confirmação para o ajuste deste número.

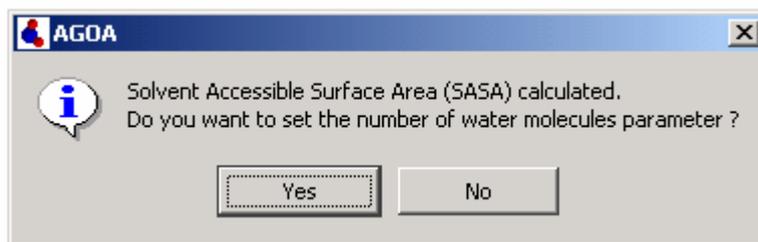


Figura 4.2 - Janela de diálogo do cálculo do SASA

O próximo passo é definir os raios de corte tanto para o soluto quanto para o solvente. No caso do soluto, foram utilizados os raios de corte *Default* (Padrão – ver Tabela 3.2). Para o raio de corte do solvente, foi escolhida a opção *Single Value* (Valor Único) ajustado para 1,4 Å, valor típico para o solvente escolhido (água). A Figura 4.3 ilustra esta etapa.

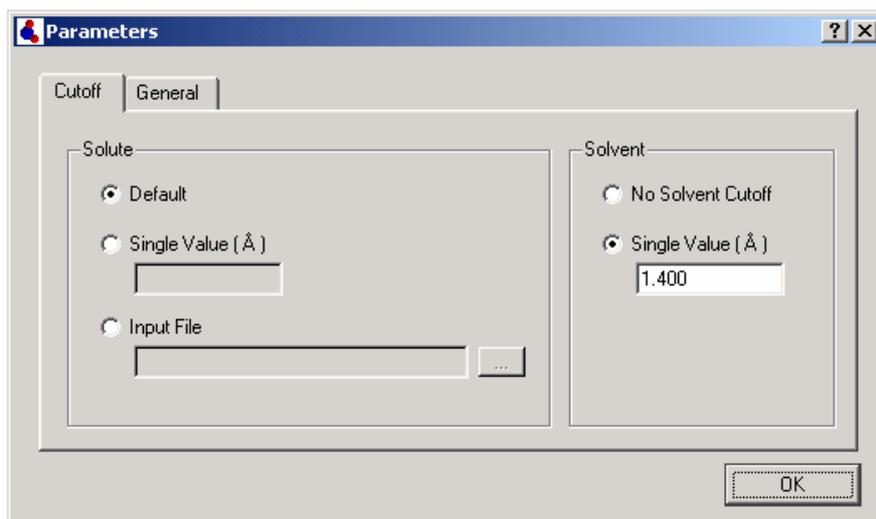


Figura 4.3 - Janela de diálogo Parameters contendo opções de raio de corte

Neste ponto, as estruturas de hidratação poderão ser calculadas através da ação da geração dos *clusters* do programa e posteriormente a saída dos arquivos que são selecionados pelo menu ilustrado pela Figura 4.4. Neste menu também será escolhido o local da geração dos arquivos, além do formato do arquivo.

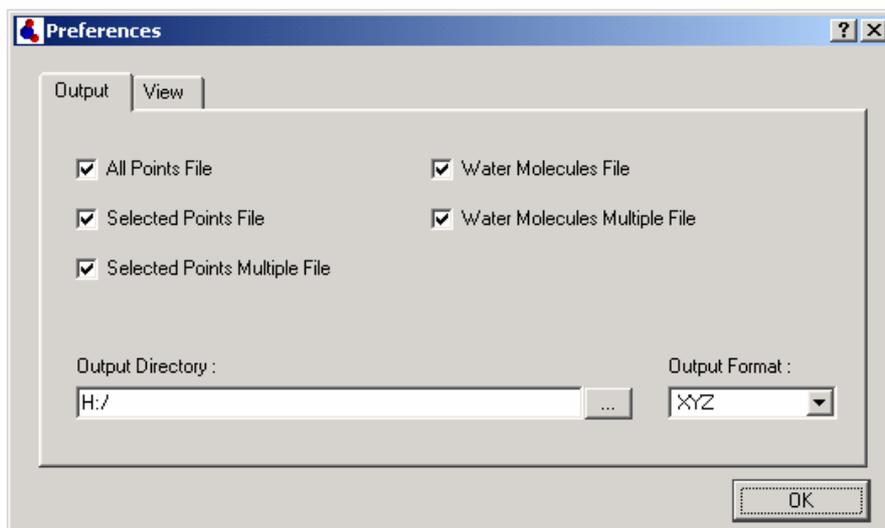


Figura 4.4 - Janela de diálogo *Preferences* contendo opções da saída dos arquivos

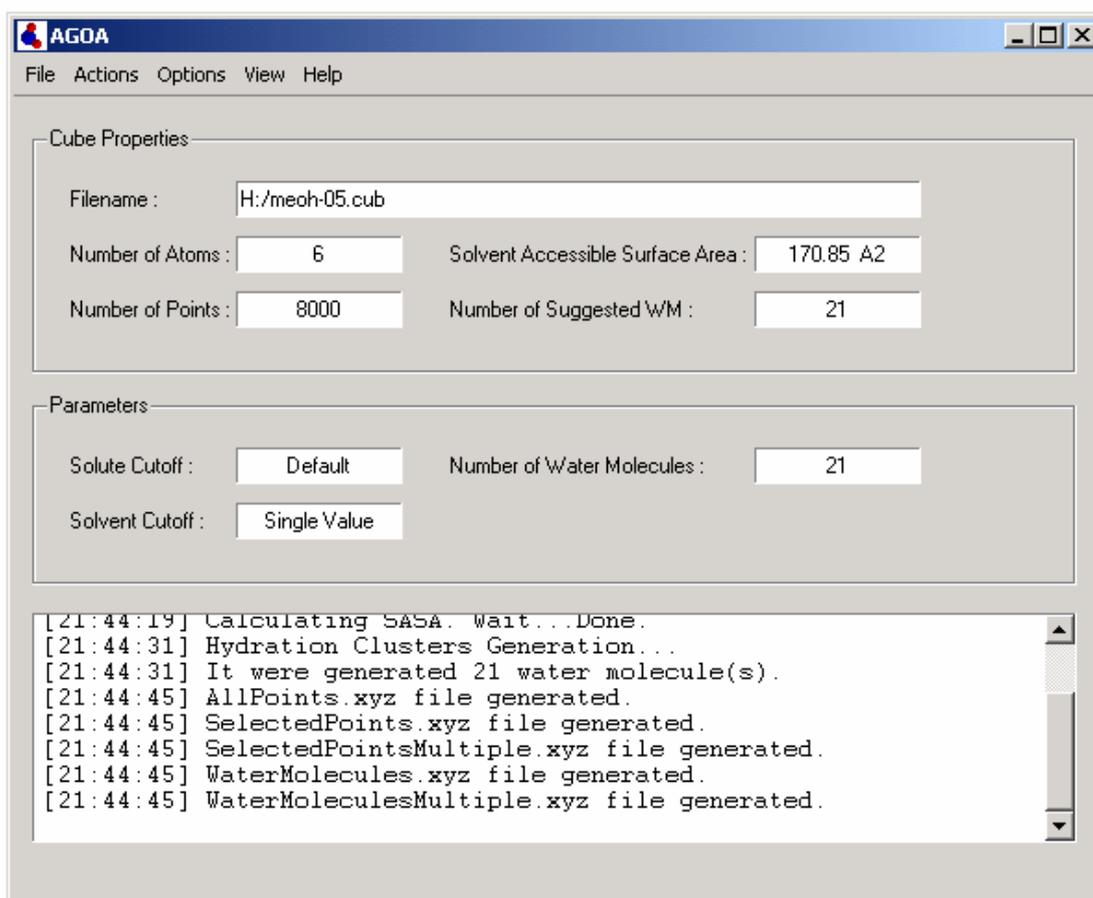


Figura 4.5 - Janela principal após geração das estruturas de hidratação e geração dos arquivos de saída

### 4.3 RESULTADOS OBTIDOS

Serão apresentadas a seguir, as figuras correspondentes à interpretação visual dos arquivos gerados pelo programa. A Figura 4.5 está

relacionada com o arquivo *SelectedPointsMultiple* que contém os pontos de máximo ou mínimo e seus respectivos pontos de gradiente numérico e a Figura 4.6 permite visualizar o arquivo *WaterMoleculesMultiple* que possui as estruturas de hidratação propriamente ditas.

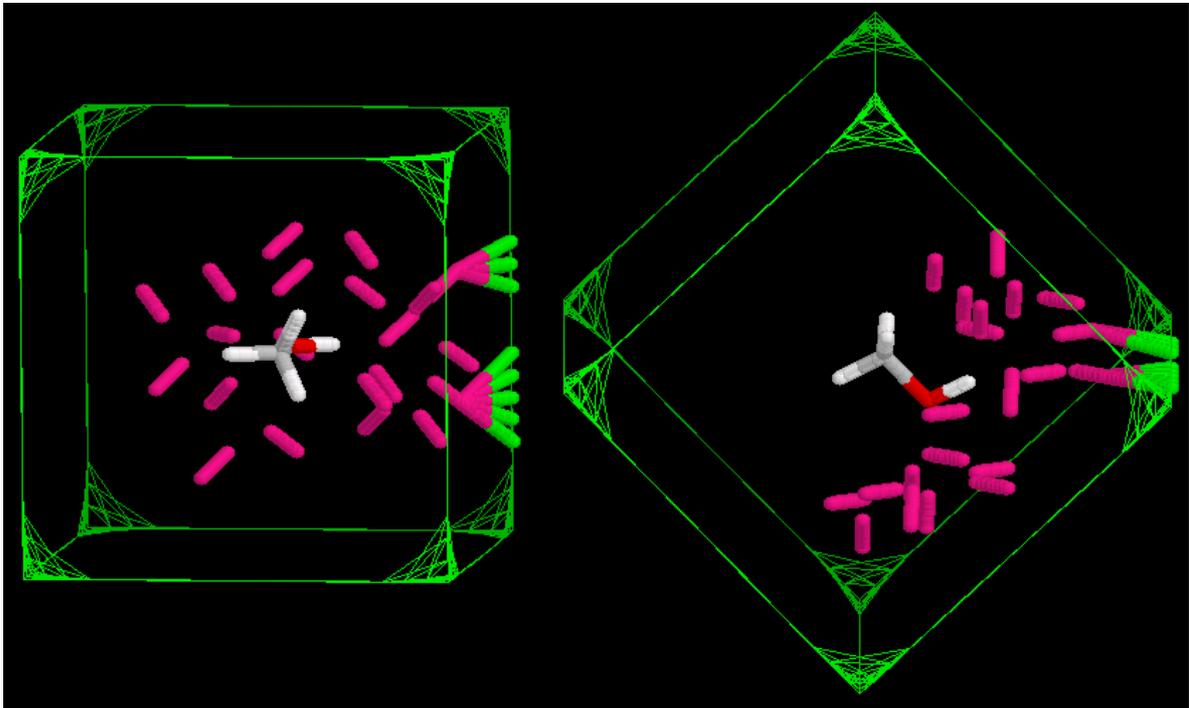


Figura 4.6 - Visualização do arquivo *SelectedPointsMultiple*

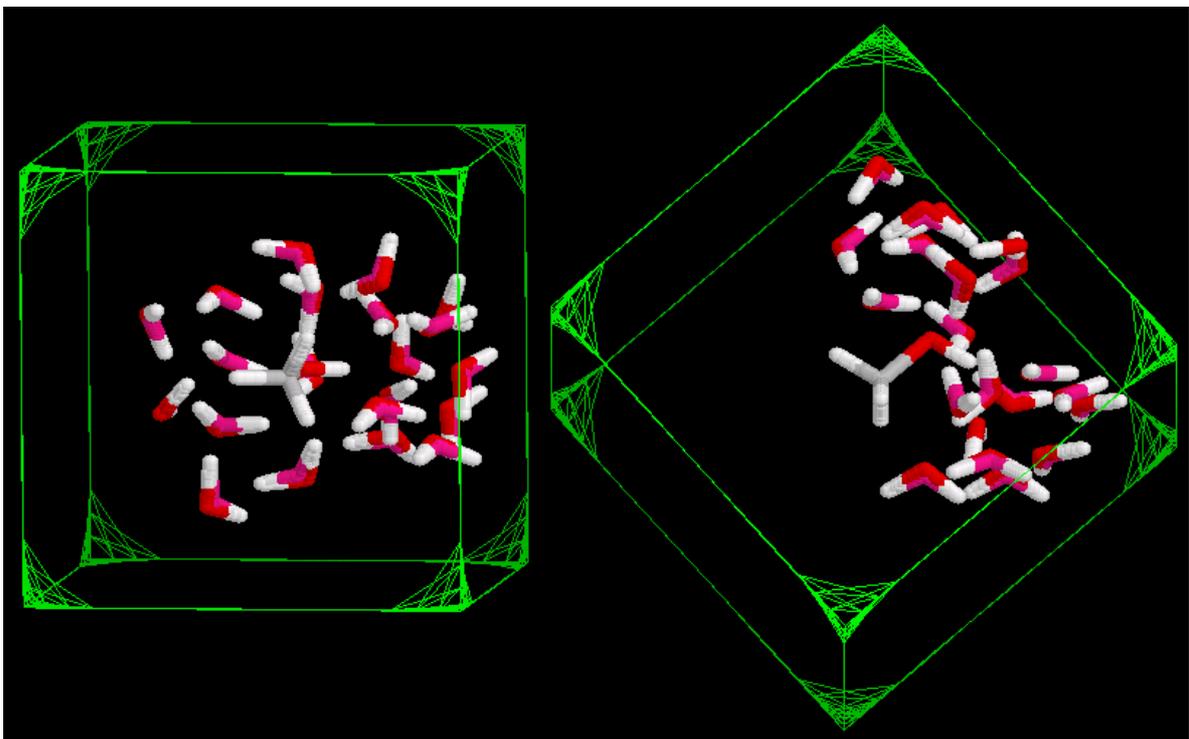


Figura 4.7 - Visualização do arquivo *WaterMoleculesMultiple*

## CAPÍTULO 5

# CONCLUSÕES E TRABALHOS FUTUROS

Ao final deste trabalho, foi possível empreender um projeto de desenvolvimento de software por completo, através de suas etapas de modelagem, projeto, implementação e testes. Esta oportunidade trouxe maturidade ao conhecimento adquirido durante o curso de graduação, uma vez que este conhecimento foi colocado em prática.

O envolvimento com uma nova área de pesquisa, neste caso, a modelagem molecular aplicada ao desenvolvimento racional de fármacos, representou um grande desafio, porém despertou profundo interesse pelo assunto e motivação para a criação de softwares para aplicações científicas.

Cabe ressaltar ainda que o acesso a este tipo de tecnologia geralmente é restrito em nosso país, pois os softwares usados nesta área do conhecimento são muitas vezes onerosos e provenientes, em sua maioria, de empresas estrangeiras. Desta forma, existe uma grande demanda para o desenvolvimento destas ferramentas a nível nacional e o grupo de pesquisa do qual faço parte tem procurado dar uma contribuição pioneira neste sentido.

O programa AGOA, desenvolvido neste trabalho, já se encontra em fase de utilização e teste pelos usuários colaboradores membros do grupo de pesquisa GPIT [GPIT05] (Grupo de Pesquisa em Inovação Terapêutica). Isto reafirma a importância deste trabalho como ferramenta computacional que provê suporte aos procedimentos de modelagem molecular realizados pelo grupo.

A principal dificuldade encontrada durante a realização deste projeto, foi a ausência de dados sobre estruturas de hidratação calculadas com outros métodos, para fins de validação e análise de desempenho do programa AGOA em relação a estes. Portanto, não foi possível realizar tal atividade considerada de grande importância, no decorrer deste trabalho.

Como sugestões de aperfeiçoamentos, podem ser citados alguns acréscimos e modificações, tais como:

1. Inclusão de um componente de visualização molecular, a fim de tornar o programa AGOA independente de quaisquer programas de visualização externos.
2. Inclusão de novos formatos de arquivo usados por diversos programas de modelagem molecular. (Hyperchem, Sybyl, Tripos, Wavefunction, entre outros.)
3. Generalização da leitura do Potencial Eletrostático Molecular (MEP ou SPE) provenientes de outros programas além do Gaussian, como por exemplo o MOPAC, que é inclusive gratuito para utilização acadêmica.

## 5.1 PERSPECTIVAS

Como o potencial eletrostático é uma propriedade localizada e aditiva [PM01, BST76, BST07], seria possível, *a priori*, estender esta aproximação (AGOA) para o tratamento de fragmentos de biomacromoléculas como proteínas e ácidos nucleicos e depois combiná-los para obter estruturas de aglomerados de hidratação adequadas para estes biopolímeros. Este é a idéia central por traz da proposta de generalização do programa AGOA para o BioAGOA.

Além disso, existe uma demanda crescente pelo desenvolvimento e implementação de outras ferramentas computacional para a realização de procedimentos de modelagem molecular utilizados para o planejamento racional de fármacos, como por exemplo na área de QSAR (“Quantitative Structure-Activity Relationship”), portanto existe a perspectiva de continuação destes trabalhos em conjunto com o grupo de pesquisa já citado.

## APÊNDICE A

# FORMATO DOS ARQUIVOS

### A.1 O FORMATO XYZ

Este formato foi originalmente utilizado em um programa de visualização denominado XMol [Xmo05]. Possui uma estrutura bem simples, pois utiliza basicamente as coordenadas cartesianas dos átomos da molécula ou de quaisquer estruturas contidas no arquivo.

A primeira linha contém o número total de entradas existentes, ou seja, apresenta a quantidade de linhas que deverão ser lidas em sequência. A segunda linha possui um texto explicativo sobre o conteúdo do arquivo ou mesmo qualquer tipo de identificação sobre a estrutura que segue. A terceira linha em diante contém as entradas que apresentam em geral a mesma forma básica: Um símbolo para a representação do ponto ou átomo, geralmente um ou dois caracteres (geralmente os símbolos atômicos), seguido de suas coordenadas cartesianas (em Angstroms) em três colunas.

Existe ainda a possibilidade de se adicionar uma quinta coluna contendo o valor de uma propriedade definida para cada ponto em particular. Vale lembrar que as colunas não possuem uma formatação em especial, ou seja, podem vir em qualquer posição, desde que sigam esta ordem e exista, no mínimo, um espaço em branco entre elas.

O formato fornece ainda a capacidade de multiplicidade, que é a possibilidade de sequenciar mais de uma estrutura num mesmo arquivo. Para isso, basta adicionar uma estrutura ao final de outra.

Abaixo pode ser visto um exemplo do formato XYZ, que neste caso, possui duas estruturas, uma com seis entradas e outra com cinco entradas. Na primeira estrutura, a quarta e quinta entradas possuem uma quinta coluna contendo um valor de uma propriedade (MEP), acontecendo o mesmo para a segunda estrutura em sua segunda e terceira entradas.

```

6
Configuration # 1.
C   0.000   0.000   0.000
O   0.000   0.000   1.424
H   0.925   0.000   1.714
XX  2.000  -0.000   2.500   0.052
XE  2.500  -0.500   3.000   0.020
H   0.484   0.892  -0.416
5
Configuration # 2.
C  -1.046   0.000  -0.305
XX  3.000  -1.000   2.500   0.016
XE  3.500  -1.500   3.000   0.008
O   0.000   0.000   1.424
H   0.925   0.000   1.714

```

## A.2 O FORMATO PDB

O formato PDB (Protein Data Bank) [Pro05] é bastante conhecido no meio acadêmico e tecnológico, principalmente no que diz respeito a estrutura de bio-macromoléculas (proteínas e ácidos nucléicos). Seu principal uso é o armazenamento de dados sobre estas moléculas apresentando tanto sua estrutura quanto informações referentes à fonte e procedimentos de obtenção.

CABEÇALHO	SIGNIFICADO
<b>REMARK</b>	Comentários gerais
<b>ATOM</b>	Coordenada atômica para grupos “padrão”
<b>HETATM</b>	Coordenada atômica para grupos “não-padrão”
<b>MODEL</b>	Especificação do número do modelo para estruturas múltiplas em uma única entrada de dados
<b>ENDMDL</b>	Marcação para o fim do modelo para estruturas múltiplas
<b>TER</b>	Finalizador de seqüência de dados

**Tabela A.1 – Cabeçalhos do formato PDB**

Cada linha do formato é iniciada por um cabeçalho que possui função específica dentro da estrutura do arquivo. O formato PDB possui inúmeros cabeçalhos e o objetivo desta seção não é esgotar o assunto, portanto, uma

breve descrição será dada apenas daqueles cabeçalhos utilizados neste trabalho. Na Tabela A.1 encontra-se o mapeamento entre cabeçalho e seu respectivo significado.

As linhas do formato PDB contêm exatamente 80 caracteres, e são preenchidas com caracteres em brancos, se necessário, até seu final. Os campos de determinados cabeçalhos possuem posicionamento pré-definido e extensão limitada. Para se aprofundar mais sobre o assunto, recomenda-se a leitura da descrição oficial do formato PDB [PDB05]. A seguir, um simples exemplo do formato.

```
REMARK  Universidade Federal de Pernambuco (UFPE)
MODEL 1
HETATM  1 C  MOL  0.000  0.000  0.000
HETATM  2 O  MOL  0.000  0.000  1.424
HETATM  3 H  MOL  0.926  0.000  1.714
HETATM  4 H  MOL  0.485 -0.892 -0.416
TER
ENDMDL
MODEL 2
HETATM  5 C  MOL  4.500 -1.500  4.500
HETATM  6 H  MOL -1.040  0.000 -0.300
HETATM  7 H  MOL  0.484  0.892 -0.410
HETATM  8 H  MOL  1.000  4.500  4.500
TER
ENDMDL
```

### A.3 O FORMATO CUBE

Este formato é utilizado pelo programa Gaussian [Gau05] na geração do arquivo *cube* que contém uma grade tridimensional de pontos devidamente espaçados entre si. Estes pontos possuem um valor de potencial eletrostático (MEP) calculado pelo programa para uma dada molécula.

As duas primeiras linhas são comentários, e geralmente são ignorados pelos aplicativos, mas em certos casos podem servir como identificadores do conteúdo do arquivo. A terceira linha possui o número de átomos incluídos no arquivo seguido da origem da grade tridimensional. As próximas três linhas contêm a quantidade de pontos em cada direção seguido do vetor axial. As próximas linhas contêm as informações sobre cada átomo consistindo de cinco números: o primeiro é o número atômico, o segundo geralmente é a

carga atômica (mas pode vir com 0.0 para todos os átomos), e os últimos três são as coordenadas cartesianas dos átomos.

Os dados referentes ao MEP da molécula são apresentados em seguida, na forma de números em ponto flutuante, separados por espaços em branco, sendo que cada número está associado a um ponto da grade tridimensional. Tradicionalmente a grade é lida com o eixo X como o laço exterior, e o eixo Z como o laço interior. Um exemplo de acesso a estes dados, em linguagem C++, pode ser visto a seguir.

```
for ( int x = 1; x <= numeroPontosDirecaoX; x++ ) {
    for ( int y = 1; y <= numeroPontosDirecaoY; y++ ) {
        for ( int z = 1; z <= numeroPontosDirecaoZ; z++ ) {
            cout << matriz[x][y][z] << endl;
            //imprime na tela
        }
    }
}
```

Logo abaixo, um exemplo deste formato:

```
meoh
MP2 Total Density
  6  -9.448630  -9.448630  -9.448630
 20  0.944863  0.000000  0.000000
 20  0.000000  0.944863  0.000000
 20  0.000000  0.000000  0.944863
  6  6.000000  0.000000  0.000000  0.000000
  8  8.000000  0.000000  0.000000  2.691663
  1  1.000000  1.749615  0.000000  3.239155
  1  1.000000  0.915727 -1.685718 -0.786748
  1  1.000000 -1.977223  0.000035 -0.577693
  1  1.000000  0.915423  1.685718 -0.786689
 3.70317E-05 -4.47852E-05 -1.47119E-04 -2.73023E-04 -4.25196E-04
-6.05499E-04
-8.14346E-04 -1.05003E-03 -1.30809E-03 -1.58101E-03 -1.85827E-03
-2.12710E-03
-2.37376E-03 -2.58532E-03 -2.75132E-03 -2.86519E-03 -2.92479E-03
-2.93230E-03
-2.89333E-03 -2.81575E-03
...
...
...
 1.05670E-04  2.09517E-05 -8.93394E-05 -2.30406E-04 -4.07523E-04
-6.25368E-04
-8.87003E-04 -1.19255E-03 -1.53773E-03 -1.91272E-03 -2.30172E-03
```

-2.68395E-03  
-3.03593E-03 -3.33501E-03 -3.56299E-03 -3.70900E-03 -3.77064E-03  
-3.75337E-03  
-3.66850E-03 -3.53055E-03

## A.4 O FORMATO INP

Este formato foi criado para conter os dados referentes aos raios de corte dos átomos do soluto do programa AGOA. Como mencionado no Capítulo 3, seu uso está na inclusão de raios de corte para novos átomos e modificação daqueles já existentes.

A primeira linha contém o número total de entradas existentes, ou seja, apresenta a quantidade de linhas que deverão ser lidas em sequência. A segunda linha possui um texto explicativo sobre o conteúdo do arquivo. A terceira linha em diante contém as entradas que apresentam em geral a mesma forma básica: O símbolo atômico do referido átomo, seguido do valor de seu raio de corte em Angstroms.

As colunas dos símbolos atômicos e valores de raio de corte não possuem formatação especial, ou seja, basta que exista entre elas, pelo menos, um espaço em branco. A seguir, um exemplo do formato poderá ser visto.

```
10
# External input for cutoff parameters:
H    1.30
C    2.00
N    2.00
O    2.00
F    1.80
P    2.00
S    2.00
Br   2.20
Cl   1.90
XX   2.00
```

## APÊNDICE B

# GUIA DO USUÁRIO

Neste Apêndice, será apresentado o Guia do Usuário que servirá como referência principal para a utilização do programa AGOA. Seu objetivo é descrever as funcionalidades do programa, incluindo detalhes e opções presentes na interface gráfica do mesmo.

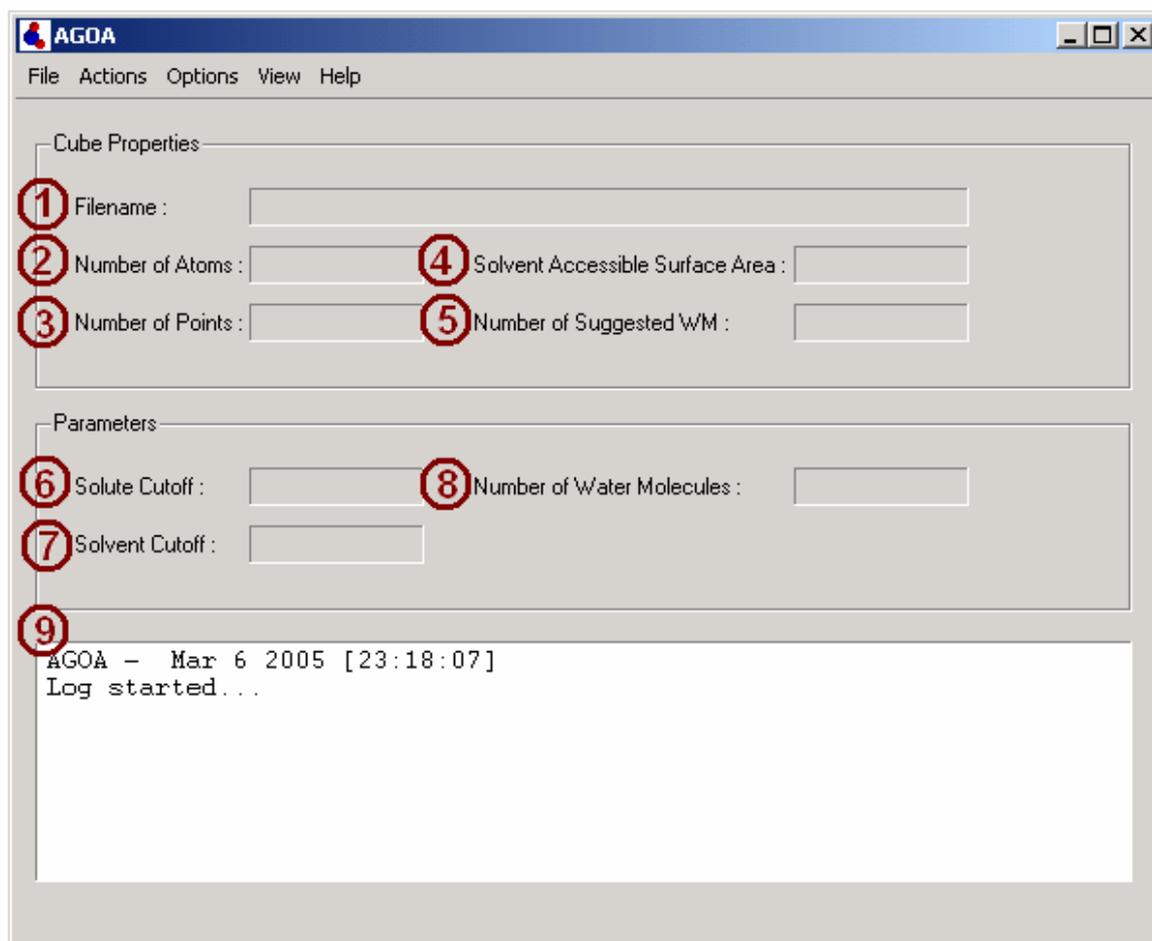


Figura B.5 - Janela principal do programa AGOA

Legenda da Figura B.1:

1. Este campo informa o nome e caminho do arquivo *cube* carregado

- pelo programa.
2. Este campo informa o número de átomos da molécula contida no arquivo.
  3. Este campo informa o número de pontos da grade tridimensional contida no arquivo *cube*.
  4. Este campo informa o valor da área superficial de acessibilidade ao solvente (SASA), sendo que a unidade utilizada é o Ångstron quadrado (Å<sup>2</sup>).
  5. Este campo informa o número de moléculas de água sugeridas pelo programa para a geração das estruturas de hidratação.
  6. Este campo informa o modo em que os valores de raio de corte para o soluto serão informados, são eles:
    - a. *Default* (Padrão): Os raios de corte dos átomos assumem o valor definido pela Tabela 3.2.
    - b. *Single Value* (Valor Único): Neste caso, os raios de corte de todos os átomos assumem o valor definido pelo usuário.
    - c. *Input File* (Arquivo de Entrada): Os valores dos raios de corte dos átomos são fornecidos através de um arquivo externo.
  7. Este campo informa a utilização ou não de um raio de corte para o solvente.
    - a. *No Solvent Cutoff* (Nenhum): O programa não utilizará um raio de corte para o solvente.
    - b. *Single Value* (Valor Único): O raio de corte terá o valor definido pelo usuário.
  8. Este campo informa o número de moléculas de água que serão geradas pelo programa.
  9. Este campo gera um arquivo de registro (*log*), ou seja, a seqüência das ações utilizadas no programa.

## B.1 VISÃO GERAL

Ao ser executado, o programa AGOA exibirá sua janela principal semelhante à Figura B.1. Cada componente da área de trabalho da janela

possui propósito informativo, ou seja, todas as funcionalidades do programa necessárias para a geração das estruturas de hidratação, visualização dos dados obtidos e saída dos mesmos em arquivo, encontram-se nos menus do programa. Estes menus serão vistos com mais detalhes na próxima seção.

## B.2 MENUS E OPÇÕES

Todas as ações por parte do usuário, como por exemplo, a leitura do arquivo *cube* e a geração dos arquivos de saída, partem dos menus do programa. A seguir estes menus serão brevemente discutidos.

### B.2.1 MENU *FILE*

O menu *File* visualizado pela Figura B.2 possui três ações descritas a seguir.

1. *Load Cube File*: Carrega o arquivo *cube* que contém a molécula e a malha tridimensional necessária para a execução do programa.
2. *Reset*: Reinicializa o sistema descartando todos os dados existentes.
3. *Exit*: Finaliza o programa.

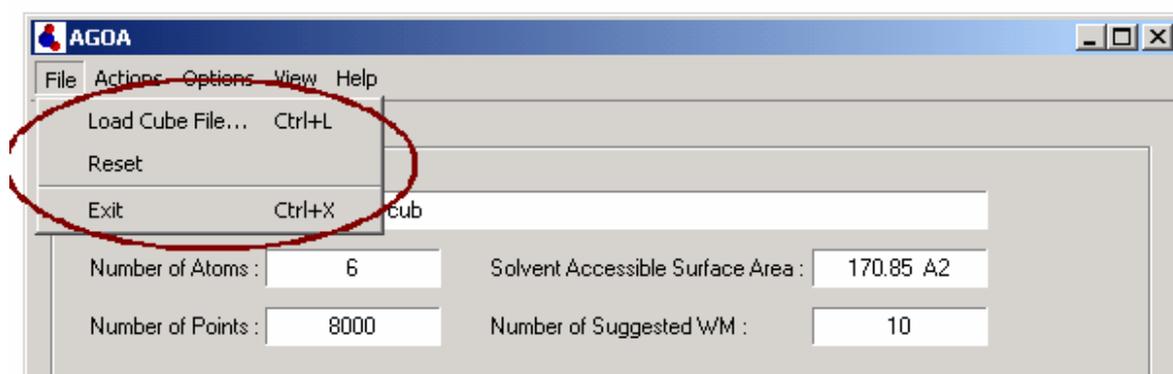
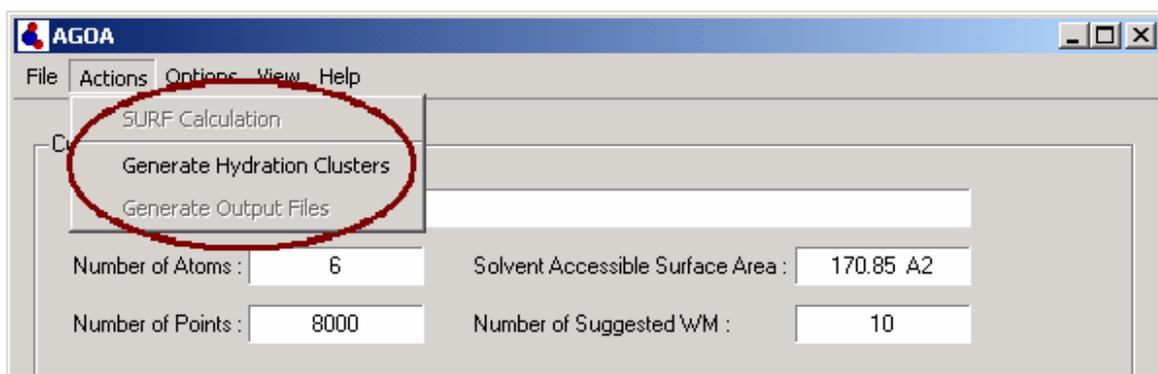


Figura B.6 - O menu *File* do programa AGOA

### B.2.2 MENU *ACTIONS*

O menu *Actions* possui as principais ações do programa e pode ser visto na Figura B.3. Abaixo, seguem as descrições das ações deste menu.

1. *SURF Calculation*: Determina a Área Superficial de Acessibilidade ao Solvente (SASA) que é utilizada no cálculo do número de moléculas de água a serem geradas.
2. *Generate Hydration Clusters*: Define as coordenadas das moléculas de água e gera as estruturas de hidratação a partir da malha tridimensional previamente carregada.
3. *Generate Output Files*: Gera os arquivos de saída para posterior visualização de suas propriedades.



**Figura B.7 - O menu *Actions* do programa AGOA**

### **B.2.3 MENU *OPTIONS***

O menu *Options* é responsável por manter as configurações necessárias para o funcionamento correto do programa. Através dele, o usuário modifica os valores de parâmetros de todas as funcionalidades do programa. Por exemplo, o valor de raio de corte para o soluto utilizado no procedimento de geração das estruturas de hidratação e a seleção dos arquivos a serem gerados pelo programa. A Figura B.4 apresenta o menu *Options* e nas Figuras B.5, B.6-7 serão mostradas as janelas de diálogo *Parameters* e *Preferences* respectivamente.

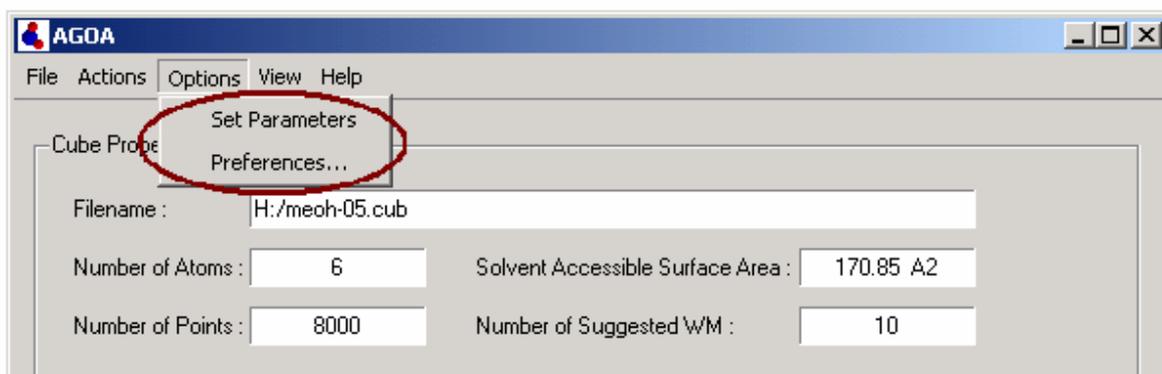


Figura B.8 - O menu *Options* do programa AGOA

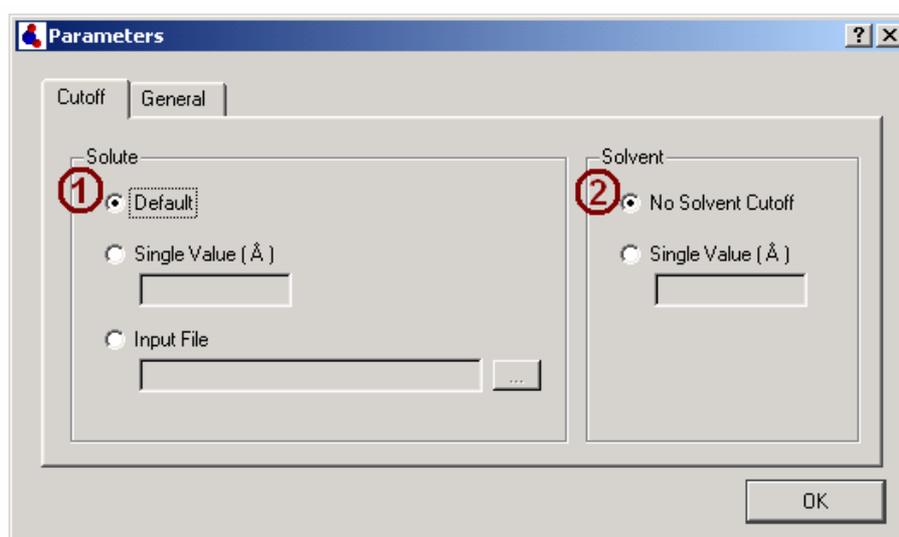


Figura B.5 - A janela de diálogo *Parameters* do menu *Options*

Legenda da Figura B.5:

1. Este campo de seleção é responsável pela escolha do raio de corte para o soluto. As opções existentes são: *Default* (Padrão), *Single Value* (Valor Único) e *Input File* (Arquivo de Entrada), já comentados anteriormente.
2. Este campo de seleção é responsável pela escolha do raio de corte para o solvente. As opções existentes são: *No Solvent Cutoff* (Nenhum) e *Single Value* (Valor Único), já comentados anteriormente.

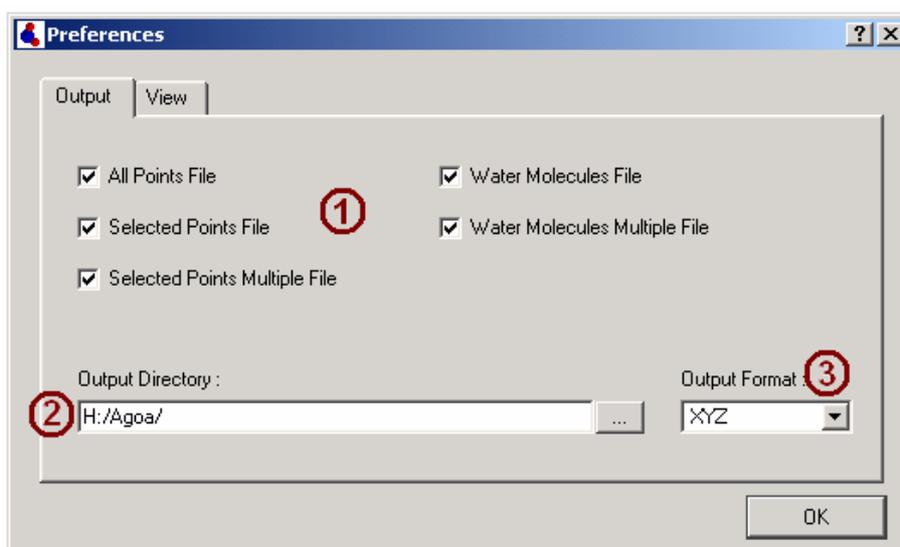


Figura B.6 - A aba *Output* pertencente à janela de diálogo *Preferences* do menu *Options*

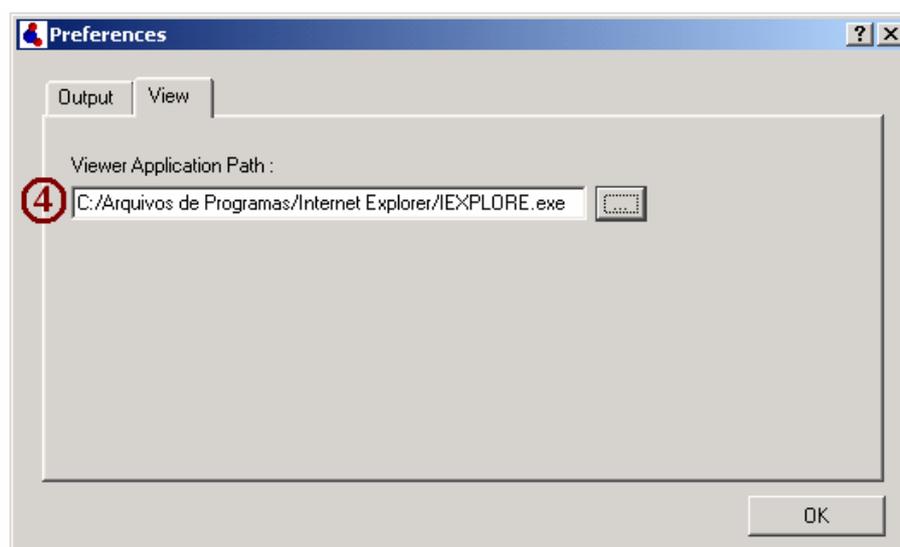


Figura B.7 - A aba *View* pertencente à janela de diálogo *Preferences* do menu *Options*

Legenda das Figuras B.6 e B.7:

1. Esta área é responsável pela seleção dos arquivos de saída a serem gerados pelo programa.
2. Este campo informa o local da geração dos arquivos de saída
3. Este campo fornece o formato do arquivo de saída. É possível escolher entre os formatos XYZ e PDB.

4. Este campo define o caminho do arquivo executável do programa de visualização utilizado externamente pelo programa AGOA.

### B.2.4 MENU VIEW

Através do menu *View*, é possível visualizar as estruturas de hidratação previamente calculadas com o auxílio de um programa de visualização externo que interpreta os formatos de saída gerados pelo programa AGOA.

Ao realizar a leitura do arquivo cube, já é possível visualizar a molécula de soluto dentro dos limites da grade tridimensional. Após a geração das moléculas de água, o programa disponibiliza as opções de visualização da molécula de soluto com: (i) os pontos que restaram após as exclusões de pontos deste procedimento, (ii) os pontos selecionados nesta geração (pontos de máximo ou mínimo e seus respectivos pontos de gradiente numérico) e (iii) as próprias moléculas de água geradas.

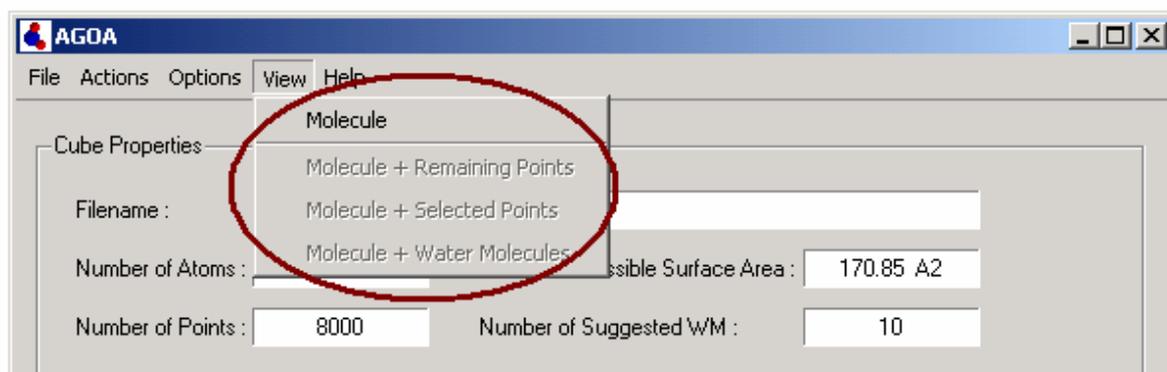


Figura B.8 - O menu *View* do programa AGOA

## B.3 ARQUIVOS DE SAÍDA

Nesta seção serão descritos os arquivos de saída gerados pelo programa AGOA. Em todos estes arquivos, a molécula de soluto e as bordas da grade tridimensional sempre são incluídas com o conteúdo específico de cada um deles.

Também é importante salientar que todos os arquivos de entrada e saída envolvidos tem o formato ASCII e, portanto, são completamente acessíveis para os usuários.

### B.3.1 O ARQUIVO *ALLPOINTS*

O arquivo *AllPoints* contém todos os pontos da grade tridimensional de potencial eletrostático do arquivo *cube* lido pelo programa. Através deste, é possível visualizar a malha original sem as exclusões de pontos durante o procedimento de geração das estruturas de hidratação. Um exemplo deste arquivo pode ser visto a seguir.

```
8006
All Points.
C 0.000000 0.000000 0.000000
O 0.000000 0.000000 1.424367
H 0.925856 0.000000 1.714087 //
H 0.484582 -0.892044 -0.416329 // Molécula
H -1.046301 0.000019 -0.305702 //
H 0.484421 0.892044 -0.416298
XX -5.000000 -5.000000 -5.000000 0.000037
XX -4.500000 -5.000000 -5.000000 0.000129 // Pontos
... // da grade
XX -2.000000 -5.000000 -5.000000 0.000757
```

### B.3.2 O ARQUIVO *SELECTEDPOINTS*

O arquivo *SelectedPoints* contém os pontos de máximo e mínimo selecionados pelo programa juntamente com seus respectivos pontos de gradiente numérico. A seguir, um exemplo deste arquivo.

```
252
Maxima points with gradients.
C 0.000000 0.000000 0.000000
O 0.000000 0.000000 1.424367
H 0.925856 0.000000 1.714087 // Molécula
H 0.484582 -0.892044 -0.416329
H -1.046301 0.000019 -0.305702
H 0.484421 0.892044 -0.416298
XX 2.000000 -0.000000 2.500000 0.052911
XE 2.500000 -0.500000 3.000000 0.020721 // Pontos
... // selecionados
XX 2.000000 -2.000000 0.500000 0.011828
XE 1.500000 -2.500000 1.000000 0.004911
```

```

B   -5.000000   -5.000000   -5.000000
B   -5.000000   -5.000000   -4.500000   // Borda
...                                     // da grade
B    4.000000    4.500000    4.500000
250
Minima points with gradients.
C    0.000000    0.000000    0.000000
O    0.000000    0.000000    1.424367
H    0.925856    0.000000    1.714087   //
H    0.484582   -0.892044   -0.416329   // Molécula
H   -1.046301    0.000019   -0.305702   //
H    0.484421    0.892044   -0.416298
XX   -1.500000   -1.000000    2.500000   -0.043452
XE   -2.000000   -1.500000    3.000000   -0.021820   // Pontos
...                                     // selecionados
XX   -1.500000    2.500000    2.000000   -0.017465
XE   -2.000000    3.000000    1.500000   -0.010203
B   -5.000000   -5.000000   -5.000000
B   -5.000000   -5.000000   -4.500000   // Borda
...                                     // da grade
B    4.000000    4.500000    4.500000

```

### B.3.3 O ARQUIVO *SELECTEDPOINTS*MULTIPLE

Possui o mesmo conteúdo do arquivo *SelectedPoints*, mas em formato múltiplo. A seguir, um exemplo deste arquivo.

```

232
Configuration # 1 maxima.
C    0.000000    0.000000    0.000000
O    0.000000    0.000000    1.424367
H    0.925856    0.000000    1.714087   //
H    0.484582   -0.892044   -0.416329   // Molécula
H   -1.046301    0.000019   -0.305702   //
H    0.484421    0.892044   -0.416298
XX    2.000000   -0.000000    2.500000    0.052911   // Pontos
XE    2.500000   -0.500000    3.000000    0.020721   // selecionados
B   -5.000000   -5.000000   -5.000000
B   -5.000000   -5.000000   -4.500000   // Borda
...                                     // da grade
B    4.000000    4.500000    4.500000
...
...
...
232
Configuration # 10 minima.
C    0.000000    0.000000    0.000000
O    0.000000    0.000000    1.424367
H    0.925856    0.000000    1.714087   //
H    0.484582   -0.892044   -0.416329   // Molécula
H   -1.046301    0.000019   -0.305702   //
H    0.484421    0.892044   -0.416298
XX   -1.500000    2.500000    2.000000   -0.017465   // Pontos
XE   -2.000000    3.000000    1.500000   -0.010203   // selecionados
B   -5.000000   -5.000000   -5.000000
B   -5.000000   -5.000000   -4.500000   // Borda

```

```

... // da grade
B 4.000000 4.500000 4.500000

```

### B.3.4 O ARQUIVO WATERMOLECULES

Contém as geometrias TIP4P das moléculas de água geradas pelo programa. A seguir, um exemplo deste arquivo.

```

274
Water molecules (maxima).
C 0.000000 0.000000 0.000000
O 0.000000 0.000000 1.424367 //
H 0.925856 0.000000 1.714087 // Molécula
H 0.484582 -0.892044 -0.416329 //
H -1.046301 0.000019 -0.305702
H 0.484421 0.892044 -0.416298
O 2.000000 -0.000000 2.500000
XX 2.086603 -0.086603 2.586603
H 2.539232 0.268293 3.243494
H 2.137057 -0.945111 2.432794 //
... // Moléculas de água
O 2.000000 -2.000000 0.500000 //
XX 1.913397 -2.086603 0.586602
H 1.181363 -1.761341 0.934984
H 2.142349 -2.914948 0.741834
B -5.000000 -5.000000 -5.000000
B -5.000000 -5.000000 -4.500000 // Borda
... // da grade
B 4.000000 4.500000 4.500000
274
Water molecules (minima).
C 0.000000 0.000000 0.000000
O 0.000000 0.000000 1.424367 //
H 0.925856 0.000000 1.714087 // Molécula
H 0.484582 -0.892044 -0.416329 //
H -1.046301 0.000019 -0.305702
H 0.484421 0.892044 -0.416298
O -2.000000 -1.500000 3.000000
XX -1.913398 -1.413398 2.913397
H -1.209409 -1.022682 3.252947
H -2.113773 -1.301029 2.070765 //
... // Moléculas de água
O -2.000000 3.000000 1.500000 //
XX -1.913398 2.913397 1.586603
H -1.409967 3.276231 2.201689
H -1.913215 2.047481 1.475129
B -5.000000 -5.000000 -5.000000
B -5.000000 -5.000000 -4.500000 // Borda
... // da grade
B 4.000000 4.500000 4.500000

```

### B.3.5 O ARQUIVO WATERMOLECULESMULTIPLE

Possui o mesmo conteúdo do arquivo *WaterMolecules*, mas em formato múltiplo. A seguir, um exemplo deste arquivo.

```

234
Water Configuration (Maximum): 1
C 0.000000 0.000000 0.000000
O 0.000000 0.000000 1.424367 //
H 0.925856 0.000000 1.714087 // Molécula
H 0.484582 -0.892044 -0.416329 //
H -1.046301 0.000019 -0.305702
H 0.484421 0.892044 -0.416298
O 2.000000 -0.000000 2.500000
XX 2.086603 -0.086603 2.586603 // Molécula
H 2.539232 0.268293 3.243494 // de água
H 2.137057 -0.945111 2.432794
B -5.000000 -5.000000 -5.000000
B -5.000000 -5.000000 -4.500000 // Borda
... // da grade
B 4.000000 4.500000 4.500000
...
...
...
234
Water Configuration (Minimum): 10
C 0.000000 0.000000 0.000000
O 0.000000 0.000000 1.424367 //
H 0.925856 0.000000 1.714087 // Molécula
H 0.484582 -0.892044 -0.416329 //
H -1.046301 0.000019 -0.305702
H 0.484421 0.892044 -0.416298
O -2.000000 3.000000 1.500000
XX -1.913398 2.913397 1.586603 // Molécula
H -1.409967 3.276231 2.201689 // de água
H -1.913215 2.047481 1.475129
B -5.000000 -5.000000 -5.000000
B -5.000000 -5.000000 -4.500000 // Borda
... // da grade
B 4.000000 4.500000 4.500000

```

## REFERÊNCIAS BIBLIOGRÁFICAS

- [All87] Allen, M. P.; Tildesley, D. J.; “*Computer Simulation of Liquids*”, Clarendon Press; Oxford, **1987**.
- [BST76] Bonaccorsi, R.; Scroco, E.; Tomasi, J.; *Am. Chem. Soc.* **1976**, 98, 4049.
- [BST77] Bonaccorsi, R.; Scroco, E.; Tomasi, J.; *Am. Chem. Soc.* **1977**, 99, 4546.
- [CC97] Coutinho, K.; Canuto, S.; *Adv. Quantum Chem.*, **1997**, 28, 89.
- [CSJ85] Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L.; *J. Am. Chem. Soc.*, **1985**, 107, 154.
- [CT95] Cramer, C. J.; Truhlar, D. G.; *Rev. Comput. Chem.* **1995**, 6, 72.
- [Eri05] Eric Martz, <http://www.umass.edu/microbio/chime> (acessado em janeiro/**2005**).
- [FLS92] Freitas, L. C. G.; Longo, R. L.; Simas, A. M.; *J. Chem. Soc., Faraday Trans.* **1992**, 189.
- [Gau05] The Gaussian System, <http://www.gaussian.com>, copyright Gaussian Inc., (acessado em janeiro/**2005**).
- [GPIT05] Grupo de Pesquisa em Inovação Terapêutica, <http://www.ufpe.br/gpit>, (acessado em fevereiro/**2005**).
- [GT92] Gaudio, A. C.; Takahata, Y.; *Computers & Chemistry.*, **1992**, 16, 277.
- [HLS02] Hernandez, M. Z.; Longo, R. L.; Silva, J. B. P. da., “AGOA: A Hydration Procedure and Its Application to the 1-Phenyl-Carboline Molecule”. *J. Braz. Chem. Soc.*, **2002**, 13, 1, 36.
- [JM85] Jorgensen, W. L.; Madura, J. D.; *Mol. Phys.*, **1985**, 56, 1381.
- [LNB01] Longo, R. L.; Nunes, R. L.; Bieber, L. W.; *J. Braz. Chem. Soc.* **2001**, 12, 52.
- [Mic05] Microsoft, <http://msdn.microsoft.com/visualc/>, (acessado em janeiro/**2005**).

- [MJ86] Madura, J. D.; Jorgensen, W. L.; *J. Am. Chem. Soc.*, **1986**, 108, 2517.
- [OMG03] OMG. Unified modeling language. Specification v1.5, Object Management Group, March 2003. <http://www.omg.org/>.
- [PDB05] PDB Format Description Version 2.2, [http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2\\_frame.html](http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2_frame.html), (acessado em janeiro/2005).
- [PM01] Politzer, P.; Murray, J. S.; *Rev. Comput. Chem. Soc.* **2001**, 12, 52.
- [Pro05] Protein Data Bank, <http://www.rcsb.org/pdb> (acessado em janeiro/2005).
- [Rei90] Reichardt, C.; “*Solvents and Solvent Effects in Organic Chemistry*”, VCH, 2<sup>nd</sup>ed., New York, NY, **1990**.
- [Rog05] Roger Sayle, <http://www.umass.edu/microbio/rasmol> (acessado em janeiro/2005).
- [TCM99] Tomasi, J.; Cammi, R.; Mennucci, B.; *Int. J. Quantum Chem.* **1999**, 75, 783.
- [TG75] Tapia, O.; Goscinski, O.; *Mol. Phys.* **1975**, 29, 1653.
- [Tro05] Trolltech, <http://trolltech.com>, (acessado em janeiro/2005).
- [Xmo05] XMol Web Page, <http://www.msc.edu/msc/docs/xmol/XMol.html>, (acessado em fevereiro/2005).