



UNIVERSIDADE FEDERAL DE PERNAMBUCO
Centro de Ciências Exatas e da Natureza
Departamento de Química Fundamental
Programa de Pós-Graduação em Química

Tese de Doutorado

Estruturas e Propriedades de Fluidos Supercríticos e Líquidos Iônicos

Vivianni Marques Leite dos Santos

Recife-PE Brasil

Julho / 2004



*UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA FUNDAMENTAL
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA*

Estruturas e Propriedades de Fluidos Supercríticos e Líquidos Iônicos

Vivianni Marques Leite dos Santos*

Tese apresentada ao Programa de Pós-Graduação em Química da UFPE como parte dos requisitos para a obtenção do título de Doutor em Química.

Área de concentração: Físico-Química

Orientador: Prof. Dr. Ricardo Luiz Longo

***Bolsista CNPq**

Recife-PE Brasil

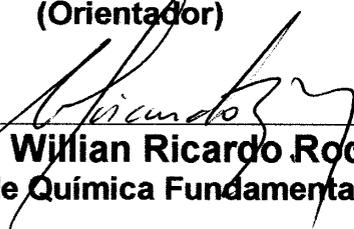
Julho / 2004

Tese submetida ao Corpo Docente do Programa de Pós-Graduação em Química do Departamento de Química Fundamental da Universidade Federal de Pernambuco como parte dos requisitos necessários para a obtenção do Grau de Doutor em Química.

Aprovada:



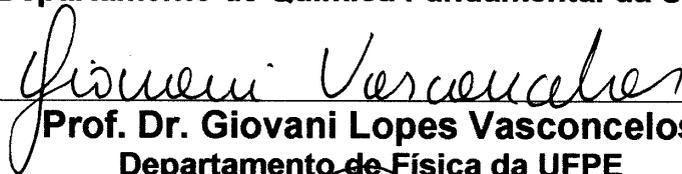
Prof. Dr. Ricardo Luiz Longo
Departamento de Química Fundamental da UFPE
(Orientador)



Prof. Dr. Willian Ricardo Rocha
Departamento de Química Fundamental da UFPE



Prof. Dr. Ricardo de Carvalho Ferreira
Departamento de Química Fundamental da UFPE



Prof. Dr. Giovani Lopes Vasconcelos
Departamento de Física da UFPE



Prof. Dr. Wagner Batista de Almeida
Departamento de Química da UFMG

“Estruturas e Propriedades de Fluidos Supercríticos e Líquidos Iônicos”

por

VIVIANNI MARQUES LEITE DOS SANTOS

Departamento de Química Fundamental
Centro de Ciências Exatas e da Natureza
Universidade Federal de Pernambuco
Recife-PE-Brasil
19 de agosto de 2004

*Aos meus pais, Ataíde e Aída,
ao meu marido, Edilson, e aos
Meu Queridos Filhos
Gabriel e o que está a caminho,
que já faz parte de nossas vidas.*

AGRADECIMENTOS

A Deus, por este momento presente. Por manter-me sempre firme e determinada frente aos obstáculos. Por rodear-me de familiares e amigos nas horas mais difíceis e fazer-me enxergar o lado bom em todas as situações. Enfim, por dar-me esperança, paz e alegria pela vida.

Ao Prof. Dr. Ricardo Longo, pela valiosa formação científica, permitindo-me explorar, com segurança, novos conhecimentos. E também imensamente, pela amizade e incentivo;

Aos Professores do Programa de Pós-Graduação deste Departamento, pelo embasamento teórico dado nas disciplinas;

Aos meus pais, Ataíde e Aída, pelos princípios e educação que tenho hoje, e também pela confiança, amor e carinho que nunca me deixaram faltar;

Ao meu marido, José Edilson Jr., pelo amor, paciência e apoio, necessários ao bom andamento da pesquisa;

Ao meu amado filho, Gabriel, que trouxe muito mais alegria e amor para minha vida;

Ao meu bebê, que já sinto em mim e embora tão pequenino, já mudou pra sempre a minha vida.

Ao meu irmão, Athayde, que esteve mais próximo neste período, pelo estímulo e carinho;

As minhas irmãs, Mylene e Bartira, pela amizade, carinho e estímulo de sempre;

A toda a minha família, pelo apoio e torcida, principalmente Tia Nalda, que por muitas vezes, brincou com Gabriel enquanto eu trabalhava;

Aos meus amigos do Departamento de Química Fundamental pelos momentos de descontração;

Aos meus amigos do Laboratório de Química Teórica e Computacional (LQTC) pelo incentivo e interação durante o andamento do trabalho de pesquisa;

Aos meus amigos do ECC por compartilharem comigo momentos de fé, paz e felicidade;

A todos aqueles que direta ou indiretamente participaram da minha formação na pós-graduação e torceram por mim sem que mesmo eu soubesse;

Aos meus mais sinceros amigos.

RESUMO

Simulações computacionais com o método de Monte Carlo foram realizadas para a água, para o metanol e mistura água-metanol, em condições ambiente e algumas próximas da crítica, utilizando os modelos de quatro (TIP4P) e cinco (TIP5P) sítios para a água e um modelo de três sítios para o metanol. As estruturas geradas nas simulações, foram então utilizadas para análise das ligações de hidrogênio e presença de padrões de *small-world*. O comportamento de *small-world* foi caracterizado para água supercrítica, ao passo que para a água em condições ambiente um cluster gigante é formado e observa-se a percolação no sistema. De forma análoga a mistura água-metanol percola no sistema em condições ambiente e apresenta padrão de *small-world* na condição supercrítica (500 bar e 400°C). No caso do metanol verificamos uma estrutura que o distancia do comportamento randômico, entretanto, a presença ou não de padrões de *small-world* ainda é inconclusiva.

Simulações computacionais com o método de Monte Carlo também foram realizadas para sais contendo o cátion BMI⁺ (1-Butil-3-Metilimidazólio) e os ânions Cl⁻ e BF₄⁻. Os sais deste cátion e ânions são líquidos à temperatura ambiente, sendo então denominados de líquidos iônicos. Verificou-se que a flexibilidade da cadeia n-butil é provavelmente a responsável pela desordem dos sais que os mantém no estado líquido.

As simulações de líquidos, geralmente, necessitam de um grande número de moléculas, de modo que o cálculo das interações pode tornar-se impraticável. Sendo assim, foi testado um método alternativo baseado em redes neurais artificiais (RNAs). Esta técnica foi testada para a água e estendida para descrever as interações entre as espécies iônicas formadoras dos líquidos iônicos. Superfícies de energia potencial (SEPs) foram então preditas por estes modelos baseados em RNAs, com pequenos erros randômicos, tanto para água líquida, como para o par [BMI⁺][BF₄⁻]. Contudo, estes erros ainda são suficientemente grandes para produzir simulações de Monte Carlo da água líquida inconsistentes com os resultados obtidos usando o modelo TIP5P. É importante citar que as simulações de Monte Carlo realizadas usando RNAs são 2 a 12 vezes mais rápidas do que aquelas usando os potenciais analíticos.

Palavras-chave: simulação computacional, fluidos supercríticos, líquidos iônicos, redes neurais artificiais, método de Monte Carlo, *small-world*.

ABSTRACT

Monte Carlo simulations were performed for the water, methanol and water-methanol 1:1 mixture at room and near critical conditions; using four- (TIP4P) and five-sites (TIP5P) models for water and a three-sites model for methanol. The structures of the liquids generated in the simulations were then used in the topological analysis of the hydrogen bond networks in order to detect the presence of small-world patterns and percolation. The small-world behavior was characterized for supercritical water (500 bar and 400°C); however, for water at room conditions a giant cluster was formed that percolated the system. In a similar analysis the water-methanol mixture percolated in the system at room conditions and presented small-world pattern at the condition supercritical (500 bar and 400°C). In the case of the methanol, the behavior was quite different from the random behavior, however, the presence or not of small-world patterns it is still inconclusive.

Monte Carlo simulations were performed on salts composed of the BMI⁺ cation (1-n-butyl-3-methylimidazolium) and the anions Cl⁻ and BF₄⁻, which are liquid at room temperature, thus being called ionic liquids. It was verified that the flexibility of the n-butyl chain is probably the responsible for the disorder of these salts leading to a liquid state.

Computer simulations of liquid usually need a large number of molecules, so the computation of the molecular interactions might be prohibitive for complicated potentials. Thus, an alternative approach was tested where the interaction potentials were modeled by artificial neural network (ANNs). It was tested for water and extended to describe the interactions among the species of the ionic liquids. Intermolecular potential energy surfaces (IPESs) were predicted by ANNs with small random errors for water-water and for the BMI⁺ and BF₄⁻. However, it seems that these errors are still large enough to produce Monte Carlo simulations of liquid water that are inconsistent with the analytical TIP5P model. It is noteworthy to mention that the Monte Carlo simulations performed with ANNs are 2 to 12 times faster than those with the analytical potential.

Key-words: computer simulation, supercritical fluids, ionic liquids, artificial neural networks, Monte Carlo method, small-worlds.

SUMÁRIO

	Página
1. INTRODUÇÃO	1
1.1. Novos Meios Reacionais: Fluidos Supercríticos e Líquidos Iônicos	1
1.1.1. Fluidos Supercríticos	2
1.1.1.a. Fundamentos	2
1.1.1.b. Propriedades dos Fluidos Supercríticos	4
1.1.1.c. Água Supercrítica	4
1.1.2. Líquidos Iônicos	9
1.1.2.a. Fundamentos	9
1.1.2.b. Exemplos de Cátions e Ânions que Formam Líquidos Iônicos e a Variação de suas Propriedades Físicas e Químicas	10
1.1.2.c. Aplicações de Líquidos Iônicos como Solventes em Química	15
1.2. Redes Neurais Artificiais	16
1.2.1. Fundamentos	16
1.2.2. Aplicações de Redes Neurais Artificiais em Química	18
1.2.3. Redes Neurais Artificiais em Ajustes de Superfícies de Energia Potencial (SEP)	20
1.3. Análise Topológica de Redes de Ligações de Hidrogênio	22
1.4. Objetivos	23
1.4.1. Objetivos Gerais	23
1.4.2. Objetivos Específicos	23
2. METODOLOGIA	24
2.1. Introdução	24
2.2. Métodos Experimentais e Propriedades Estruturais da Água	24
2.3. Simulação Computacional	25
2.3.1. Método de Monte Carlo (MC)	26
2.3.2. Determinação das Propriedades Termodinâmicas e Estruturas à partir de Simulações de Monte Carlo	28
2.3.3. Modelos de Potencial Intermolecular da Água	29
2.3.4. Modelos de Potencial Intermolecular do Metanol	31
2.3.5. Potenciais de Interação – Tipos de Parametrização	31

	Pag.
2.3.6. Ligações de Hidrogênio entre Moléculas de Água	32
2.3.7. Ligações de Hidrogênio entre Moléculas de Metanol	33
2.3.8. Ligações de Hidrogênio entre as Moléculas na Mistura Água:Metanol	34
2.4. Métodos de Química Quântica <i>ab initio</i>	34
2.5. Análise de Componentes Principais e Análise de Sensibilidade	36
2.6. Redes Neurais Artificiais	37
2.6.1. Topologia	38
2.6.2. Equações de Redes Neurais Multicamadas	39
2.6.3. Tipos de Redes Neurais Artificiais	41
2.6.4. Método de Retropropagação (<i>BACKPROPAGATION</i>)	43
2.7. Análise Topológica de Redes de Ligações de Hidrogênio	47
3. RESULTADOS E DISCUSSÕES	50
3.1. Introdução	50
3.2. Simulação Computacional da Água usando os Modelos TIP4P e TIP5P.	51
3.2.1. Propriedades Termodinâmicas a 25°C e 1 atm	51
3.2.2. Propriedades Termodinâmicas em condições extremas	52
3.2.3. Funções de Distribuição Radial	55
3.2.4. Estatística das Ligações de Hidrogênio	60
3.2.5 Análise Topológica das Redes de Ligações de Hidrogênio em Água	68
3.3. Ajuste de Superfícies de Energia Potencial para a Água Líquida	76
3.3.1. Análise de Componentes Principais e Análise de Sensibilidade	77
3.3.2. Ajuste inicial da energia de interação usando RNAs	79
3.3.3. Ajuste da energia de interação usando RNAs	82
3.3.4. Aplicação de Modelos de potencial da Água baseados em RNAs em Simulações de Monte Carlo	91
3.4. Simulação Computacional do Metanol	97
3.4.1. Propriedades Termodinâmicas	97
3.4.2. Funções de Distribuição Radial	99
3.4.3. Análise Topológica das Redes de Ligações de Hidrogênio no Metanol	101
3.5. Simulação Computacional da Mistura Água-Metanol 1:1	108

	Pag.
3.5.1. Propriedades Termodinâmicas	108
3.5.2. Funções de Distribuição Radial	109
3.5.3. Análise Topológica das Redes de Ligações de Hidrogênio na Mistura equimolar de Água : Metanol	114
3.6. Líquidos Iônicos	121
3.6.1. Análise conformacional do cátion BMI (1,3 - <i>n</i> -butilmetilimidazólio)	121
3.6.2. Simulações de Monte Carlo Preliminares	123
3.6.3. Ajuste da energia de interação usando RNAs	126
4. CONCLUSÕES E PERSPECTIVAS	131
4.1. Simulação Computacional da água em diferentes condições de temperatura e pressão	131
4.2. Simulação Computacional do Metanol e da Mistura Água-Metanol 1:1 nas Condições Ambiente e Extrema	132
4.3. Análise topológica da mistura água:metanol 1:1 e dos líquidos puros	133
4.4. Simulação Computacional de Líquidos Iônicos	134
4.5. Ajuste de Superfícies de Energia Potencial usando Redes Neurais	135
4.6. Modelos baseados em Redes Neurais Artificiais em Simulações de Monte Carlo	136

APÊNDICE 1 – Participação em *Workshops* e Trabalhos Enviados para Congressos durante o desenvolvimento deste trabalho.

APÊNDICE 2 – Artigo publicado: Topology of the Hydrogen Bond Network in Liquid Water at Room and Supercritical Conditions: A Small-World Structure. *Chem. Phys. Lett.* 390: 157–161 (2004).

LISTA DE ABREVIACOES E SIGLAS

AICl ₃	– Cloreto de Alumnio
AI	– <i>Artificial Intelligence</i> ou Inteligncia artificial
ANN	– <i>Artificial Neural Network</i>
BPN	– <i>BackPropagation Network</i> – Rede <i>Backpropagation</i>
BMI ⁺ ≡ BMI	– Butil metil Imidazlio
BSSE	– <i>Basis Set Superposition Error</i> – Erro de superposio de base
CH ₃	– Metil
Cl	– Cloreto
CO	– Monxido de Carbono
EMI	– Etil metil Imidazlio
GRNN	– <i>Generalized Regression Neural Networks</i> - Redes Neurais de Regresso Generalizada
HF	– Hartree-Fock
MC	– Monte Carlo
MD	– <i>Molecular Dynamic</i> – Dinmica Molecular
MLP	– <i>Multilayer Perceptron</i> – Rede Perceptron Multicamada
MPn	– Teoria de perturbao de Mller e Plesset de ordem n
NCC	– Modelo de Nieser, Corongiu e Clementi
NpT	– <i>ensemble</i> isobrico-isotrmico cannico (N = nmero de molculas, p = presso, T = temperatura)
NO ₃ ⁻	– Nitrato
NVT	– <i>ensemble</i> cannico
OPLS	– <i>Optimized Parameters for Liquid Simulations</i>
PCA	– Principal Components Analysis
PF ₆	– Hexafluoreto de fsforo
PLS	– <i>Partial Least Square</i>
PNN	– <i>Probabilistic Neural Networks</i> - Redes Neurais Probabilsticas
PSP	– <i>Post-Synaptic Potential</i> - potencial ps-sinptico
RBF	– <i>Radial Basis Function</i> – Redes de Funo de Base Radial
RDFs	– <i>Radial Distribution functions</i> – Funes de Distribuio Radial
RNAs	– Redes Neurais Artificial
SEP	– Superfcie de Energia Potencial

SPC	– <i>Simple Point Charge</i> – Modelo de carga pontual simples
SW	– <i>Small-world</i>
TIPnP	– <i>Transferable intermolecular potential n-polarization</i>
UCP	– Unidade Central de Processamento
UV-VIS	– Ultra-violeta - Visível
X ⁻	– Ânions de líquidos iônicos
μVT	– <i>ensemble</i> gran-canônico

LISTA DE SÍMBOLOS

Símbolos Latinos:

C	– coeficiente de agregação ou <i>clustering</i>
C_p	– Capacidade calorífica a pressão constante
C_v	– Capacidade calorífica a volume constante
E	– função objetivo para treinamento
E_{ab}	– Energia de interação entre duas moléculas, a e b
E_p	– erro em relação a cada padrão
f	– função de ativação – Referida a cada neurônio da rede neural
$\langle k \rangle$	– Grau médio (<i>average degree</i>) ou conexões, ou <i>links</i> ou ligações de hidrogênio da rede
L	– camada de saída da rede neural <i>feedforward</i> – RNAs
L	– comprimento do caminho médio – Topologia de redes
n_h	– número de neurônios na camada escondida
n_i	– número de neurônios na camada de entrada
n_k	– número de neurônios na camada k
n_L	– número de neurônios na camada de saída
N_p	– número total de padrões para treinamento - RNAs
N	– número de sítios ou nós de uma rede – Topologia de redes
$P(k)$	– grau de distribuição – Topologia de redes
q	– carga do átomo
S	– direção de busca para otimização
$S_{pi,k}$	– saída (ou ativação) de um neurônio i da camada k , para um padrão p
t_{pl}	– saída alvo de um neurônio “ l ” da camada de saída, para um padrão p

$U(\bar{r})$	– Energia de interação total de uma dada configuração
$w_{ji,k}$	– peso entre o neurônio “j” da camada (k+1) e o neurônio “i” da camada precedente

Símbolos Gregos:

ρ	– densidade
$\rho(\lambda)$	– espectro do gráfico
dH_{na}	– mudança de entalpia envolvida na mudança do estado “u” para o estado “n”.
ε	– Parâmetro de energia de <i>Lennard-Jones</i>
ε	– constante de momento no método do <i>gradiente descendente</i> - RNAs
$\lambda_{pj,k+1}$	– função auxiliar que indica a soma ponderada de todas as entradas de um neurônio “j” da camada (k+1), para um padrão p, mais o seu “bias”
μ_{ji}	– centros adaptativos da função de ativação gaussiana de um neurônio “j”
η	– taxa de aprendizado no método do <i>gradiente descendente</i>
σ_j	– largura (ou desvio-padrão) da função de ativação gaussiana de um neurônio “j” - RNAs
σ	– Parâmetro de distância de <i>Lennard-Jones</i>
$\theta_{j,k+1}$	– <i>bias</i> (ou limite interno de ativação) do neurônio “j” da camada (k+1)

LISTA DE TABELAS

	Página
Tabela 1.1 - Pontos críticos para alguns fluidos de interesse. ...	3
Tabela 1.2 - Propriedades da água em diferentes condições de temperatura e pressão. ...	5
Tabela 1.3 - Exemplos de aplicações de água em condições extremas. ...	6
Tabela 1.4 - Temperatura de transição de fase para compostos do tipo [1-alkyl-3-methylimidazolium] ⁺ [anion] ⁻	12
Tabela 1.5 - Massas molares, temperaturas de transição de fase e as densidades e viscosidades à 30°C para o líquido iônico 1- <i>n</i> -butil-3-metilimidazólio·X [BMI·X]. ...	14
Tabela 1.6 - Exemplos de aplicações de líquidos iônicos como solventes para química. ...	15
Tabela 1.7 - Erro absoluto médio na energia do dímero para os conjuntos de treinamento e teste. ...	21
Tabela 2.1 - Parâmetros e geometria do monômero para as funções de potencial dos modelos TIP4P e TIP5P. ...	30
Tabela 2.2 - Parâmetros e geometria das funções de potencial para o modelo do metanol. ...	31
Tabela 3.1 - Propriedades termodinâmicas calculadas e experimentais para a água líquida nas condições ambiente (1 atm e 25°C). ...	51
Tabela 3.2 - Protocolos de simulação incluindo razão de aceitação A (%), ângulo máximo de rotação θ_{rot} (°), deslocamento máximo de translação $R_{m\acute{a}x}$ (?), razão de aceitação das variações de volume A_v (%), máxima variação de volume $V_{m\acute{a}x}$ e dimensão final da caixa cúbica. ...	52
Tabela 3.3 - Algumas propriedades termodinâmicas obtidas nas simulações de Monte Carlo da água usando os modelos TIP4P e TIP5P. ...	52
Tabela 3.4 - Densidade (g cm^{-3}) da água em condições extremas. Os valores entre parênteses são os erros percentuais com relação aos dados experimentais. ...	54
Tabela 3.5 - Número médio de ligações de hidrogênio sobre 500 configurações para cada condição de temperatura e pressão simulada. ...	64
Tabela 3.6 - Energia média das interações com ligações de hidrogênio e energia de coesão para cada condição de temperatura e pressão simulada. ...	68

LISTA DE TABELAS	Pag.
Tabela 3.7 - Comprimento do caminho médio (L and L_{rand}), coeficiente de agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio.	70
Tabela 3.8 - Tamanho máximo médio das ilhas para as 500 caixas obtidas nas duas condições analisadas.	74
Tabela 3.9 - Resultados da análise de sensibilidade.	79
Tabela 3.10 - Dados de Teste. As quatro primeiras colunas são as entradas para a RNA e a última coluna apresenta a saída desejada.	80
Tabela 3.11 - Restrição de máximos e mínimos para as entradas e saída das redes treinadas usando-se 486 dados.	81
Tabela 3.12 - Restrição de máximos e mínimos para as entradas e saída da redes treinadas usando-se 405 dados.	82
Tabela 3.13 - Erro quadrático médio (kcal mol^{-1}) ² para os conjuntos de treinamento, validação e teste.	83
Tabela 3.14 - Escalonamento linear interno para os dados de entrada para a RNA 4:4:1.	88
Tabela 3.15 - Pesos (w_{ij}) ajustados para as conexões entre a camada de entrada e a escondida.	89
Tabela 3.16 - Pesos (w_{ij}) ajustados para as conexões entre a camada escondida e a camada de saída.	89
Tabela 3.17 - Limiares ou <i>bias</i> (θ_i) ajustados para os neurônios das camadas escondida e de saída.	89
Tabela 3.18 - Energias e resíduos (kcal mol^{-1}) para três dados de teste variando-se apenas a distância entre os centros de massa.	90
Tabela 3.19 - Demanda computacional para os modelos TIP5P e aquele baseado em RNAs.	90
Tabela 3.20 - Resultados de simulações de Monte Carlo usando RNAs distintas e usando o modelo TIP5P.	93
Tabela 3.21 - Energias totais de interação (kcal mol^{-1}) para uma caixa com 256 (L1) + 256 (L2) moléculas de água, usando o modelo baseado em RNAs e o modelo TIP5P.	95
Tabela 3.22 - Energias médias por molécula (L1L1 + L2L2 + L1L2) para uma caixa com 512 moléculas de água utilizada em simulações de Monte Carlo.	95

LISTA DE TABELAS

	Pag.
Tabela 3.23 - Protocolos de simulação incluindo razão de aceitação A (%), ... ângulo máximo de rotação θ_{rot} ($^{\circ}$), deslocamento máximo de translação $R_{m\acute{a}x}$ (?), razão de aceitação das variações de volume A_v (%), máxima variação de volume $V_{m\acute{a}x}$ e dimensão final da caixa cúbica.	98
Tabela 3.24 - Propriedades termodinâmicas obtidas nas simulações de Monte ... Carlo do metanol.	98
Tabela 3.25 - Comprimento do caminho médio (L and L_{rand}), coeficiente de ... agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio no metanol puro.	104
Tabela 3.26 - Tamanho máximo médio das ilhas para as 500 caixas obtidas nas ... duas condições analisadas.	107
Tabela 3.27 - Simulação de Monte Carlo para o sistema água-metanol 1:1. ...	108
Tabela 3.28 - Propriedades termodinâmicas nas condições ambiente (1 atm ... 25°C).	108
Tabela 3.29 - Propriedades termodinâmicas à 500 bar e 300°C	109
Tabela 3.30 - Propriedades termodinâmicas à 500 bar e 400°C	109
Tabela 3.31 - Densidade e energia média por molécula da mistura água-metanol ... nas condições extremas analisadas.	109
Tabela 3.32 - Comprimento do caminho médio (L and L_{rand}), coeficiente de ... agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio na mistura água-metanol 1:1.	115
Tabela 3.33 - Tamanho máximo médio das ilhas para as 500 caixas obtidas nas ... duas condições analisadas.	120
Tabela 3.34 - Resultados da análise conformacional do BMI usando MM^+	123
Tabela 3.35 - Protocolos de simulação incluindo razão de aceitação para ... mudanças translacionais ou rotacionais A (%), razão de aceitação das variações de volume A_v (%), máxima variação de volume $V_{m\acute{a}x}$ e dimensão final da caixa cúbica.	124
Tabela 3.36 - Propriedades termodinâmicas obtidas nas simulações de Monte ... Carlo de líquidos iônicos [BMI 1][BF_4^-].	125
Tabela 3.37 - Demanda computacional para os modelos analítico e aquele ... baseado em RNAs.	130

LISTA DE FIGURAS

	Página
Figura 1.1 - Diagrama de fases da água (H ₂ O) e dióxido de carbono (CO ₂).....	... 3
Figura 1.2.a - Mecanismo A _{ac} 2 da hidrólise do acetato de etila.	... 7
Figura 1.2.b - Mecanismo da hidrólise do acetato de etila não catalisada.	... 8
Figura 1.3 - Alguns exemplos de cátions em líquidos iônicos	11
Figura 1.4 - Alinhamento de cadeias alquílicas no sal hexafluorofosfato de 1- dodecil-3-metilimidazólio	... 13
Figura 1.5 - Rede Neural <i>feedforward</i> 17
Figura 2.1 - Geometria do monômero da água pelo Modelo TIP4P.	... 30
Figura 2.2 - Geometria do monômero da água pelo Modelo TIP5P.	... 30
Figura 2.3 - Parâmetros geométricos utilizados na determinação de ligações de hidrogênio para o dímero da água.	... 33
Figura 2.4 - Parâmetros geométricos utilizados na determinação de ligações de hidrogênio para o dímero de metanol.	... 33
Figura 2.5 - Diagrama para representação das componentes principais.	... 36
Figuras 2.6 - Intracamada	... 38
Figura 2.7 - Intercamada	... 38
Figura 2.8 - Recorrentes	... 38
Figura 2.9 - <i>feedback</i>	... 38
Figura 2.10 - <i>feedforward</i>	... 38
Figura 2.11 - j-ésimo neurônio na camada (k+1).	... 40
Figura 2.12 - Rede neural <i>feedforward</i> com 3 camadas.	... 43
Figura 3.1 - Variação da densidade da água com o aumento da temperatura usando os modelos TIP5P e TIP4P.	... 55
Figura 3.2 - RDFs O-H para a água a 1 atm 25°C.	... 56
Figura 3.3 - RDFs O-O para a água a 1 atm 25°C.	... 56
Figura 3.4 - RDFs H-H para a água a 1 atm 25°C.	... 57
Figura 3.5 - RDFs O-H e O-O para a água a 400°C e 500 bar.	... 57
Figura 3.6 - RDFs H-H para a água a 400°C e 500 bar.	... 58
Figura 3.7 - RDFs O-H para a água a 300°C e pressões iguais a 500, 1100 e 1970 bar.	... 59
Figura 3.8 - RDFs O-O para a água a 300°C e pressões iguais a 500, 1100 e 1970 bar.	... 59

LISTA DE FIGURAS	Pag.
Figura 3.9 - RDFs H-H para a água a 300°C e pressões iguais a 500, 1100 e 1970 bar. ...	59
Figura 3.10 - RDFs H-H para a água a 48 bar e temperaturas 50 e 125°C. ...	60
Figura 3.11 - Distribuição angular das ligações de hidrogênio O-H...O da água em concordância com os critérios geométricos e energéticos. ...	61
Figura 3.12 - Porcentagem de moléculas que formam de uma a sete ligações de hidrogênio. Condições ambiente. Total de ligações de hidrogênio: 861. ...	62
Figura 3.13.a-i - Porcentagem de moléculas que formam de uma a sete ligações de hidrogênio. Diferentes condições de temperatura e pressão. ...	62-63
Figura 3.14 - Dependência do número de ligações de hidrogênio com a pressão para T = 300°C. ...	65
Figura 3.15 - Dependência do número de ligações de hidrogênio com a temperatura para p = 1 atm. ...	65
Figura 3.16.a-i - Porcentagem de moléculas que formam ligações de hidrogênio em função das energias de interação de ligação de hidrogênio. ...	66-67
Figura 3.17 - Caixa de simulação com 512 moléculas após equilíbrio à T = 25°C e p = 1 atm. O volume da caixa é (24,75 Å) ³	68
Figura 3.18 - Caixa de simulação com 512 moléculas após equilíbrio. p = 500 bar e T = 400°C. O volume da caixa é (34,84 Å) ³	69
Figure 3.19 - Rede de ligações de hidrogênio da água líquida em condições ambiente. Cada círculo (sítio) representa uma molécula de água e as conexões representam as ligações de hidrogênio. ...	71
Figure 3.20 - Rede de ligações de hidrogênio (esquerda) e o maior agregado da rede (direita) para a água supercrítica. Cada círculo (sítio) representa uma molécula de água e as conexões representam as ligações de hidrogênio. ...	71
Figura 3.21 - Grau de distribuição para rede de ligações de hidrogênio da água líquida nas condições ambiente. ...	72
Figura 3.22 - Grau de distribuição para rede de ligações de hidrogênio da água supercrítica. ...	72
Figura 3.23 - Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação da água nas condições ambiente. Para N = 512 formam-se apenas três ilhas com 1, 2 e 509 componentes. Para N = 300 formam-se duas ilhas com 1 e 299 componentes. ...	73

LISTA DE FIGURAS

	Pag.
Figura 3.24 - Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação da água na condição supercrítica. ...	74
Figura 3.25a - Densidade espectral média reescalada para as redes de ligações de hidrogênio da água nas condições ambiente para caixas de simulação contendo 512 e 300 moléculas. ...	75
Figura 3.25b - Densidade espectral média reescalada para as redes de ligações de hidrogênio da água na condição extrema 500 bar 400°C, para caixas de simulação contendo 512 moléculas. ...	75
Figura 3.26 - Diversas configurações superpostas para o dímero da água (molécula 1 no centro, e a molécula 2 ocupando cada posição superposta ao redor da molécula central). ...	76
Figura 3.27 - Dímero da água. Modelo TIP5P. ...	77
Figura 3.28 - Análise de componentes principais para o conjunto de treinamento inicial contendo 486 dados. A abscissa representa a primeira componente e a ordenada a segunda componente principal. ...	77
Figura 3.29 - Análise de componentes principais para o conjunto de treinamento contendo 405 dados. A abscissa representa a primeira componente e a ordenada a segunda componente principal. X-expl: 98%, 1% significa que a primeira componente captura 98% e a segunda 1% da variância. ...	78
Figura 3.30 - Conjunto de teste para predição da energia potencial usando-se o ajuste por RNAs. Entradas para os modelos são os quadrados das distâncias. ...	81
Figura 3.31 - Conjunto de teste para predição da energia potencial usando-se o ajuste por RNAs. Entradas para os modelos são o inverso dos quadrados das distâncias. ...	81
Figura 3.32 - Residual (valor predito – valor TIP5P) para as RNAs analisadas. ...	82
Figura 3.33 - Representações esquemáticas de RNAs utilizadas. ...	84
Figura 3.34 - Superfície de energia potencial para o dímero da água usando o modelo TIP5P. ...	85
Figura 3.35 - Superfícies de energia potencial para o dímero da água, obtidas usando os modelos de redes neurais a serem testados. ...	85
Figura 3.36a - Residuais em kcal mol ⁻¹ gerados pelos modelos baseados em RNAs. ...	86

LISTA DE FIGURAS	Pag.
Figura 3.36b - Residuais em kcal mol ⁻¹ gerados pelos modelos baseados em RNAs.	87
Figura 3.37 - Predições da energia para os quatro modelos baseados em RNAs comparados com TIP5P.	87
Figura 3.38 - Rede neural <i>feedforward</i> para predição da superfície de energia potencial.	88
Figura 3.39 - Superfícies de energia potencial para o dímero da água, obtidas com o TIP5P e a RNA 4:4:1.	89
Figura 3.40 - Energia do conjunto de dados de treinamento (Sistemático 2) com relação a distância R16 (O ₁ ... O ₂ : oxigênios da moléculas 1 e 2)..	92
Figura 3.41 - Configuração de pares de moléculas de água usadas para o treinamento de RNAs (caixa equilibrada em condição ambiente de temperatura e pressão).	94
Figura 3.42 - Comportamento da energia média por molécula (E/N) durante a simulação de Monte Carlo em funções de número de configurações geradas.	96
Figura 3.43 - Caixa de simulação equilibrada em condições ambiente, contendo 512 moléculas de metanol (três sítios: CH ₃ , O, H). Volume da caixa é (32,82 Å) ³ .	97
Figura 3.44 - RDFs O-H para o metanol em condições ambiente e extremas.	99
Figura 3.45 - RDFs O-O para o metanol em condições ambiente e extremas.	99
Figura 3.46 - RDFs H-H para o metanol em condições ambiente e extremas.	100
Figura 3.47 - RDFs O-O experimental e simulada para o metanol na condição ambiente.	101
Figura 3.48 - Distribuição angular das ligações de hidrogênio em metanol puro nas condições ambiente.	102
Figura 3.49 - Ligação de hidrogênio rejeitada pelo critério energético.	102
Figura 3.50 - Rede de ligações de hidrogênio no metanol puro nas condições ambiente. Maior agregado da rede à direita. Cada círculo representa uma molécula de metanol. O volume da caixa é (32,82 Å) ³ .	103
Figura 3.51 - Rede de ligações de hidrogênio (esquerda) e o maior agregado da rede (direita) no metanol puro à 500 bar 400°C. Cada círculo representa uma molécula de metanol. O volume da caixa é (42,85 Å) ³ .	103
Figura 3.52 - Grau de distribuição das ligações de hidrogênio do metanol puro nas condições ambiente.	105

LISTA DE FIGURAS	Pag.
Figura 3.53 - Grau de distribuição das ligações de hidrogênio do metanol na condição supercrítica (500 bar 400°C). ...	105
Figura 3.54 - Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação do metanol nas condições ambiente. ...	106
Figura 3.55 - Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação do metanol à 500 bar 400°C. ...	106
Figura 3.56 - Densidade espectral média reescalada para as redes de ligações de hidrogênio do metanol nas condições ambiente. ...	107
Figura 3.57 - Densidade espectral média reescalada para as redes de ligações de hidrogênio do metanol na condição extrema (500 bar 400°C). ...	107
Figura 3.58 - RDFs O-H para a água pura e na mistura 1:1 nas condições ambiente. ...	110
Figura 3.59 - RDFs O-O para a água pura e na mistura 1:1 nas condições ambiente. ...	110
Figura 3.60 - RDFs H-H para a água pura e na mistura 1:1 nas condições ambiente. ...	110
Figura 3.61 - RDFs O-H para o metanol puro e na mistura 1:1 nas condições ambiente. ...	110
Figura 3.62 - RDFs O-O para o metanol puro e na mistura 1:1 nas condições ambiente. ...	110
Figura 3.63 - RDFs HH para o metanol puro e na mistura 1:1 nas condições ambiente. ...	111
Figura 3.64 - RDFs O-H para a água pura e na mistura 1:1 em 500 bar 300°C. ...	112
Figura 3.65 - RDFs O-O para a água pura e na mistura 1:1 em 500 bar 300°C. ...	112
Figura 3.66 - Valores das RDFs HH para a água pura e na mistura 1:1 em 500 bar 300°C. ...	112
Figura 3.67 - RDFs O-H para o metanol puro e na mistura 1:1 em 500 bar 300°C. ...	112
Figura 3.68 - RDFs O-O para o metanol puro e na mistura 1:1 em 500 bar 300°C. ...	112
Figura 3.69 - RDFs H-H para o metanol puro e na mistura 1:1 em 500 bar 300°C. ...	113
Figura 3.70 - RDFs entre o átomo O da <u>água</u> e o átomo H do <u>metanol</u> em condições ambiente e críticas. ...	113

LISTA DE FIGURAS	Pag.
Figura 3.71 - RDFs entre o átomo O da <u>água</u> e o átomo O do <u>metanol</u> em ... condições ambiente e críticas.	113
Figura 3.72 - RDFs entre o átomo H da <u>água</u> e o átomo H do <u>metanol</u> em ... condições ambiente e críticas.	114
Figure 3.73 - Rede de ligações de hidrogênio da mistura água-metanol 1:1 em ... condição ambiente. O volume da caixa é $(30,10 \text{ \AA})^3$.	115
Figure 3.74 - Rede de ligações de hidrogênio (esquerda) e o maior agregado da ... rede (direita) para a mistura água-metanol em condição supercrítica (500 bar 400°C). O volume da caixa é $(40,19 \text{ \AA})^3$.	116
Figura 3.75 - Grau de distribuição das ligações de hidrogênio das moléculas de ... água (esquerda) e metanol (direita) na mistura água-metanol 1:1 nas condições ambiente.	117
Figura 3.76 - Grau de distribuição das ligações de hidrogênio de todas as ... moléculas presentes na mistura água-metanol 1:1 nas condições ambiente.	117
Figura 3.77 - Grau de distribuição das ligações de hidrogênio das moléculas de ... água (esquerda) e metanol (direita) na mistura água-metanol 1:1 à 500 bar e 400°C.	118
Figura 3.78 - Grau de distribuição das ligações de hidrogênio de todas as ... moléculas presentes na mistura água-metanol 1:1 à 500 bar e 400°C.	118
Figura 3.79 - Número de ilhas com uma quantidade de componentes entre: 2^{bin-1} ... – $(2^{bin} - 1)$ para uma das caixas de simulação da mistura binária água-metanol nas condições ambiente (esquerda) e na condição supercrítica (direita).	119
Figura 3.80 - Número de ilhas com uma quantidade de componentes entre: 2^{bin-1} ... – $(2^{bin} - 1)$ para caixas de simulação contendo líquidos puros e a mistura à temperatura ambiente (esquerda) e à 500 bar e 400°C (direita).	119
Figura 3.81 - Densidade espectral média reescalada para as redes de ligações ... de hidrogênio da mistura água-metanol 1:1 nas condições ambiente.	120
Figura 3.82 - Densidade espectral média reescalada para as redes de ligações ... de hidrogênio da mistura água-metanol 1:1 na condição extrema (500 bar 400°C).	120
Figura 3.83 - Estrutura do cátion 1,3 – <i>n</i> -butil-metilimidazólio (BMI) - MP2/6- ... 31G(d, p).	122
Figura 3.84 - Estrutura do ânion BF_4^- - MP2/6-31+G(d, p). ...	122
Figura 3.85 - Estrutura dos ânions PF_6^- (esquerda) e CH_3COO^- (direita). ...	122

LISTA DE FIGURAS	Pag.
Figura 3.86 - Ângulos diédricos f_1 , f_2 e f_3 para análise conformacional do cátion BMI.	123
Figura 3.87 - Cargas ajustadas usando o módulo ESP do programa <i>NWCHEM</i> 4.1.	124
Figura 3.88 - Caixa de simulação equilibrado contendo os líquidos iônicos [BMI 1][BF ₄ ⁻] (esquerda) e [BMI 2][BF ₄ ⁻] (direita) nas condições ambiente.	125
Figura 3.89 - Caixa de simulação equilibrada contendo o líquido iônico [BMI 1][Cl] (esquerda) e [BMI 3][Cl] (direita).	126
Figura 3.90 - Diversas configurações superpostas para o par cátion BMI (centro – conformação 1), e para o ânion BF ₄ ⁻ , ocupando várias posições ao redor do cátion.	127
Figura 3.91 - Par [BMI][BF ₄ ⁻] com numeração dos átomos, cujas distâncias quadradas foram utilizadas como entrada para o modelo baseado em RNAs.	127
Figura 3.92 - Superfície de energia potencial a partir de energias <i>ab initio</i> .	128
Figura 3.93 - Superfície de energia potencial a partir de energias previstas pela RNA 4:3:1 (esquerda) e pela RNA 4:4:1 (direita).	128
Figura 3.94 - Energias <i>ab initio</i> e previstas pela RNA 4:3:1 (esquerda) e pela RNA 4:4:1 (direita).	129
Figura 3.95 - Superfície de energia potencial para o par [BMI][BF ₄ ⁻] usando energias <i>ab initio</i> (conjunto de teste).	129
Figura 3.96 - Superfícies de energia potencial para o par [BMI][BF ₄ ⁻], obtidas usando os modelos de redes neurais.	130

CAPÍTULO 1

INTRODUÇÃO

1. Introdução

Esta tese será apresentada em quatro capítulos. Neste capítulo, detalha-se alguns fundamentos, como: fluidos supercríticos com ênfase para a água supercrítica, líquidos iônicos, redes neurais artificiais e topologia de redes (ligações de hidrogênio) e apresenta-se os objetivos gerais e específicos deste trabalho. No capítulo 2, são descritas as ferramentas utilizadas no desenvolvimento desta tese, como os modelos disponíveis para a simulação da água e mistura água-metanol, métodos para pré-processamento de dados, método de simulação computacional de Monte Carlo, métodos *ab initio* para cálculos de energia de interação, redes neurais artificiais e as técnicas de aprendizado *backpropagation*, de validação cruzada e as propriedades topológicas de redes, incluindo padrões de *small-world*.

No Capítulo 3, são apresentados e discutidos os resultados obtidos. No Capítulo 4, as conclusões e perspectivas sobre os resultados obtidos serão apresentados.

Nos apêndices encontram-se publicações e apresentações e/ou participação em congressos.

1.1. Novos Meios Reacionais: Fluidos Supercríticos e Líquidos Iônicos

Os avanços da ciência e tecnologia são importantes para o aprimoramento da humanidade. Entretanto as novas descobertas podem causar problemas ambientais. Estudos que viabilizem a preservação do meio ambiente estão ligados ao ramo da Química denominado de Química Verde¹, que procura utilizar um conjunto de princípios que reduza ou elimine o uso ou geração de substâncias perigosas durante o projeto, manufatura e aplicação de produtos químicos¹.

A utilização de novos meios reacionais com características atóxicas, tal como água supercrítica, e de não volatilidade como líquidos iônicos, torna-se importante. Dessa forma, são cada vez maiores os estudos das propriedades de solventes alternativos, como a água, que é um fluido com características atípicas tanto em condições ambiente como

¹ P. T. Anastas, J. C. Warner, *Green Chemistry: Theory and Practice*. Oxford University Press. (1998).

naquelas próximas da condição crítica². O entendimento de fluidos supercríticos do ponto de vista microscópico (molecular) tem conduzido ao projeto de novas tecnologias de extração³, separação e metodologias de síntese e reações⁵. Esta possibilidade de aplicação da água partindo-se de previsões, aliada a contribuição de trabalhos teóricos computacionais⁴ que permitiram a detecção de erros em resultados experimentais⁵ ressaltam o valor preditivo das simulações computacionais da água, apesar deste valor preditivo ter sido questionado por Brodsky em 1996⁶.

Nos últimos 10 anos, os líquidos iônicos foram também reconhecidos como uma nova classe de solventes com menor impacto ambiental. Dentre os aspectos que tornam os líquidos iônicos atrativos pode-se citar a baixa volatilidade e a possibilidade de ajustar suas propriedades físicas e químicas. Combinações de cátions e ânions permitem ajustar, de maneira fina, as propriedades necessárias para uma dada reação, criando a possibilidade de inúmeros conjuntos de solventes especificamente projetados, como por exemplo modificar a atividade e seletividade de uma dada reação catalítica⁷. A baixa volatilidade também permite a possibilidade de reciclo do solvente e dos sistemas catalíticos.

A compreensão da estrutura destes líquidos é escassa e são praticamente inexistentes os trabalhos de simulação computacional destes fluidos. Logo, dada a sua importância, pretende-se realizar simulações de Monte Carlo para a predição de propriedades de interesse.

1.1.1. Fluidos Supercríticos

1.1.1.a. Fundamentos

Aquecendo-se um fluido confinado, onde a fase líquida e de vapor coexistem em equilíbrio, as suas propriedades intensivas (densidade, viscosidade, condutividade térmica, etc.) convergem igualando-se no ponto crítico. O ponto crítico assinala a temperatura mais alta de coexistência das duas fases.

Um fluido é chamado supercrítico quando sua pressão e temperatura estão acima do seu ponto crítico (Tabela 1.1). A água apresenta as seguintes constantes críticas:

² A. G. Kalinichev, S. V. Churakov, *Chem. Phys. Lett.* 302, 411-417 (1999).

³ N. L. Rozzi, R. K. Singh, *Comprehensive reviews in food science and food safety*. 1, 33-34 (2002).

⁴ P. Jedlovsky, R. Vallauri, *J. Chem. Phys.* 105, 2391-2398 (1996).

⁵ K. Soper, M. G. Phillips, *Chem. Phys.* 107: 47-60 (1986); A. K. Soper, C. Andreani, M. Nardone, *Phys. Rev. E* 47, 2598-2605 (1993).

⁶ A. Brodsky, *Chem. Phys. Lett.* 261, 563-568 (1996).

⁷ H. Olivier-Bourbigou, L. Magna, *J. Mol. Catal. A: Chem.* 3484, 1-19, (2002).

temperatura 647 K (374°C), pressão 22,08 MPa (220,8 bar) e densidade 0,32 g cm⁻³ (0,0108 moléculas/Å³). Dentre os fluidos supercríticos, além da água, também destacam-se o dióxido de carbono, etano e hidrofluorcarbonos, como o 1,1,1,2-tetrafluoroetano e outros. Os diagramas de fases esquemáticos da água e dióxido de carbono são ilustrados na Figura 1.1.

Tabela 1.1: Pontos críticos para alguns fluidos de interesse⁸.

Fluido	Temperatura Crítica (K)	Pressão Crítica (MPa)	
		MPa	atm
Água	647,00	22,08	218,0
Metanol	513,75	7,69	78,9
Metano	190,55	4,595	45,35
Dióxido de carbono	304,17	7,386	72,89
Propano	369,85	4,247	41,91
Acetileno	308,70	6,247	61,65
Etileno	282,35	5,040	49,74

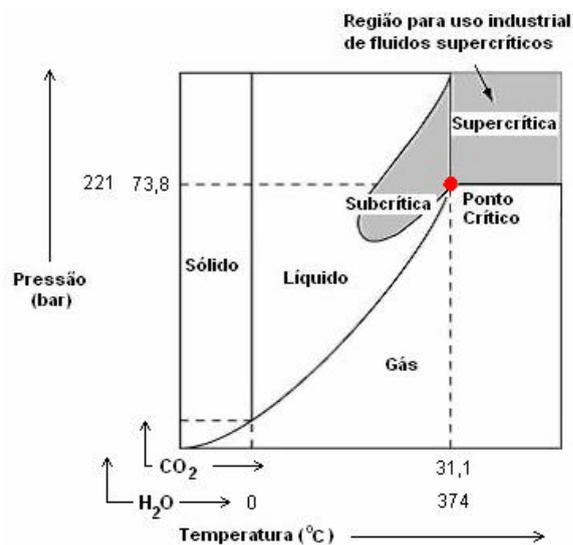


Figura 1.1. Diagrama de fases da água (H₂O) e dióxido de carbono (CO₂).

Fluidos supercríticos estão sendo usados e serão cada vez mais usados, como solventes benignos para o meio ambiente em diversos processos. A importância dos fluidos

⁸ N. L. Rozzi, R. K. Singh, *Comprehensive Reviews in Food Science and Food Safety* 1, 33-34 (2002).

supercríticos para sínteses orgânicas, extrações, separações e outras aplicações é destacada constantemente.

1.1.1.b. Propriedades dos Fluidos Supercríticos

Os fluidos supercríticos apresentam propriedades físico-químicas intermediárias entre aquelas dos gases e dos líquidos. Algumas características dos fluidos supercríticos são:

- 1) Baixa densidade. Por exemplo, a água em condições supercríticas apresenta densidade variando de 0,2 a 0,9 g cm⁻³, que são maiores que a do vapor (10⁻⁴ a 10⁻³ g cm⁻³) e menores que a do líquido (1 g cm⁻³);
- 2) Alta viscosidade dinâmica. Na vizinhança do ponto crítico, o coeficiente de difusão chega a ser cerca de 10 vezes maior do que a do líquido.

A importância dos fluidos supercríticos vêm ainda da possibilidade de sua reutilização ou reciclo, devido a facilidade de recuperação após as reações. Mudando-se as condições de pressão e temperatura, as propriedades dos fluidos são extremamente variadas possibilitando, por exemplo, que o solvente seja separado dos componentes da reação.

Rozzi & Singh (2002)⁹ analisaram a aplicação de fluidos supercríticos em vários aspectos da indústria de alimentos e ressaltaram cinco vantagens para a extração usando fluidos supercríticos:

- 1) Fluidos supercríticos têm altos coeficientes de difusão e menores viscosidades do que os líquidos;
- 2) A menor tensão superficial permite sua mais rápida penetração em poros de matrizes heterogêneas, o que aumenta a eficiências das extrações;
- 3) O controle da seletividade durante a extração pode ser realizado variando as condições de temperatura e pressão, afetando a solubilidade dos vários componentes no fluido;
- 4) Ausência de resíduos;
- 5) Pode ser reciclado e usado novamente, como parte da unidade de operação.

1.1.1.c. Água Supercrítica

A aplicação de água como um meio reacional supercrítico ou subcrítico tem vantagens adicionais. A água tem baixo custo, não é tóxica, nem combustível e nem explosiva. Alcanos e gases como O₂ ou N₂ dissolvem bem em água supercrítica¹⁰.

Estudos mostram que as características da água são peculiares em altas temperaturas e

⁹ N. L. Rozzi, R. K. Singh, *Comprehensive Reviews in Food Science and Food Safety* 1, 33-34 (2002).

¹⁰ P. Krammer, H. Vogel, *J. Supercrit. Fluids* 16, 189-206 (2000).

pressões, como por exemplo, aumento de seu caráter iônico e diminuição de sua constante dielétrica (Tabela 1.2), com conseqüente aumento da solubilidade de compostos orgânicos.

Tabela 1.2: Propriedades da água em diferentes condições de temperatura e pressão¹¹.

Fluido	Água Ambiente	Água Subcrítica	Água	Supercrítica	Vapor
	$T < 150^{\circ}\text{C}$	$150 < T < 350^{\circ}\text{C}$	$T < 370^{\circ}\text{C}$		
	$p < 0,4 \text{ MPa}$	$0,4 < p < 20 \text{ MPa}$	25 MPa	50 MPa	
Temperatura T ($^{\circ}\text{C}$)	25	250	400	400	400
Pressão p (bar)	1	50	250	500	1
Densidade ρ (g cm^{-3})	1	0,80	0,17	0,58	3×10^{-3}
Constante Dielétrica ϵ	78,5	27,1	5,9	10,5	~ 1
pK_w	14,0	11,2	19,4	11,9	-
Capacidade Calorífica C_p ($\text{kJ kg}^{-1} \text{ K}^{-1}$)	4,22	4,86	13,0	6,8	2,1
Viscosidade Dinâmica η (mPa s)	0,89	0,11	0,03	0,07	0,02
Condutividade térmica λ ($\text{mW m}^{-1} \text{ K}^{-1}$)	608	620	160	438	55

Em condições extremas, a água reúne propriedades de seu estado gasoso (como densidade e viscosidade) e de seu estado líquido (capacidade de dissolução) (Tabela 1.2). Estes importantes aspectos para um meio reacional como a água supercrítica vêm sendo demonstrados em diversas reações orgânicas (hidratação, hidrogenação, etc.) e conversões químicas (decomposição e oxidação de compostos orgânicos) (Tabela 1.3).

¹¹ A. N. Jingyi, L. Bagnell, T. Cablewski, C. R. Strauss, R. W. Trainor, *J. Org. Chem.* 62, 2502-2511 (1997); P. Krammer, H. Vogel, *J. Supercrit. Fluids.* 16, 189-206 (2000).

Tabela 1.3: Exemplos de aplicações de água em condições extremas¹².

Reação	Referência
Hidrogenação e desidrogenação de complexos de metais de transição	13
Arilação de Heck	14
Hidratação e desidratação de álcool e olefina	15
Oxidação parcial – Catálises homogêneas e heterogêneas	16
Decomposição de compostos orgânicos	17
Oxidação	12

Krammer & Vogel¹⁸, estudaram a hidrólise de ésteres em água sub- e supercrítica e propuseram um mecanismo de hidrólise na ausência de catalisadores (ácidos ou bases), ilustrando assim uma aplicação para reações com atividade catalítica da água. No mecanismo de catálise ácida do tipo $A_{ac}2$, onde “A” refere-se a catálise ácida, “2” indica um mecanismo bimolecular e “ac” refere-se a ligação que é quebrada durante o decorrer da reação (Figura 1.2.a). Primeiramente a molécula de água age como nucleófilo, atacando o éster protonado. Em seguida, uma outra molécula de água é responsável pela transferência de próton formando o íon hidrônio. No mecanismo não catalisado (Figura 1.2.b), ocorre o ataque direto de uma molécula de água em condições neutras. Nesta etapa uma molécula de água ataca o átomo de carbono da carbonila. A hidrólise é completada via a transferência de prótons, formando álcool e ácido. Este segundo mecanismo explica a hidrólise seletiva para o álcool e ácido sem uso de catalisadores. Provavelmente a nucleoficidade das moléculas de água é aumentada em consequência da fraca rede tridimensional gerada em baixas densidades.

¹² P. E. Savage, *Chem. Rev.* 99, 603-621 (1999).

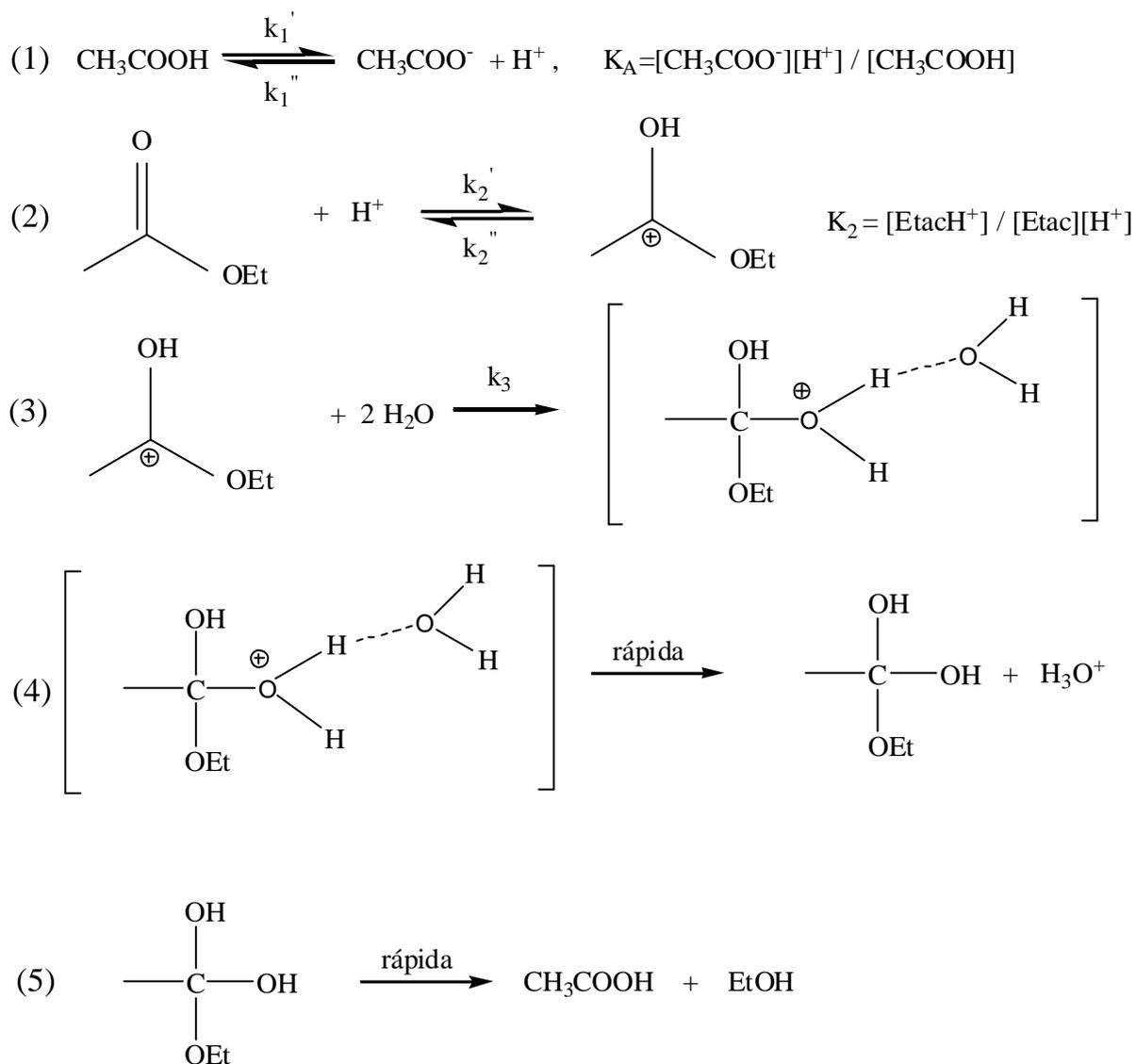
¹³ R. C. Crittendon, E. J. Parsons, *J. Organometal.* 13, 2587-2591 (1994).

¹⁴ P. Reardon, S. Metts, C. Crittendon, P. Daugherty, E. Parsons, *J. Organomet. Chem.* 14, 3810-3816 (1995).

¹⁵ X. Xu, M. J. Antal Jr., *AIChE J.* 9, 1524-1531 (1994).

¹⁶ P. E. Savage, R. Li, J. T. Santini Jr., *J. Supercrit. Fluids.* 7, 135-144 (1994); S. N. V. K. Aki, M. A. Abraham, *J. Supercrit. Fluids.* 7, 250-263 (1994).

¹⁷ M. Nakahara, T. Tenuh, V. Wakai, E. Fujita, H. Enomoto, *Chem. Lett.* 2, 163-164 (1997); T. Minowa, F. Zhen, T. Ogi, *J. Supercrit. Fluids.* 13, 253-259 (1998).

Figura 1.2.a: Mecanismo $A_{ac}2$ da hidrólise do acetato de etila.

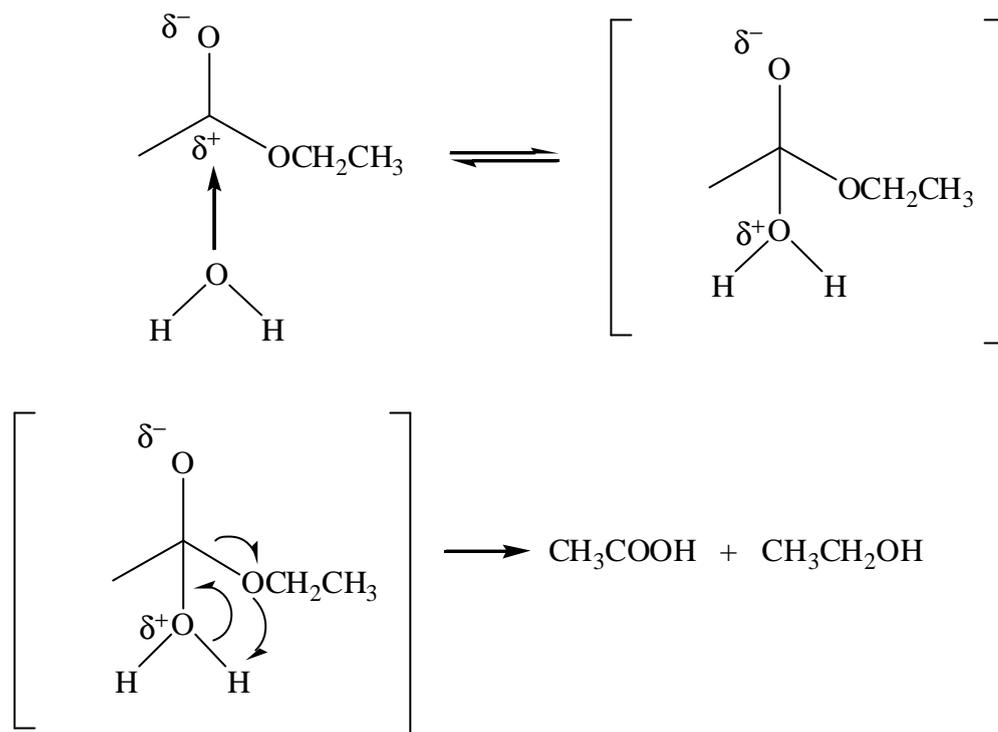


Figura 1.2.b: Mecanismo da hidrólise do acetato de etila não catalisada.

Os resultados obtidos na região subcrítica (250-350°C) confirmam o mecanismo $A_{ac}2$ da hidrólise. O pequeno valor de pK_w (Tabela 1.2) refere-se a este caminho de reação. Acima da temperatura crítica, o mecanismo de reação dominante muda para o ataque nucleofílico direto da água (Tabela 1.2). A baixa densidade e conseqüente diminuição de ligações de hidrogênio deixam moléculas de água livres, aumentando seu caráter nucleofílico.

Condições de altas pressões e temperaturas são difíceis de serem obtidas experimentalmente, exigindo a disponibilidade de equipamentos especiais, com altas resistências térmica e mecânica. Isto pode ser remediado com a utilização de ferramentas computacionais para predição de propriedades de interesse. Controvérsias na literatura sobre a determinação experimental da estrutura da água em condições ambiente também enfocam a preocupação com a confiabilidade de dados experimentais. Soper *et al.*, 1997¹⁹, sugeriram como uma possível explicação de discrepâncias em regiões próximas da crítica, como provenientes das incertezas na correção da inelasticidade da difração de nêutrons, sendo particularmente maior para a água leve. Por isto, estes autores descrevem um novo método de obtenção das funções de correlação de pares, para o qual encontraram resultados mais consistentes. Alguns destes resultados foram utilizados para investigar o

¹⁹ A. K. Soper, F. Bruni, M. A. Ricci, *J. Chem. Phys.* 106, 247-254 (1997).

desempenho²⁰ do modelo TIP4P e estão sendo utilizados neste trabalho para comparações e análises para as simulações com o modelo mais recente, TIP5P²¹.

1.1.2. Líquidos Iônicos

1.1.2.a. Fundamentos

Líquidos iônicos são sais (pares de cátions e ânions) líquidos à temperatura ambiente. Apesar dos líquidos iônicos serem conhecidos desde o começo do século, por exemplo, o [EtNH₃⁺]-[NO₃⁻], o qual tem ponto de fusão igual a 12°C e portanto líquido à 25°C, somente na década de setenta estes líquidos iônicos passaram a ser utilizados como catalisadores em reações orgânicas de alquilação e acilação de olefinas e como solventes para diferentes reações²², emergindo como uma alternativa para a substituição dos solventes utilizados tradicionalmente para diversas reações orgânicas e catalíticas. Apesar do desenvolvimento acelerado de trabalhos experimentais, ainda são escassos os estudos de simulação computacional destes compostos, de modo que é pouco conhecido seu comportamento físico-químico fundamental. Meng *et al.*²³ realizaram cálculos semi-empíricos e *ab initio* para otimização das estruturas em fase gasosa do par [1-butil-3-metilimidazólio⁺][PF₆⁻].

Abaixo estão enumeradas algumas das propriedades físicas e químicas de interesse para aplicações de líquidos iônicos como solventes em sínteses²⁴:

- 1) São bons solventes tanto para materiais orgânicos e inorgânicos. Também combinações incomuns de reagentes podem ser mantidos em uma mesma fase;
- 2) Os líquidos iônicos são polares, mas freqüentemente compostos por íons não coordenantes, constituindo uma alternativa para estudo de propriedades de coordenação de íons metálicos;
- 3) São imiscíveis com vários solventes orgânicos e proporcionam uma alternativa polar não aquosa para sistemas com duas fases. Líquidos iônicos hidrofóbicos também podem ser usados como fases polares imiscíveis com água;
- 4) Apresentam baixa volatilidade e portanto podem ser usados em sistemas a alto vácuo e eliminam muitos problemas de retenção.

²⁰ J. M. M. Cordeiro, *Quím. Nova* 21, 698-701 (1998).

²¹ M. W. Mahoney, W. L. Jorgensen, *J. Chem. Phys.* 112, 8910 (2000).

²² C. S. Consorti, R. F. Souza, J. Dupont, P. A. Z. Suarez, *Quím. Nova* 24, 830-837 (2001).

²³ Z. Meng, A. Dolle, W. Robert Carper, *J. Mol. Struct. (Theochem)* 585, 119-128 (2002).

Até meados de 1998, líquidos iônicos eram considerados raros, mas descobriu-se que muitos sais formam líquidos na temperatura ambiente ou próximo dela, ou seja, são líquidos iônicos²⁵.

Atualmente existem dois métodos gerais e práticos para a preparação de líquidos iônicos. É importante mencionar a presença de impurezas durante sua preparação, visto que suas propriedades físicas e químicas são alteradas pelas impurezas, sendo importante realizar a devida purificação. Os principais contaminantes são ânions haleto ou bases orgânicas que geralmente provêm do material de partida não reagido e também a água, que é difícil de ser removida devido existência de ligações de hidrogênio. Para resolver problemas gerados pelo ânion haleto, como em reações catalisadas por metais de transição, tem-se proposto métodos alternativos de preparação, tal como a alquilação direta de derivados alquilimidazólio²⁶.

1.1.2.b. Exemplos de Cátions e Ânions que Formam Líquidos Iônicos e a Variação de suas Propriedades Físicas e Químicas

Alguns exemplos de combinações de cátions e ânions que formam líquidos iônicos mais comuns são resumidos por Olivier-Bourbigou & Magna (2002)²⁷. Na Figura 1.3 estão ilustrados alguns exemplos de cátions formadores de líquidos iônicos.

Os sais baseados no cátion N, N-dialquilimidazólio (Figura 1.3) são os mais utilizados²⁸, devido à variedade de propriedades físico-químicas disponíveis para esta classe, bem como sua fácil obtenção. Sais de imidazólio líquido são geralmente obtidos pela troca de ânions a partir dos precursores de haletos imidazólio. Olivier-Bourbigou & Magna (2002), descreveram a síntese de brometos de 1-alquil-3-metilimidazólio com tempos de reação curtos e com altos rendimentos (94-99%) sem a etapa de purificação.

²⁴ T. Welton, *Chem. Rev.* 99, 2071-2083 (1999).

²⁵ P. E. Savage, R. Li, J. T. Santini Jr., *J. Supercrit. Fluids.* 7, 135-144 (1994)

²⁶ M. Yoshizawa, W. Ogihara, H. Ohno, *Electrochem. Solid State Lett.* 4 E25 (2001).

²⁷ H. Olivier-Bourbigou, L. Magna, *J. Mol. Catal. A: Chem.* 3484, 1-19, (2002)

²⁸ C. S. Consorti, R. F. Souza, J. Dupont, P. A. Z. Suarez, *Quím. Nova.* 24, 830-837 (2001); Z. Meng, A. Dolle, W. Robert Carper, *J. Mol. Struct. (Theochem)* 585, 119-128 (2002); P. A. Z. Suarez, J. E. L. Dullius, S. Einloft, R. F. De Souza; J. Dupont, *Polyhedron* 15, (1996).

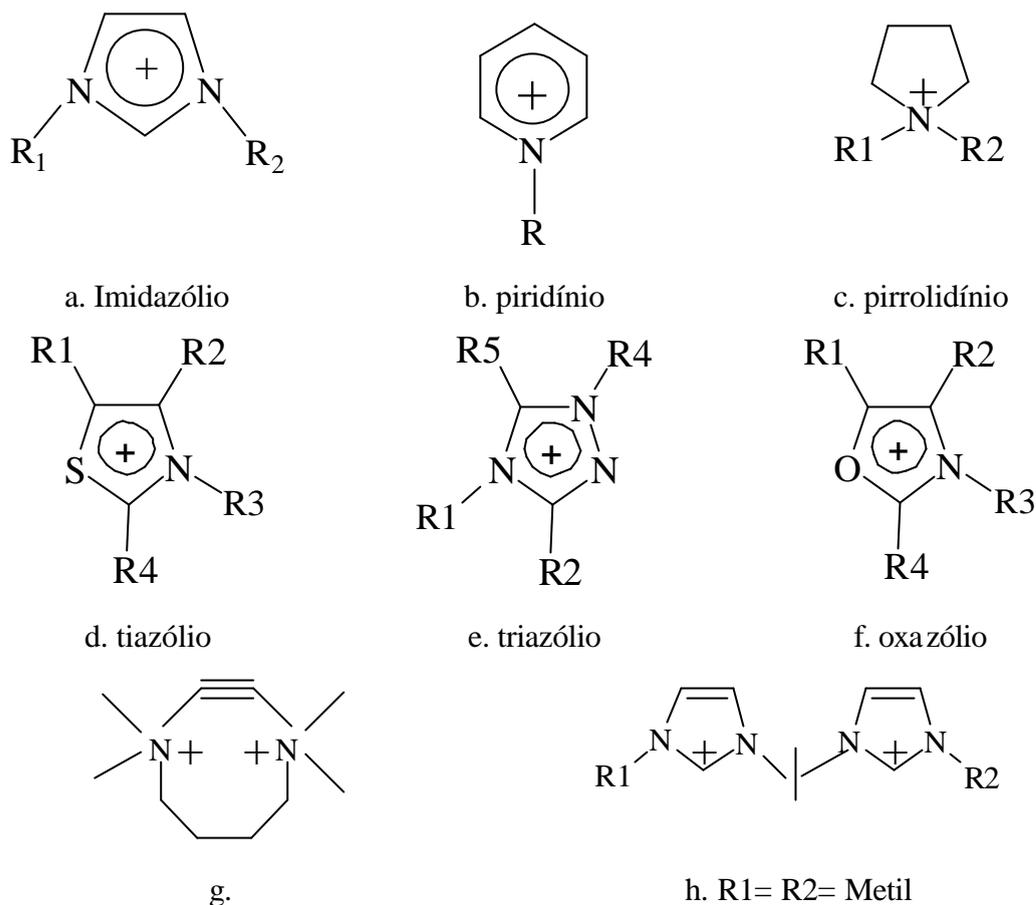


Figura 1.3 Alguns exemplos de cátions em líquidos iônicos

Cátions N, N-diaquil imidazólio assimétricos frequentemente apresentam temperaturas de fusão mais baixas, sendo líquidos à temperatura ambiente. Também alguns simétricos foram encontrados líquidos nestas mesmas condições.

Os dois últimos cátions (g e h) são exemplos de policátions orgânicos e novas categorias de líquidos iônicos vêm sendo descritas como bons candidatos para processos eletroquímicos orgânicos.

Os líquidos iônicos podem ser classificados quanto aos ânions em poli e mononucleares. Exemplos de ânions mais comuns: BF_4^- , AlCl_4^- , Cl^- , Br^- , I^- , PF_6^- , SbF_6^- , CF_3CO_2^- (mononucleares, ou seja, com um átomo central) e Al_2Cl_7^- , $\text{Al}_3\text{Cl}_{10}^-$, Au_2Cl_7^- , Fe_2Cl_7^- , $\text{Sb}_2\text{F}_{11}^-$ (polinucleares, ou seja, mais de um átomo central).

Propriedades Físicas e Químicas

Os líquidos iônicos derivados do 1,3-dialquilimidazólio apresentam uma grande variação nas suas propriedades físico-químicas em função do ânion e dos substituintes

presentes no anel imidazólio. Ainda não foram estabelecidas correlações entre a estrutura e suas propriedades.

Na Tabela 1.4 estão apresentados os valores das temperaturas de transição de fase (fusão) para compostos com cátions do tipo 1-alkil-3-metilimidazólio em função do grupo alquila e do ânion²⁹.

Tabela 1.4: Temperatura de transição de fase para compostos do tipo [1-alkil-3-metilimidazólio⁺][ânion⁻].

Substituinte	Temperatura de Transição de Fase (°C)				
	Ânion				
alquila	Cl ⁻	Br ⁻	I ⁻	BF ₄ ⁻	PF ₆ ⁻
Metil	125	-	-	103,4	-
Etil	84	81	81	15	65
n-propil	66	-	-	-14	-
n-butil	65	61	-	-71	10
n-C ₁₂ H ₂₅	50	-	-	30	60
n- C ₁₄ H ₂₉	57	-	-	42	74
C ₁₆ H ₃₃	60	-	-	50	75
C ₁₈ H ₃₇	77	-	-	67	80

(-) valor não disponível na literatura.

Analisando-se a série com ânion cloreto, coluna 2 da Tabela 1.4, nota-se que a temperatura de transição de fase depende do tamanho da cadeia alquílica. A medida que o tamanho aumenta, verifica-se um decréscimo acentuado do ponto de fusão do sal, porém, quando o número de carbonos torna-se muito elevado (acima de 12), nota-se uma leve tendência de aumento na temperatura de transição.

Consorti *et al.* (2001)²⁹ argumentaram que o decréscimo inicial do ponto de fusão com o aumento da cadeia pode ser explicado pelo aumento do número de graus de liberdade, provocando distúrbios na cela cristalina e dificultando seu empacotamento. Ao passo, que quando o número de carbonos é maior que 12, começa a aparecer o fenômeno de formação de mesofases, sendo gerados cristais líquidos. Este fenômeno sugere que interações fracas do tipo van der Waals começam a orientar as cadeias levando a sistemas organizados³⁰.

²⁹ C. S. Consorti, R. F. Souza, J. Dupont, P. A. Z. Suarez *Quím. Nova* 24, 830-837 (2001).

³⁰ C. J. Bowles, D. W. Bruce, K. R. Seddon, *J. Chem. Soc., Chem. Commun.* 1625 (1996).

Esta tendência em formar mesofases também foi verificada para outros compostos derivados do 1-alkil-3-metilimidazólio, utilizando os ânions PF_6^- e BF_4^- .

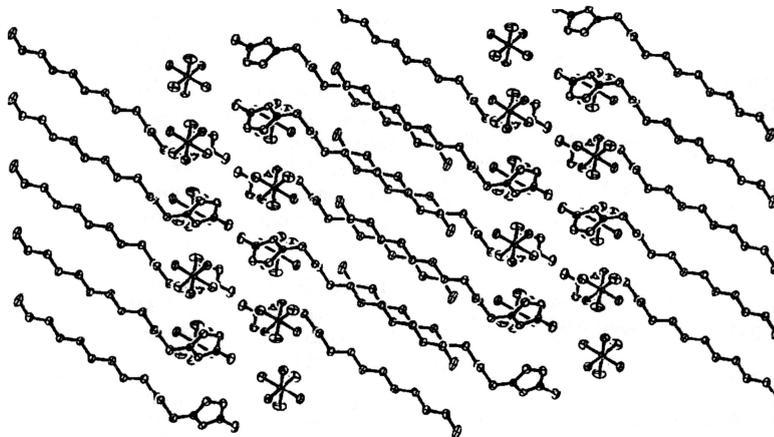


Figura 1.4.: Alinhamento de cadeias alquílicas no sal hexafluorofosfato de 1-dodecil-3-metilimidazólio.

Em geral, a estrutura consiste de líquidos iônicos consiste em contatos interiônicos dominados por forças Coulombianas entre cátion e ânion, com ligações de hidrogênio fracas. Deve-se salientar que a cadeia alquílica possui uma distorção próxima ao anel do cátion, adquirindo uma conformação curva. No caso do composto com substituinte dodecil, verifica-se um alinhamento dos cátions via interações do tipo van der Waals entre as cadeias alquílicas, conforme ilustrado na Figura 1.4. No caso do composto com substituinte etila, este alinhamento não foi verificado.

A densidade dos sais 1-3-dialquilimidazólio em função da temperatura pode ser ajustada por equações do tipo:

$$\rho = a + b(\theta - 60) \quad 1.1$$

onde a e b são constantes e θ é a temperatura em $^{\circ}\text{C}$.

Fannin Jr. *et al.* 1984³¹, estudaram a variação da densidade com o mudança do número de carbonos do grupo alquila R1 do anel imidazólio (Figura 1.3.a) e com a fração molar do alumínio no ânion (AlCl_4^- , Al_2Cl_7^- , ...).

Observaram uma variação linear, em que, aumentando-se a massa do cátion tem-se uma diminuição da densidade. Ao passo que aumentando-se a massa do ânion (fração molar em alumínio) tem-se um aumento da densidade. A correlação entre a massa molecular do ânion e a densidade também pode ser verificada na Tabela 1.5 (linhas 1 e 2)

para ânions distintos. A exceção é o ânion PF_6^- com densidade acima da esperada (1,37 g/mL), que foi explicada pelo argumento de que o empacotamento das espécies é maior, diminuindo o volume específico do sal.

A viscosidade absoluta (η) em sais derivados do cátion imidazólio pode ser ajustada por equações do tipo,

$$\ln(\eta) = k_n/(T/T_0) + 0,5 \ln(T) + \ln(A_n) \quad 1.2$$

onde A_n é um fator dependente da escala na qual a viscosidade absoluta é expressa, T_0 é a temperatura de transição vítrea ‘ideal’ e k_n é uma constante característica do material³¹.

A estrutura do cátion influencia a viscosidade dos líquidos iônicos. Aumentando-se o comprimento do substituinte alquila no cátion 1,3-dialquilimidazólio Fannin Jr. *et al.* (1984)³¹ verificaram um aumento da viscosidade. Este aumento foi relacionado com o aumento das interações do tipo van der Waals entre estas cadeias.

A dependência da viscosidade com a natureza do ânion em sais BMI·X está ilustrada na Tabela 1.5.

Tabela 1.5: Massas molares, temperaturas de transição de fase e as densidades e viscosidades à 30°C³² para o líquido iônico 1-*n*-butil-3-metilimidazólio·X [BMI·X].

Propriedade	Ânion (X)			
	AlCl_4^-	CF_3SO_3^-	BF_4^-	PF_6^-
Massa molar (g/mol)	306	288	226	284
Densidade (g/mL) (30°C)	1,23	1,22	1,17	1,37
Temperatura de transição de fase (°C)			-71	10
Viscosidade (P) (30°C)	-	1,27	2,33	3,12

Nota-se um paralelo entre os valores das viscosidades e a temperatura de transição de fase. Quanto maior a força das interações do tipo ligações de hidrogênio, as espécies estarão mais fortemente empacotadas, acarretando uma menor difusão, aumentando a viscosidade do líquido.

³¹ A. A. Fannin Jr., D. A. Floreani, L. A. King, J. S. Landers, B. J. Piersma, D. J. Stech, R. L. Vaughn, J. S. Wilkes, J. L. Williams, *J. Phys. Chem.* 88, 2614 (1984).

³² P. A. Z. Suarez *Dissertação de mestrado*. UFRGS, Porto Alegre, RS (1996).

1.1.2.c. Aplicações de Líquidos Iônicos como Solventes em Química

Inicialmente desenvolvidos com o objetivo de servirem como eletrólitos para baterias, os líquidos iônicos vêm sendo utilizados com sucesso em várias aplicações. Além de eletrólitos para dispositivos ou processos eletroquímicos, também estão sendo aplicados para reações orgânicas e catalíticas, produção de novos materiais e solvente para processos de extração e separação.

A aplicação de líquidos iônicos em diversas reações continua expandindo. A Tabela 1.6³³ apresenta algumas destas aplicações.

Tabela 1.6: Exemplos de aplicações de líquidos iônicos como solventes para química.

Reação	Líquido Iônico	Catalisador	Referência
Reações de Diels-Alder	[EtNH ₃][NO ₃]	-	34
Síntese organometálica de complexos de ferro	[BMI][Cl]/AlCl ₃	[BMI][HC ₂] como fonte de H ⁺	35
Síntese de complexos de metais de transição	[BMI][Cl]/AlCl ₃	-	36
Hidrogenação de olefinas e diolefinas	[BMI][BF ₄], [BMI][PF ₆]	Pd(acac) ₂	37
Esterificação	[BMI][BF ₄]	PdCl ₂ (PhCN ₂), (+)-NMDPP/TsOH	38
Dimerização de olefinas	[BMI][Cl]/EtAlCl ₂	WC ₆ com anilina/EtAlCl ₂	39
Acilação de Friedel-Crafts	[EMI][I]/AlCl ₃	-	40
Síntese de Z-aspartame	[BMI][PF ₆]	Enzima	41

³³ H. Olivier-Bourbigou, L. Magna, *J. Mol. Catal. A: Chem.* 3484, 1-19, (2002)

³⁴ D. A. Jaeger, C. E. Tucker, *Tetrahedron Lett.* 30, 1785 (1989).

³⁵ P. J. Dyson, M. C. Grossel, N. Srinivasan, T. Vine, T. Welton, D. J. Williams, A. J. P. White, T. Zigras, *J. Chem. Soc., Dalton Trans.* 3465 (1997).

³⁶ D. Crofts, P. J. Dyson, K. M. Sanderson, N. Srinivasan, T. Welton, *J. Organomet. Chem.* 573, 292 (1999).

³⁷ J. Dupont, P. A. Z. Suarez, A. P. Umpierre, R. F. De Souza, *J. Braz. Chem. Soc.* 11, 293 (2000).

³⁸ D. Zim, R. F. De Souza, J. Dupont, A. Monteiro, *Tetrahedron Lett.*, 39, 7071 (1998).

³⁹ H. Olivier, P. Gérot-Laurent, *J. Mol. Catal.* 148, 43 (1999).

⁴⁰ A. Stark, B. L. MacLean, R. D. Singer, *J. Chem. Soc., Dalton Trans.* 63 (1999); J. K. D. Surette, L. Green, R. D. Singer, *Chem. Commun.* 1996, 2753 (1996).

⁴¹ M. Erbedinger, A. Mesiano, A. Russell, *Biotechnol. Prog.* 16, 1129 (2000).

Do ponto de vista químico, o potencial benéfico na utilização de líquidos iônicos encontra-se no aumento das velocidades das reações e da regioselectividade, quando comparadas a outros solventes orgânicos. No sentido prático e de custo, o uso destes líquidos pode beneficiar trazendo melhores custos para a etapa de separação e de recuperação.

1.2. Redes Neurais Artificiais

1.2.1. Fundamentos

Redes Neurais Artificiais – RNAs (*Artificial Neural Networks* – ANNs ou ainda *Computational Neural Networks* - CNN) são cada vez mais utilizadas nesta última década. Isto se deve, principalmente, por serem uma ferramenta poderosa no tratamento de problemas complexos e não lineares, sem a necessidade de um modelo explícito para as relações entre as entradas e saídas do processo. Vários problemas em química têm utilizado com sucesso a flexibilidade e desempenho oferecidos por RNAs⁴².

RNAs constituem uma técnica baseada em inteligência artificial (AI - *Artificial Intelligence*), associada ao desenvolvimento de sistemas computacionais inteligentes, os quais exibem características associadas com o comportamento humano⁴³. Estas características fornecem às redes neurais capacidade de adaptação, analogamente ao aspecto criativo do cérebro humano, sem a necessidade de uma formulação matemática explícita entre as entradas e saídas dos dados.

Uma rede neural artificial é formada por uma rede de neurônios ou elementos de processamento interconectados (Figura 1.5), que operam paralelamente, processando informações a partir de entradas, para a obtenção de uma resposta dinâmica. A rede então armazena conhecimentos e os torna disponível para uso. Assim como no cérebro, são possíveis diversas interconexões, tais como, entre os neurônios de camadas adjacentes, dentro da própria camada ou conexão de um neurônio para ele mesmo (a topologia da rede será discutida no capítulo de metodologia). No caso da representação abaixo, a rede neural é chamada de *feedforward*. As conexões ocorrem apenas intercamada e para “frente” (em direção a camada de saída).

⁴² J. Zupan, J. Gasteiger, *Neural Networks for chemist: An introduction*. New York:VCH (1993); J. A. Burns, G. M. Whitesides, *Chem. Rev.* 93, 2583-2601 (1994).

⁴³ D. R. Baughman, Y. A. Liu. *Neural Networks in Bioprocessing and Chemical Engineering*. Academic Press, Inc. (1995).

A relação entre a entrada e a saída da rede é desenvolvida na fase denominada de aprendizagem ou de fase de treinamento. O método de treinamento da rede pode ser classificado em dois tipos: supervisionado e não supervisionado. Sendo que os métodos supervisionados são os mais utilizados. No aprendizado supervisionado, as entradas e saídas são fornecidas a rede neural, definindo qual o comportamento deverá desenvolver. Ou seja, apresentam-se os padrões de entrada e saída, e a rede ajusta seus pesos e *biases* de modo a atingirem as saídas desejadas. Nos métodos não supervisionados, a rede aprende de forma autônoma. O aprendizado supervisionado é também chamado de aprendizado por correção de erro.

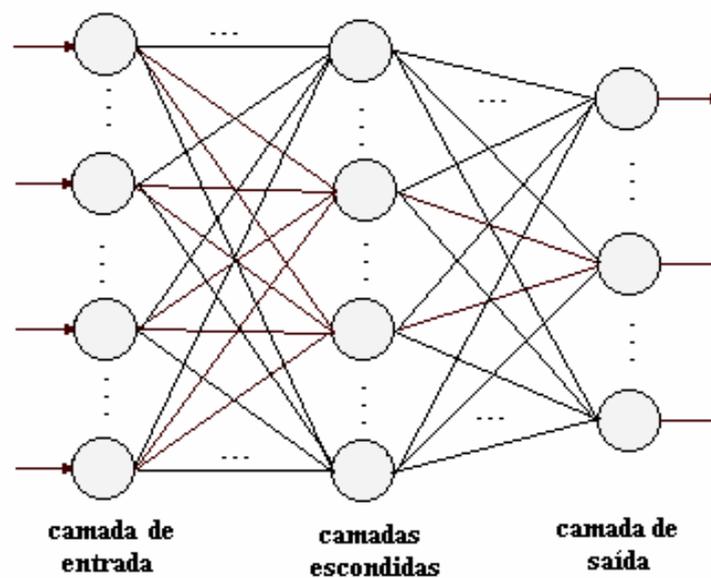


Figura 1.5: Rede Neural *feedforward*.

O método de treinamento está diretamente ligado a topologia da rede escolhida e já que os aprendizados supervisionados são os mais utilizados, também as redes mais conhecidas usam este tipo de aprendizado, entre elas, as redes *feedforward*: *perceptron*, *adaline*, *madaline* e *backpropagation* (BPN). A rede *backpropagation* será utilizada neste trabalho.

Redes neurais têm propriedades que fornecem vantagens em relação a outras técnicas computacionais, entre elas:

1. A informação é distribuída em todos os neurônios da rede. Esta distribuição fornece maior flexibilidade em relação às técnicas de processamento simbólico (sistemas especialistas), onde a informação tem localização fixa.

2. Capacidade de aprender. Ocorrendo um erro ou uma nova situação que produza resultados falsos, pode-se usar uma técnica de treinamento para corrigir, ajustando os pesos dos sinais emitidos pelos neurônios até que o erro na saída desapareça.

3. Melhor adaptação para processamento com dados ruidosos, incompletos ou inconsistentes. Cada neurônio é responsável apenas por um microaspecto da relação entre a entrada e a saída. O conceito de microaspecto implica que cada neurônio afeta cada padrão entrada-saída apenas suavemente, minimizando os efeitos de dados ruidosos ou incompletos em um dado neurônio. Apenas quando são reunidos todos os neurônios na rede, estes microaspectos traçam o padrão macroscópico de entrada-saída.

1.2.2. Aplicações de Redes Neurais Artificiais em Química

Apesar de existirem vários outros métodos, principalmente técnicas estatísticas e métodos estatísticos multivariados, que podem modelar problemas em química, as RNAs constituem mais uma ferramenta de escolha, que pode ser complementar ou essencial para o estudo de alguns destes problemas.

As aplicações incluem a obtenção de relações estrutura-atividade e estrutura-propriedades, predição de reatividade, determinação da estrutura de proteínas, controle de processos, reconhecimento de padrões, classificação de espectros moleculares e análise de dados⁴⁴. Alguns exemplos de aplicações de RNAs em química, são: reconhecimento de padrões (espectro de infravermelho e espectrometria de massas)⁴⁵, modelagem de dados espectrais usando infravermelho próximo ou UV-VIS⁴⁶, determinação da estrutura de proteínas⁴⁷, controle de processos⁴⁸, predição de reatividade química⁴⁹, simulação computacional de dinâmica molecular e Monte Carlo⁵⁰.

Tanabe *et al.* (1992)⁴⁵ usaram RNAs para armazenar 1000 espectros de infravermelho dentro de 40 categorias e em seguida treinou uma segunda RNA para predizer o tipo de molécula para o espectro original. Outra utilização de RNAs para reconhecimento de padrões foi realizada por Curry & Rumelhart, que usaram RNAs com algoritmo de

⁴⁴ B. G. Sumpter, C. Getino, D. W. Noid, *Annu. Rev. Phys. Chem.* 45, 439-481 (1994).

⁴⁵ K. Tanabe, T. Tamura, H. Uesaka, *Appl. Spectrosc.* 46, 807-810 (1992); B. Curry, D. E. Rumelhart, *Tetrahedron Comput. Methodol.* 3, 213-237 (1990).

⁴⁶ J. R. Long, V. G. Gregoriou, P. J. Gemperline, *Anal. Chem.* 62, 1791-1797 (1990).

⁴⁷ N. Qian, T. J. Sejnowski, *J. Mol. Biol.* 202, 865-884 (1988); S. Pascarella, F. Bossa, *Comput. Appl. Biosci.* 5, 319-320 (1989); L. H. Holley, M. Karplus, *Methods Enzymol.* 202, 204-224 (1991).

⁴⁸ V. M. L. Santos, F. R. Carvalho, M. B. De Souza Jr. *Braz. J. Chem. Engineer.* 17, 897-905 (2000).

⁴⁹ V. Simon, J. Gasteiger, J. Zupan, *J. Am. Chem. Soc.* 115, 9148-9159 (1993); V. Kvasnicka, J. Pospichal, *Theochem* 81, 227-242 (1991); V. Kvasnicka, S. Sklenak, J. Pospichal, *J. Am. Chem. Soc.* 115, 1495-1500 (1993).

⁵⁰ D. W. Noid, J. A. Darsey, *Comput. Polym. Sci.* 1, 157-160 (1991); J. A. Darsey, A. Soman, D. W. Noid, *Makromol. Chem. Theory Simul.* 2, 711-719 (1993).

aprendizagem *backpropagation* para relacionar espectros de massas às estruturas para um conjunto de 493 dados provenientes de espectrometria de massa.

Long *et al.*⁴⁶ realizaram a modelagem baseada em RNAs de dados espectrais produzindo estimativas das concentrações de misturas com dois componentes usando espectroscopia de infravermelho próximo ou UV-VIS. Eles concluíram que o modelo de RNAs revelou vantagens sobre um método de regressão de componentes principais. De um modo geral, encontrou-se que os modelos de RNAs são mais exatos do que mínimos quadrados parciais ou análise de componentes principais.

Qian & Sejnowski⁴⁷ foram os pioneiros na utilização de RNAs para predição de estruturas secundárias de proteínas.

O interesse por RNAs vem crescendo tanto no meio acadêmico como no meio industrial. As aplicações incluem o monitoramento e controle de processos químicos⁵¹ e reações, análise de composições químicas, reconhecimento de padrões para predição de propriedades de materiais, caracterização e modelagem de processos em reatores químicos, projeto de produtos e a classificação de misturas multicomponentes. A capacidade das RNAs para determinar as correlações de interesse que descrevem um processo particular e em seguida usá-los para fazer extrapolações e principalmente interpolações, torna as RNAs uma poderosa ferramenta para modelagem de fenômenos químicos.

As RNAs foram aplicadas com sucesso na predição de reatividade química. Um importante avanço em química está na capacidade de determinar se um processo em particular passará por uma dada reação química, quais produtos serão formados e seu rendimento. RNAs tem a capacidade de suprir este desejo, fazendo uma análise rápida destas muitas possibilidades.

Noid & Darsey (1991)⁵⁰ usaram RNAs para determinar relações entre as propriedades de uma superfície de energia potencial e as posições relativas dos núcleos atômicos. Naquele estudo, as variáveis (posições) foram passadas por resfriamento randômico como descrito por dinâmica molecular. RNAs foram usadas para determinar a informação necessária para fazer predições da taxa na qual o processo de resfriamento ocorreu, como função da posição e da temperatura.

Darsey *et al.* (1991)⁵² usaram RNAs para obter a relação entre a capacidade calorífica e a temperatura para vários polímeros. Os resultados permitiram a determinação desta medida em baixas temperaturas, cuja obtenção experimental é difícil.

⁵¹ V. M. L. Santos, F. R. Carvalho, M. B. De Souza Jr., *Braz. J. Chem. Engineer.* 17, 897-905 (2000).

⁵² J. A. Darsey, D. W. Noid, B. Wunderlich, L. Tsoukalas, *Makromol. Chem. Rapid. Commun.* 12, 325-330 (1991).

1.2.3. Redes Neurais Artificiais em Ajustes de Superfícies de Energia Potencial (SEP)

O potencial de interação inter- e intramolecular é determinante no sucesso de simulações computacionais de processos e fenômenos químicos. Logo, a busca de potenciais mais realísticos e de menor demanda computacional é uma constante nesta área de simulação computacional atomística. Sumpter & Noid (1992)⁵³ obtiveram RNAs capazes de correlacionar informações proveniente de espectros vibracionais com os parâmetros da superfície de energia potencial do polietileno.

Dispõe-se de duas opções para ajuste dos potenciais de interação. Ajuste para reproduzir as propriedades dos líquidos (OPLS) ou o ajuste a partir de cálculos *ab initio*. O modelo TIP5P⁵⁴, por exemplo, foi ajustado para reproduzir as propriedades da água líquida (OPLS), em particular, a temperatura de densidade máxima da água.

Hobday *et al.* (1999)⁵⁵ aplicaram algoritmos genéticos e RNAs para determinar potenciais de interação. Mostrou-se que estas técnicas permitem determinar o mínimo de energia de agregados de átomos, descritos por funções de potencial interatômicos. Também mostrou-se que potenciais de interação com termos de muitos corpos para sistemas multi-componentes podem ser obtidos a partir de RNAs.

Tai No *et al.* (1997)⁵⁶ realizaram a descrição da superfície de energia potencial (SEP) do dímero da água com uma RNA treinada a partir de dados de química quântica *ab initio*. O ajuste obtido foi realizado a partir de 3331 configurações. A arquitetura da RNA final contendo quatro camadas 6:12:12:1 (total de 31 neurônios), com 253 pesos e *biases*. Os erros absolutos médios para energia do dímero obtidos a partir dos conjuntos de treinamento e teste estão apresentados na Tabela 1.7.

⁵³ B.G. Sumpter, D. W. Noid, *Chem. Phys. Lett.* 192, 455-462 (1992).

⁵⁴ M. W. Mahoney & W. L. Jorgensen, *J. Chem. Phys.* 112, 8910 (2000).

⁵⁵ S. Hobday, R. Smith, J. BelBruno, *Nucl. Instr. And Meth. In Phys. Res.* B 153, 247-263 (1999).

⁵⁶ K. Tai No, B. H. Chang, S. Y. Kim, M. S. Jhon, H. A. Scheraga, *Chem. Phys. Lett.* 271, 152-156 (1997).

Tabela 1.7: Erro absoluto médio na energia do dímero para os conjuntos de treinamento e teste.

Intervalo de energia do dímero (kcal mol ⁻¹)	Conjunto de treinamento		Conjunto de teste	
	Nº Configurações	Erro absoluto ^a	Nº Configurações	Erro absoluto
< -5	33	0,042	5	0,059
-5 a -4	243	0,060	16	0,092
-4 a -3	480	0,106	30	0,156
-3 a -2	590	0,134	42	0,170
-2 a -1	507	0,165	33	0,115
-1 a 0	786	0,134	48	0,146
0 a 10	420	0,408	23	0,482
10 a 40	39	2,066	2	0,556
40	33	5,904	1	1,445
Total	3131	0,250 ^b	200	0,190 ^b

$$^a \text{ Erro absoluto médio do } j\text{th intervalo: } erro = \frac{1}{n_j} \sum_{i=1}^{n_j} |V_i^{ab} - V_i^{NN}|$$

$$^b \text{ Erro absoluto médio total : } erro\ total = \frac{1}{n} \sum_{i=1}^n |V_i^{ab} - V_i^{NN}|, \text{ onde } n_j \text{ e } n \text{ são os número de configurações no } j\text{th intervalo e no conjunto total, respectivamente, dos conjuntos de treinamento e teste.}$$

Prudente & Soares Neto (1998)⁵⁷ utilizaram uma RNA do tipo perceptron multicamada com algoritmo de aprendizagem *backpropagation* para ajustar as funções de momento de dipolo e superfície de energia potencial de HCl⁺ e H₃. Os resultados indicaram que a rede neural pode descrever as funções de momento de dipolo e superfície de energia potencial para o caso unidimensional (HCl⁺) e multidimensional (H₃).

Neste trabalho RNAs serão utilizadas para representação das SEPs para a água e líquidos iônicos. Além disso, um dos aspectos originais desse trabalho consistiu, pela primeira vez, na utilização de RNAs para o cálculo de potencial de pares em simulações de Monte Carlo da água líquida.

⁵⁷ F. V. Prudente, J. J. Soares Neto, *Chem. Phys. Lett.* 287, 585-589 (1998).

1.3. Análise Topológica de Redes de Ligações de Hidrogênio

A análise topológica de modelos de redes e de redes reais mostrou, em alguns casos, a existência de padrões de *small-world* (SW), que consistiu na existência de pequenos agregados. Esta verificação mostra que em redes complexas com diversas conexões entre os nós ou sítios (altamente agregadas), podem existir distâncias ou caminhos mais curtos entre dois sítios na rede, que diminuem o comprimento médio da rede e caracteriza a existência de pequenos mundos ou SW. Esta análise topológica teve um grande impulso com os estudos de Watts e Strogatz⁵⁸ e vem sendo utilizadas diferentes redes, como: topologia da internet, redes neurais, circuitos eletrônicos, rede de alimentos, ecossistemas, redes de interação social, redes de colaboração científica, entre outras, auxiliando na elucidação de seus aspectos dinâmicos⁵⁹.

Devido a este aspecto de universalidade⁶⁰, a análise topológica e os padrões de SW podem gerar informações sobre propriedades de interesse de redes complexas, como informações sobre sua robustez e comportamento com respeito as interferências ou flutuações externas. Para este estudo, por exemplo, a estabilidade das redes de ligações de hidrogênio com respeito às mudanças na pressão e temperatura da água, podem levar a novas informações e perspectivas sobre as relações entre as propriedades da água e as rede de ligações de hidrogênio.

Neste trabalho, caracterizamos a presença de padrões de small-world em água supercrítica através do cálculo de propriedades locais: coeficiente de agregação (*clustering*), comprimento do caminho e grau de distribuição e também da propriedade global dada pelo espectro do gráfico de conexões.

⁵⁸ D. J. Watts, S. H. Strogatz, *Nature* 393, 440-442 (1998).

⁵⁹ R. Albert, H. Jeong, A.L. Barabási, *Nature* 401, 130 (1999); M. Marchiori, V. Latora, *Physica A* 285, 539-546 (2000); R. F. Cancho, C. Janssen, R. V. Solé, *Phys. Rev. E* 64, 046119 (2001); J. M. Montoya, R. V. Solé, *J. Theor. Biol.* 214, 405-412 (2002).

⁶⁰ R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* 74, 47 (2002).

1.4. Objetivos

1.4.1. Objetivos Gerais

- 1) Determinação de propriedades estruturais e termodinâmicas por simulação computacional de fluidos e soluções em condições supercríticas;
- 2) Determinação de propriedades estruturais e termodinâmicas de líquidos iônicos por simulação computacional;
- 3) Análise topológica de líquidos usando estatística de ilhas e estudo da presença de padrões de *small world*;
- 4) Desenvolvimento de novos métodos para ajuste de potenciais intermoleculares.

1.4.2. Objetivos Específicos

- 1) Simulação computacional de Monte Carlo (MC) da água, em condições ambiente e críticas, usando os modelos TIP4P e TIP5P. Análise topológica das redes de ligações de hidrogênio em condições ambiente e extremas.
- 2) Simulação computacional de Monte Carlo (MC) de soluções água-metanol em condições ambiente e críticas;
- 3) Simulação computacional de Monte Carlo (MC) de líquidos iônicos com o cátion 1-*n*-butil-3-metilimidazólio (BMI) e ânions Cl⁻ e BF₄⁻;
- 4) Ajuste de superfícies de energia potencial entre moléculas de água com redes neurais artificiais, a partir de dados gerados pelas simulações de Monte Carlo usando o modelo TIP5P;
- 5) Inclusão de modelos baseados em redes neurais artificiais em simulações de Monte Carlo;
- 6) Ajuste de superfícies de energia potencial para líquidos iônicos usando redes neurais artificiais, a partir de dados *ab initio*.

CAPÍTULO 2

METODOLOGIA

2.1. Introdução

Neste Capítulo, serão abordados os métodos de simulação computacional (método de Monte Carlo, modelos de potencial intermolecular, técnicas de parametrização de funções de potencial), métodos de química quântica *ab initio*, métodos para pré-processamento de dados, redes neurais artificiais (RNAs) e a formulação matemática necessária para a análise topológica de redes. Contudo, iniciamos este capítulo com uma descrição sucinta dos métodos experimentais disponíveis na literatura para a determinação das propriedades estruturais da água.

2.2. Métodos Experimentais e as Propriedades Estruturais da Água

A estrutura da água, descrita em termos das correlações existentes entre os pares de átomos O-H, O-O e H-H, denominadas de funções de distribuição radial ou RDFs (*Radial Distribution functions*): $g_{OO}(r)$, $g_{OH}(r)$, $g_{HH}(r)$, pode ser estudada através de difração de raios-X, ($g_{OO}(r)$) ou de nêutrons com substituição isotópica (NDIS – *neutron diffraction with isotopic substitution*).

As funções de distribuição de pares, $g_{\alpha\beta}(r)$, fornecem a probabilidade de encontrar uma partícula β em uma camada esférica de espessura “dr” à uma distância “r” da partícula α , cuja posição é considerada a origem, relativamente à probabilidade esperada em uma distribuição completamente aleatória (ideal) na mesma densidade. A partir desta definição, alterações na densidade local nas imediações das moléculas induzem a valores de $g(r)$ diferentes da unidade, valor para o qual tende a distribuição quando “r” tende para infinito⁶¹.

No caso da determinação das RDFs por difração de nêutrons os experimentos envolveram três composições isotópicas diferentes da água (H₂O) e água pesada (D₂O): (H₂O)_{pura}, (H₂O)_{0,5}(D₂O)_{0,5} e (H₂O)_{0,003}(D₂O)_{0,997}⁶². Cada uma destas composições fornece informações estruturais como uma combinação linear dos fatores de estrutura sítio-sítio no espaço-k de Fourier. A conversão dos resultados para o espaço real, para a obtenção das

⁶¹ M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*. Clarendon Press. Oxford. (1994).

funções de correlação estão sujeitas a incertezas. Deste modo, apesar destas técnicas terem os princípios bem definidos, há importantes dificuldades experimentais intrínsecas, que podem gerar erros nas RDFs. Neste caso, a utilização de simulação computacional foi decisiva para a reavaliação dos resultados experimentais⁶³, levando em conta os efeitos e correções do espalhamento inelástico. Estes resultados reavaliados por Soper *et al.* concordaram muito bem com as simulações computacionais⁶⁴.

A determinação das RDFs de líquidos e soluções em condições de alta pressão e temperatura por métodos de difração apresenta várias dificuldades técnicas quando comparada com as condições ambientes, como por exemplo, o fraco espalhamento devido ao volume reduzido de amostra. Contudo, recentes avanços permitem realizar experimentos de difração de raios-X sob pressões de até 10,0 GPa e temperaturas de até 1000°C⁶⁵. Entretanto, para a água, a intensidade de espalhamento dos raios-X é tão fraca que não permite análise com confiabilidade razoável. Os resultados já obtidos com esta técnica são para condições ambiente e para as funções de distribuição radial oxigênio – oxigênio.

2.3. Simulação Computacional

A simulação computacional permite obter propriedades macroscópicas (coeficiente de transporte, equação de estado, energia interna, capacidade calorífica, etc.) a partir de informações microscópicas, como as posições e momentos dos átomos e moléculas e das interações entre os átomos e moléculas.

A simulação computacional consiste em gerar as configurações microscópicas necessárias para, após o tratamento da mecânica estatística, fornecer propriedades macroscópicas.

Para o estudo da estrutura da água líquida os métodos mais utilizados para a simulação computacional são: dinâmica molecular (MD – *Molecular Dynamics*)⁶⁶ e Monte Carlo (MC)⁶⁷. Em ambos os casos, cria-se uma caixa de simulação, geralmente cúbica, com condições de contorno periódicas, contendo entre 10^2 e 10^4 moléculas. O tamanho do

⁶² P. Postorino, R. H. Tromp, M-A. Ricci, A. K. Soper, G. W. Nelison, *Nature* 366, 668-670 (1993).

⁶³ A. K. Soper, M. G. Phillips, *Chem. Phys.* 107, 47-60 (1986); A. K. Soper, C. Andreani, M. Nardone, *Phys. Rev. E* 47, 2598-2605 (1993); A. K. Soper, *J. Chem. Phys.* 101, 6888-6901 (1994).

⁶⁴ A. K. Soper, F. Bruni, M. A. Ricci, *J. Chem. Phys.* 106, 247-254 (1997).

⁶⁵ H. Ohtaki, T. Radnai, T. Yamaguchi, *Chem. Soc. Rev.* 41-51 (1997).

⁶⁶ E. S. Fois, M. Sprik, M. Parrinello, *Chem. Phys. Lett.* 223, 411-415 (1994); A. G. Kalinichev, S. V. Churakov, *Chem. Phys. Lett.* 302, 411-417 (1999).

sistema é limitado principalmente pelos recursos computacional disponíveis, principalmente, tempo de UCP. A energia do sistema (intramolecular + intermolecular) é, em geral, calculada partir de funções de potencial ou campos de força⁶⁸. A escolha criteriosa de um modelo do potencial de interação é determinante para o sucesso destas simulações. Neste trabalho, as interações serão descritas somente a partir dos potenciais intermoleculares, visto que as propriedades de interesse sofrem poucos efeitos dos graus de liberdade intramolecular. O método de Monte Carlo foi adotado neste trabalho e por isso será detalhado a seguir.

Os métodos MD e MC diferenciam-se pela forma de exploração do espaço de fase, contudo, devem tender ao mesmos resultados quando o tempo de simulação MD for muito longo e número de configurações MC muito grande. A escolha de uma ou outra metodologia depende das propriedades de interesse e do sistema a simular. Por exemplo, quando se deseja calcular propriedades dependentes do tempo, como coeficientes de transporte, opta-se pelo método MD. Para investigar sistemas em equilíbrio, com pressão e temperatura muito bem definidas, opta-se pelo método MC.

No *ensemble* NpT, adotado nas simulações realizadas neste trabalho, o número de moléculas (N), a temperatura (T) e a pressão (p) são mantidas fixas, enquanto permite-se variar o volume da caixa de simulação. Este *ensemble* é particularmente apropriado para simulação de misturas, visto que os dados experimentais são freqüentemente obtidos a pressão constante e as teorias de misturas são freqüentemente formuladas com esta suposição.

2.3.1. Método de Monte Carlo (MC)

O método de Monte Carlo foi desenvolvido por Neumann, Ulan e Metropolis no final da segunda guerra mundial para o estudo da difusão de nêutrons⁶⁹. O nome do método é oriundo do uso extensivo de números aleatórios, ao qual associa-se os famosos cassinos da cidade de Monte Carlo, no principado de Mônaco. O método de Monte Carlo foi a técnica utilizada para executar a primeira simulação computacional de um sistema molecular⁶⁸.

O principal objetivo do método é realizar uma amostragem eficiente do espaço configuracional, que para o caso de líquidos consiste em:

⁶⁷ M. W. Mahoney, W. L. Jorgensen, *J. Chem. Phys.* 112, 8910-8922 (2000); X. Wu, Y. Li, J. Lu, T. Teng, *Fluid Phase Equilibria*, 77, 139-156 (1992); J. M. M. Cordeiro, *Quim. Nova* 21, 698-701 (1998).

⁶⁸ M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*. Clarendon Press. Oxford. (1994).

⁶⁹ N. Metropolis, S. Ulam, *J. Am. Stat. Ass.* 44, 335-341 (1949).

1. Preparar uma dada configuração inicial “ \vec{r}_a ”, que tenha um fator de Boltzmann $e^{-bU(\vec{r}_a)}$ não nulo, onde $b = 1/kT$.
2. Escolher (aleatoriamente ou não) uma molécula da configuração inicial e realizar um movimento (rotação ou translação) aleatório, gerando-se uma nova configuração “ \vec{r}_n ”, com $e^{-bU(\vec{r}_n)}$.
3. Aceitar ou rejeitar esta nova configuração de acordo com o algoritmo de Metropolis, isto é, se $U(\vec{r}_n) > U(\vec{r}_a)$, então o seguinte fator $e^{-[U(\vec{r}_n)-U(\vec{r}_a)]/k_B T}$ é calculado e comparado com um número aleatório $\xi \in [0, 1]$: se for maior que ξ , a nova configuração “ \vec{r}_n ” é aceita e se for menor, a nova configuração é rejeitada, ou se $U(\vec{r}_n) < U(\vec{r}_a)$ então a nova configuração é aceita.

Onde “ \vec{r}_a ” e “ \vec{r}_n ” referem-se as coordenadas das moléculas na configuração inicial e nas novas configurações geradas aleatoriamente, respectivamente. $U(\vec{r}_n)$ representa a energia potencial da configuração n .

Uma das vantagens do método MC é a facilidade para adaptação para simulações com vários *ensembles* de interesse, particularmente NVT (canônico), NpT (isobárico-isotérmico) e μ VT (gran-canônico).

Simulações de Monte Carlo usando *ensemble* NpT são de interesse, pois em geral, as medidas experimentais são realizadas a pressão constante. Para simulação usando este *ensemble* é necessário efetuar mudanças no volume para manter a pressão constante. A simulação é realizada combinando deslocamentos aleatórios das partículas com mudanças de volume da caixa de simulação. Logo, um novo estado é gerado por rotação ou translação aleatória de uma molécula e/ou por uma mudança de volume de V^a para V^n . Ambas as variações são determinadas pelas relações abaixo⁷⁰:

$$\vec{s}_i^n = \vec{s}_i^a + d\mathbf{s}_{max} (2\xi - \vec{1})$$

$$\vec{V}_i^n = \vec{V}_i^a + d\mathbf{V}_{max} (2\xi - \vec{1})$$

onde ξ é um número aleatório entre 0 e 1 e $\vec{1}$ é o vetor (1, 1, 1). $d\mathbf{s}_{max}$ e $d\mathbf{V}_{max}$ correspondem as mudanças máximas nas coordenadas das partículas e no volume da caixa de simulação, respectivamente. Estes dependem do estado (N, p e T) simulado e devem ser ajustados para manter a taxa de aceitação entre 35 e 50%⁷⁰.

⁷⁰ A. R. Leach, *Molecular Modelling. Principles and Applications*. Longman. England. (1998).

Como o volume é mudado, então todas as energias de interação do sistema devem ser recalculadas e a diferença de energia entre a configuração nova “ n ” e a anterior “ a ”, após a variação de volume, é calculada pela relação⁷¹:

$$dH_{na} = dU_{na} + p(V_n - V_a) - Nb^{-1} \ln\left(\frac{V_n}{V_a}\right)$$

Este valor é aproximadamente igual a mudança de entalpia envolvida na mudança do estado “ a ” para o estado “ n ”.

Os movimentos e/ou mudanças de volume serão aceitos ou rejeitados segundo os critérios descritos no item 3 (substituindo $U(r_n) - U(r_a)$ pelo dH_{na}) do método de Métropolis descrito acima.

Mudanças de volume requerem maior demanda computacional do que rotações e translações em uma dada molécula, porque quando uma molécula é movimentada necessita-se calcular a energia de interação desta com as demais, ou seja, $2(N-1)$ cálculos do potencial de pares, ao passo que uma mudança de volume implica no cálculo de todas as $(1/2)N(N-1)$ interações. Entretanto a variação de volume tem uma frequência muito pequena quando comparada com a aquela na qual as partículas são deslocadas.

2.3.2. Determinação das Propriedades Termodinâmicas e Estruturais por Simulações de Monte Carlo

Propriedades termodinâmicas, como entalpia de vaporização e energia interna são obtidas a partir de médias sobre as configurações. Outras propriedades são determinadas a partir de médias sobre flutuações cruzadas, como por exemplo, as capacidades caloríficas a pressão e volumes constantes: $C_p = (dH/dT)_p$ e $C_v = (dE/dT)_v$, o coeficiente de expansão térmica: $\alpha_p = V^{-1}(dV/dT)_p$, a compressibilidade isotérmica: $\beta_T = -V^{-1}(dV/dP)_T$, etc.⁷¹

A determinação de propriedades através das flutuações implica num grande número de configurações e logo, alta demanda computacional. Para contornar este problema, pode-se optar pela obtenção da propriedade através de diferenças finitas que, por exemplo, para a capacidade calorífica à pressão constante, é dada por:

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p = \left(\frac{\partial(E + pV)}{\partial T}\right)_p = \left(\frac{\partial E}{\partial T}\right)_p + p\left(\frac{\partial V}{\partial T}\right)_p \approx \frac{\Delta E}{\Delta T} + p\frac{\Delta V}{\Delta T} \quad (2.1)$$

⁷¹ M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*. Clarendon Press. Oxford. (1994).

Neste caso basta executar a simulação a temperaturas acima e abaixo da condição desejada.

2.3.3. Modelos de Potencial Intermolecular da Água

Existem inúmero modelos para a descrição das interações entre as moléculas de água. Estes modelos são, em geral, agrupados em três categorias. No modelo de sítios de interação simples cada molécula de água é mantida com uma geometria fixa e a interação entre as moléculas é descrita usando as expressões de Lennard-Jones e coulombianas. Modelos flexíveis que permitem mudanças na estrutura da molécula, e modelos que incluem os efeitos de polarização e de muitos corpos explicitamente.

Os modelos de potencial intermolecular mais utilizados para a água são: o SPC (*Simple Point Charge*) e o TIP3P (*Transferable intermolecular potential 3-polarization*) com três sítios, o TIP4P, no qual a água é descrita por quatro sítios. Estes modelos utilizam cargas parciais distribuídas numa geometria rígida para descrição das contribuições eletrostáticas e parâmetros de *Lennard-Jones* para descrever as interações de van der Waals e repulsivas de curto alcance.

Antes de 1983, cálculos MD e MC para a água líquida geralmente usavam as funções de potencial do tipo BNS, MCY e ST2⁷². A partir 1983 foram introduzidos os modelos TIP3P⁷³ e TIP4P⁷⁴, que até 2000 eram os mais utilizados juntamente com os modelos SCP⁷⁵ e SPC/E⁷⁶. A partir de 2000 surgiu o modelo TIP5P⁷² que conseguiu reproduzir a temperatura de densidade máxima da água.

Neste trabalho utilizou-se os modelos TIP4P e TIP5P. Suas geometrias são ilustradas nas Figuras 2.1 e 2.2 e seus parâmetros estão apresentados na Tabela 2.1. A energia de interação entre duas moléculas de água, a e b , é expressa como:

$$E_{ab} = \sum_{i \in a} \sum_{j \in b} \frac{q_i q_j e^2}{4\pi \epsilon_0 r_{ij}} + 4e_o \left[\left(\frac{\mathbf{s}_o}{r_{o-o}} \right)^{12} - \left(\frac{\mathbf{s}_o}{r_{o-o}} \right)^6 \right] \quad (2.2)$$

onde i e j são os sítios em a e b , respectivamente, e r_{o-o} é a distância entre os átomos de oxigênio, e $4\pi\epsilon_0$ é a constante elétrica.

⁷² M. W. Mahoney, W. L. Jorgensen, *J. Chem. Phys.* 112, 8910-8922 (2000).

⁷³ W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* 79, 926 (1983).

⁷⁴ W. L. Jorgensen, J. D. Madura, *Mol. Phys.* 56, 1381 (1985).

⁷⁵ H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, *Intermolecular Forces* (1981).

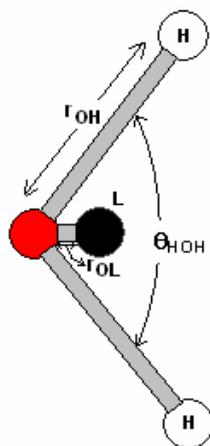


Figura 2.1: Geometria do monômero da água pelo Modelo TIP4P.

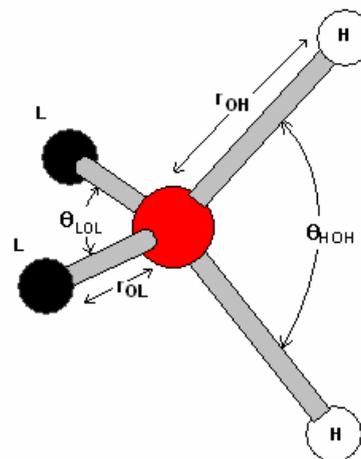


Figura 2.2: Geometria do monômero da água pelo Modelo TIP5P.

Tabela 2.1: Parâmetros e geometria do monômero para as funções de potencial dos modelos TIP4P e TIP5P.

PARÂMETROS	TIP4P (Figura 2.1)	TIP5P (Figura 2.2)
q_H (e)	0,52	0,241
σ_O (Å)	3,15365	3,12
ϵ_O (kcal/mol)	0,1550	0,16
r_{O-L} (Å)	0,15	0,7
r_{O-H} (Å)	0,9572	0,9572
q_L (e)	-1,04	-0,241
θ_{HOH} ($^\circ$)	104,52	104,52
θ_{LOL} ($^\circ$)	-	109,47

Estes parâmetros são denominados de OPLS (*Optimized Parameters for Liquid Simulations*), pois são obtidos através da reprodução de propriedades experimentais do líquido (densidade, entalpia de vaporização, etc.) por simulações MC.

⁷⁶ H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* 91, 6269 (1987).

2.3.4. Modelos de Potencial Intermolecular do Metanol

A obtenção das propriedades do metanol usando simulação computacional foi realizada usando um modelo de três sítios desenvolvido por Jorgensen *et al.* (1983)⁷⁷, que ajustaram os parâmetros às propriedades do metanol líquido (OPLS). Neste modelo, o grupo CH₃ é representado por um único sítio, modelo de átomos unidos, que continuaremos denominando de sítio CH₃. Os parâmetros do modelo estão apresentados na Tabela 2.2.

Tabela 2.2: Parâmetros e geometria das funções de potencial para o modelo do metanol.

Sítios	σ_O (Å)	ϵ_O (kcal/mol)	q_H (e)
O	3,070	0,170	-0,700
CH ₃	3,775	0,207	0,265
H	0,000	0,000	0,435
Geometria	$r(O-H)$ (Å)	$r(O-CH_3)$ (Å)	\angle_{CH_3-O-H} (°)
	0,945	1,4298	108,498

2.3.5. Potenciais de Interação – Tipos de Parametrização

A parametrização dos potenciais de interação pode ser realizada utilizando dados experimentais (OPLS) ou dados *ab initio* ou combinações destes. A parametrização do tipo OPLS é realizada através do cálculo de várias propriedades usando simulações MC e então modifica-se os parâmetros até que se tenha uma concordância aceitável entre os resultados experimentais e as simulações. As propriedades termodinâmicas comumente usadas são: densidade, entalpia de vaporização, funções de distribuição radial, capacidade calorífica. No caso da parametrização dos modelos da água, observou-se que a densidade e a entalpia de vaporização são razoavelmente bem preditas pelos modelos desenvolvidos, mas com significantes variações em outras propriedades, como a constante dielétrica.

A parametrização do tipo *ab initio* envolve cálculos de química quântica *ab initio* para pares e pequenos agregados moleculares, gerando energias que são usadas para parametrização dos potenciais. Um exemplo é o modelo de Nieser, Corongiu e Clementi (NCC) para a água, que combina um potencial de duas moléculas com um termo de

polarização⁷⁸. O modelo NCC foi parametrizado ajustando as energias e outras propriedades a partir das configurações de 250 trímeros e 350 dímeros determinadas com funções de base extendidas. Os dados dos trímeros foram usados para ajustar os termos com parâmetros que descrevem os efeitos de muitos corpos, como as localizações dos dipolos induzidos e cargas pontuais, junto com as polarizabilidades e o valor da carga do hidrogênio.

Neste trabalho estão sendo realizados cálculos de química quântica *ab initio* para a energia potencial que será utilizada no treinamento de redes neurais artificiais e não no ajuste de potenciais analíticos.

2.3.6. Ligações de Hidrogênio entre Moléculas de Água

As moléculas de água em fase líquida formam fortes ligações de hidrogênio, que conferem à água suas propriedades peculiares, por isto, neste estudo foi dada ênfase à determinação do número de ligações de hidrogênio em diferentes condições de temperatura e pressão.

Para o estudo do comportamento das ligações de hidrogênio nas diferentes condições adotou-se dois critérios: geométrico e energético. Os critérios geométricos adotados foram aqueles utilizados por Chialvo & Cummings (1996)⁷⁹, relativo aos valores de R_{O-Y-OX} , \angle_{H1Y-OX} , $\angle_{XX1Y-OX}$ e d_D (H11 refere-se ao hidrogênio 1 da molécula Y (1 ou 2) e XX2Y refere-se ao pseudo átomo 2 da molécula Y (1 ou 2). X (1 ou 2) tem que ser necessariamente diferente de Y), ilustrados na Figura 2.3. As restrições geométricas mencionadas, são as seguintes: $R_{O-O} \leq 3,3 \text{ \AA}$, $\angle_{H1Y-O} \leq 70,53^\circ$, $\angle_{XX1Y-O} \leq 70,53^\circ$, e $d_D \leq 180,0^\circ$. Neste modelo os pseudo-átomos (XX) estão localizados perpendicularmente ao triângulo H-O-H e tem as mesmas dimensões dele em uma disposição tetraédrica.

As restrições nos ângulos foram transformadas para a maior distância possível entre os átomos constituintes da ligação: a molécula doadora de hidrogênios e a receptora (pseudo-átomos ocupando a posição dos orbitais na molécula de água). As restrições obtidas são: $R_{O-O} \leq 3,3 \text{ \AA}$ e $R_{H-XX} \leq 2,74 \text{ \AA}$.

⁷⁷ W. L. Jorgensen, J. Chandrasekhar, J.D. Madura, *J. Chem. Phys.* 79, 926 (1983).

⁷⁸ U. Niesar, G. Corongiu, E. Clementi, G. R. Keller, D. K. Bhattacharya, *J. Phys. Chem.* 94, 7949-7956 (1990).

⁷⁹ A. A. Chialvo, P. T. Cummings, *J. Phys. Chem.* 100, 1309-1316 (1996).

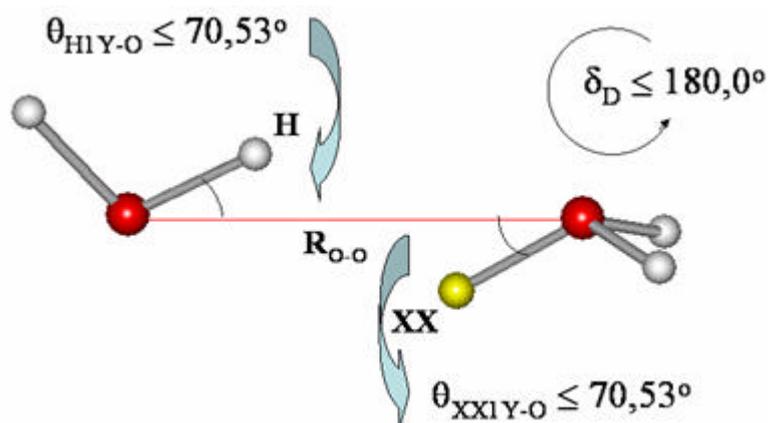


Figura 2.3: Parâmetros geométricos utilizados na determinação de ligações de hidrogênio para o dímero da água.

A existência de ligação de hidrogênio é aceita quando satisfaz os critérios geométricos descritos e a energia calculada é menor que zero ($E_{ab} < 0$).

2.3.7. Ligações de Hidrogênio entre Moléculas de Metanol

Assim como observado para a água, também o metanol é caracterizado pela formação de redes complexas de ligações de hidrogênio. As restrições para a formação de ligações de hidrogênio no metanol puro foram baseadas nas restrições propostas por Pagliai *et al.* (2003)⁸⁰, que recomendam a verificação das seguintes restrições geométricas:

- (1) Distância r (O...H) menor do que 2,6 Å,
- (2) Distância r (O...O) menor do que 3,5 Å,
- (3) Ângulo \angle (HO...O) menor do que 30°.

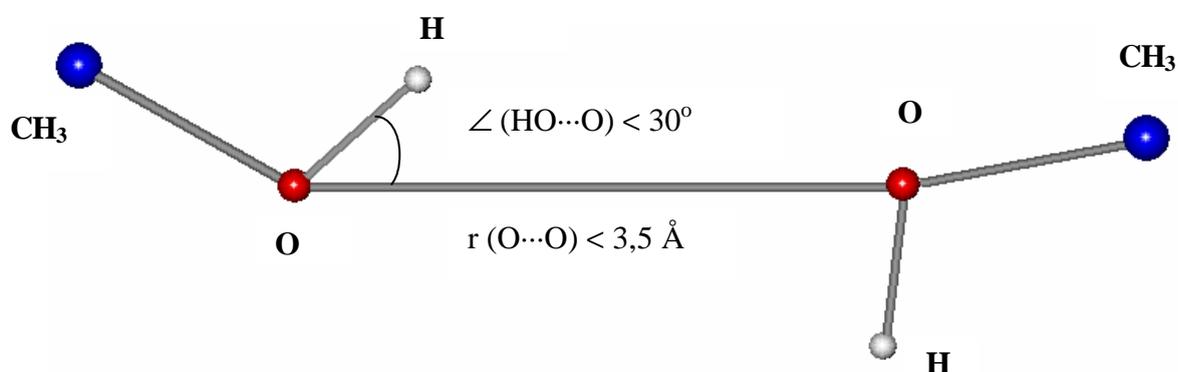


Figura 2.4: Parâmetros geométricos utilizados na determinação de ligações de hidrogênio para o dímero de metanol.

⁸⁰ M. Pagliai, G. Cardini, R. Righini, V. Schettino, *J. Chem. Physics.* 119, 6655-6662 (2003).

A existência de ligação de hidrogênio é aceita quando satisfaz os critérios geométricos descritos e a energia calculada é menor que zero ($E_{ab} < 0$).

2.3.8. Ligações de Hidrogênio entre as Moléculas na Mistura Água:Metanol

As restrições para a formação de ligações de hidrogênio entre as moléculas de água e entre as moléculas de metanol na mistura foram idênticas aquelas apresentadas nos itens 2.3.6 e 2.3.7, respectivamente. Para a existência de ligações de hidrogênio entre as moléculas de água e metanol na mistura utilizou-se as seguintes restrições geométricas:

- (1) Distância $r(\text{O}_{\text{água}} \cdots \text{O}_{\text{metanol}})$ menor do que 3,4 Å;
- (2) Distâncias $r(\text{O}_{\text{água}} \cdots \text{H}_{\text{metanol}})$ ou $r(\text{O}_{\text{metanol}} \cdots \text{H}_{\text{água}})$ menor do que 2,85 Å.

Quanto a distância O...O tomou-se a média das distâncias utilizadas para as moléculas individuais e a distância O...H foi obtida a partir da função de distribuição $\text{O}_{\text{água}} - \text{H}_{\text{metanol}}$ provenientes da simulação de Monte Carlo da mistura.

Analogamente aos critérios adotados nos itens 2.3.6 e 2.3.7, a existência de ligação de hidrogênio é aceita quando satisfaz os critérios geométricos descritos e a energia calculada é menor que zero ($E_{ab} < 0$).

2.4. Métodos de Química Quântica *ab initio*

Métodos *ab initio* são métodos de química quântica ou métodos de estrutura eletrônica utilizados em modelagem molecular. Para obter um aperfeiçoamento da energia de Hartree-Fock (HF) inclui-se os efeitos da correlação eletrônica, usando-se, por exemplo, a teoria de perturbação de Møller e Plesset (MP) no mínimo de segunda ordem (2), isto é, MP2. A vantagem da teoria de perturbação de muitos corpos é que ela é extensiva, ao contrário do método de interação de configuração (CI) truncado. Entretanto, a teoria de perturbação MP não é variacional e pode, em princípio, produzir energias menores que a exata. MP2 vem modelando com sucesso uma variedade de sistemas e propriedades⁸¹, permanecendo uma ferramenta poderosa para os químicos computacionais. Os cálculos químico-quânticos das estruturas e energias dos líquidos iônicos foram e serão realizados com este nível de teoria.

⁸¹ J. B. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc. (1996).

Para os cálculos já realizados, as estruturas dos cátions e ânions que formam os líquidos iônicos foram otimizadas usando o método MP2 com funções de base 6-31G** = 6-31G(d,p) para o cátion BMI e a base 6-31+G* = 6-31+G(d) para o ânion.

Erros de superposição de base

Quando se está interessado no cálculo da energia de interação de pares de moléculas, tal como a energia de formação do dímero da água, deve-se levar em consideração os erros gerados pela superposição de bases. Os erros de superposição de bases (ou *basis set superposition error* – BSSE), geralmente, tornam os dímeros mais estáveis do que deveriam, pois os monômeros podem utilizar as funções de base uns dos outros, causando uma estabilização exagerada⁸².

A maneira mais usual para estimar o BSSE é através do método de correção de “counterpoise” de Boys e Bernardi⁸³. Neste método o cálculo da energia de uma espécie individual é executado na presença das funções de base da outra espécie e vice-versa. Logo, a energia de interação intermolecular entre duas moléculas A e B é calculado como:

$$\Delta E = E_{AB}(AB) - E_A(AB) - E_B(AB) \quad (2.3)$$

onde $E_X(Y)$ significa a energia de X na presença das funções de base de Y.

Análise Conformacional

Além das estruturas moleculares e energias de interação, os métodos *ab initio* foram utilizados para análise conformacional dos grupos alquílicos dos anéis imidazólios. Uma exploração detalhada foi realizada com o método de mecânica molecular MM⁺ implementado no programa *HyperChem* 6.0⁸⁴. As conformações mais estáveis foram então utilizadas como estruturas de partida para as otimizações *ab initio*.

Os líquidos iônicos com cátion 1-alkil-3-metilimidazólio apresentam uma grande variação nas suas propriedades físico-químicas em função do ânion e dos substituintes presentes no anel imidazólio. Por exemplo, quanto maior o substituinte, maior o número de conformações possíveis, e menor o ponto de fusão. Dada esta importante influência nas propriedades dos líquidos iônicos, é proposta a inclusão de funções de potencial intermolecular para conformermeros distintos.

⁸² A. R. Leach, *Molecular Modelling. Principles and Applications*. Ed. Addison Wesley Longman Ltd. Edinburgh Gate, Harlow, England, (1998).

⁸³ S. F. Boys, F. Bernardi, *Mol. Phys.* 19, 553-566 (1970).

⁸⁴ Manual do Programa: *HyperChem Release 7 for Windows*, USA (2002).

2.5. Análise de Componentes Principais e Análise de Sensibilidade

A análise de componentes principais (PCA) é uma técnica de análise exploratória. Esta técnica transforma os dados na sua forma original em um espaço N-dimensional para um espaço M-dimensional, onde M é menor que N. Isto é realizado identificando-se as componentes principais⁸⁵, permitindo o reconhecimento de possíveis estruturas, com a menor perda de informação possível. A primeira componente (PC1) representa a direção de máxima variabilidade. A próxima componente (PC2) é a direção de máxima variabilidade que é ortogonal a primeira, a terceira é a direção de máxima variabilidade ortogonal as duas primeiras e assim sucessivamente. A Figura 2.5 ilustra o caso bidimensional. Em espaços de mais alta dimensionalidade, pode-se encontrar uma grande proporção de variabilidade (como 95%) representada em menos que a metade das componentes⁸⁶.

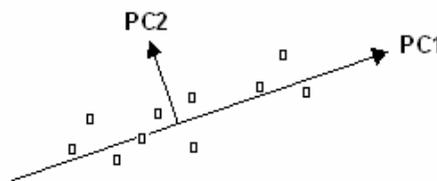


Figura 2.5: Diagrama para representação das componentes principais.

Análise de componente principais pode, portanto, reduzir a dimensionalidade dos dados nos estágios de pré-processamento para o treinamento de redes neurais artificiais. A redução permite a obtenção de redes neurais menores, e conseqüentemente, menos complexas, com treinamento e desempenho mais eficiente e rápido

O cálculo das componentes principais consiste em:

1) Cálculo da matriz de correlação R:

$$R = \frac{X^T X}{n-1} \quad (2.4)$$

onde n é o número de objetos ou dados e X e X^T é a matriz de dados e sua transposta, respectivamente.

⁸⁵ D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: a textbook*, Elsevier Science Publishers, Amsterdam, 339 (1990).

⁸⁶ *Electronic Manual for ST Neural Networks release 4.0 E* (2000).

2) A matriz R é diagonalizada para determinação das componentes principais (PC_k , $k = 1, 2, \dots, N$).

A realização da PCA provem da necessidade de redução do grande número de variáveis independentes necessárias para a descrição e o cálculo do potencial intermolecular. Para o cálculo destes potenciais, faz-se combinações das distâncias entre os sítios presentes entre pares de moléculas. Entretanto, sabe-se que orientação relativa entre moléculas pode ser completamente descrita pela distância entre os centros de massa e os ângulos de *Euler*. A PCA permite, portanto, a escolha mais apropriada das variáveis, ou seja, daquelas variáveis que capturam a maior variabilidade ou informações necessárias para o ajuste do potencial. No caso do modelo TIP5P, por exemplo, cada molécula contém cinco sítios e portanto 25 distâncias intermoleculares são utilizadas para o cálculo do potencial intermolecular analítico. Esta redução de complexidade é necessária para o ajuste de potenciais intermoleculares usando redes neurais artificiais, e portanto, reduzindo sua complexidade.

A análise de sensibilidade é uma técnica que permite a determinação dos parâmetros ou variáveis de uma dada função que mais influenciam no valor desta função. Na prática, esta ferramenta é útil para detectar as variáveis que possuem alta e baixa sensibilidade no modelo, ou seja, maior e menor importância. Apesar de necessitar cautela ao se decidir excluir variáveis, este procedimento é importante para evitar redundâncias e número excessivo de variáveis. Para este estudo, esta técnica complementa a aplicação da PCA.

O procedimento da análise de sensibilidade consiste na extração de cada uma das variáveis, no uso do modelo e posterior cálculo do erro. O erro é então comparado com aquele gerado utilizando-se o total das variáveis. Baseado na razão entre o primeiro e o segundo erros estima-se que a variável pode ser descartada, caso esta razão seja menor ou igual a unidade, sem perda de informação sobre o sistema.

2.6. Redes Neurais Artificiais

Redes neurais artificiais (RNA) são versáteis para a solução de problemas complexos e não lineares. Dependendo do problema a ser abordado, uma RNA pode possuir diversas arquiteturas. A RNA mais conhecida e utilizada é a chamada rede “perceptron” multicamada⁸⁷. A seguir serão apresentadas estas possíveis estruturas, o modelo matemático de um neurônio, os tipos de RNAs disponíveis e o método *backpropagation*

mais comumente empregados nos métodos de aprendizado supervisionado de uma rede neural.

2.6.1. Topologia

A topologia de uma ANN refere-se ao modo como os neurônios são interconectados. Existem três opções para promover a interconexão entre os neurônios. São elas:

- 1) Conexões intracamadas – Nas quais, as saídas a partir de um neurônio alimenta outros neurônios na mesma camada (Figura 2.6).
- 2) Conexões intercamadas – As saídas a partir de um neurônio em uma camada alimenta neurônios em outra camada (Figura 2.7).
- 3) Conexões recorrentes – A saída a partir de um neurônio alimenta ele mesmo (Figura 2.8).

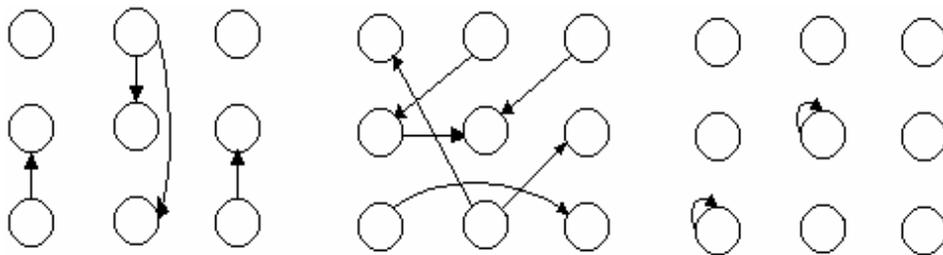


Figura 2.6: Intracamada

Figura 2.7: Intercamada

Figura 2.8: Recorrentes

As conexões intercamadas são as mais amplamente utilizadas em problemas de engenharia e podem ser *feedback* (Figura 2.9) ou *feedforward* (Figura 2.10).

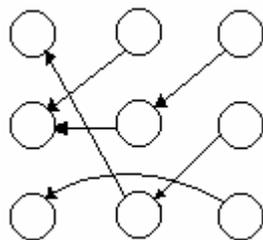


Figura 2.9: *feedback*

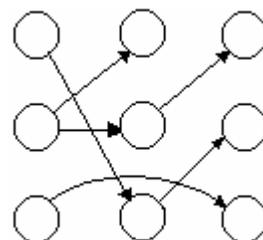


Figura 2.10: *feedforward*

⁸⁷ K. L. Kovács, *Redes Neurais Artificiais. Fundamentos e Aplicações*. pp. 31, Ed. Acadêmica. Universidade de São Paulo, São Paulo (1996).

O tipo de problema a ser tratado determina a topologia a ser usada⁸⁸. Desejando-se desenvolver uma rede neural, para usar treinamento não supervisionado, deve-se usar conexões *feedback*, e para casos em que se deseja obter a modelagem de um determinado sistema, onde o treinamento será supervisionado, para obter uma resposta baseada em um sinal de entrada, deve-se usar conexões *feedforward*. Este tipo de conexão foi escolhida para a descrição dos potenciais de interação intermoleculares.

2.6.2. Equações de Redes Neurais Multicamadas

As redes neurais processam dados numéricos, produzindo sinais de saída a partir de dados de entrada, após uma fase de treinamento, na qual a rede é submetida a diversos padrões com dados de entrada e saída. As entradas são apresentadas como um vetor, e os neurônios manipulam estas entradas, ou ativações, para gerar uma saída, que pode ser entrada para outros neurônios. As entradas são chamadas de ativações após serem processadas em outros neurônios.

A saída a partir de um neurônio “j” qualquer da camada (k+1) é afetada pela entrada ($S_{pi,k}$, $i = 1, \dots, n_k$ neurônios da camada anterior) e também pelos pesos (w_{jik}) e *biases* ($\theta_{j,k+1}$). Uma vez calculado o produto da entrada com o peso, e somado o *bias*, este valor torna-se entrada para uma função que desempenha o papel do potencial pós-sináptico, na linguagem biológica, e por isto esta função é conhecida também como função PSP, oriundo do inglês *Post-Synaptic Potential*. A Figura 2.11 ilustra este processo. Desse modo, a função de ativação é o fator governante final que produz a saída a partir de um neurônio. O comportamento de cada neurônio é expresso pela equação:

$$I_{pj,k+1} = \left[\sum_{i=1}^{n_k} w_{jik} S_{pi,k} \right] + \mathbf{q}_{j,k+1} \quad (2.5)$$

⁸⁸ D. R. Baughman, Y. A. Liu, *Neural Networks in Bioprocessing and Chemical Engineering*. Academic Press, Inc. San Diego, Califórnia. (1995). S. Haykin, *Redes Neurais: Princípios e Prática*. 2. ed. Artmed Editora Ltda. Porto Alegre, RS, Brasil (2001).

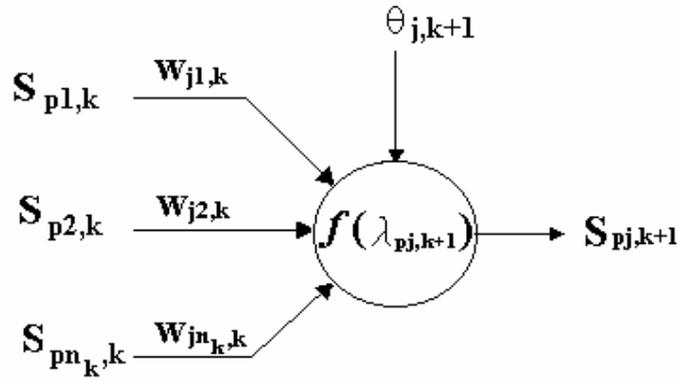


Figura 2.11: j-ésimo neurônio na camada (k+1).

São exemplos de função de ativação (PSP), com seus respectivos intervalos de aplicação:

- Função linear: $f(I_{pj,k+1}) = aI_{pj,k+1} + b$, com $a, b \in R$ $[-\infty, +\infty]$ (2.6)

- Função degrau limiar:
$$\begin{cases} f(I_{pj,k+1}) = +x & \text{Para } I_{pj,k+1} > 0 \\ f(I_{pj,k+1}) = -x & \text{Para } I_{pj,k+1} \leq 0 \end{cases}$$
 (2.7)

- Função degrau de Heaviside
$$\begin{cases} f(I_{pj,k+1}) = 1 & \text{Para } I_{pj,k+1} > VL \\ f(I_{pj,k+1}) = 0 & \text{Para } I_{pj,k+1} \leq VL \end{cases}$$
 (2.8)

- Função sigmoideal: $f(I_{pj,k+1}) = \frac{1}{1 + \exp(-I_{pj,k+1})}$ $[0, +1]$ (2.9)

- Função exponencial: $f(I_{pj,k+1}) = \exp(-I_{pj,k+1})$ $[0, +\infty]$ (2.10)

- Função tangente hiperbólica: $f(I_{pj,k+1}) = \tanh(I_{pj,k+1})$ $[-1, +1]$ (2.11)

- Função seno: $f(I_{pj,k+1}) = \text{sen}(-I_{pj,k+1})$ $[-1, +1]$ (2.12)

- Função gaussiana:
$$f_{pj} = \prod_{i=1}^{n_k} \exp\left[\frac{(s_{pi,k} - m_{ji})}{-2\sigma_j^2}\right]$$
 $[0, +1]$ (2.13)

onde σ_j representa o desvio-padrão da gaussiana e μ_{ij} , os centros adaptativos da mesma, que podem ser fixados em μ .

A função mais amplamente empregada para redes neurais do tipo MLP é a sigmoide, que limita-se em [0, 1]. Os dados de entrada para a rede devem ser normalizados para este intervalo.

2.6.3. Tipos de Redes Neurais Artificiais

RNAs podem apresentar diversas arquiteturas, cada uma delas possui diversas vantagens e desvantagens, sendo capazes de responder a certos estímulos, dependendo do problema. Entre os tipos de RNAs, destacam-se as redes Perceptron Multicamadas (ou *Multilayer Perceptron – MLP*), as redes de Função de Base Radial (ou *Radial Basis Function – RBF*), as redes do tipo Bayesianas, que se dividem em dois tipos: Redes Neurais Probabilísticas (ou *Probabilistic Neural Networks – PNN*), a qual é usada somente para propósitos de classificação e Redes Neurais de Regressão Generalizada (ou *Generalized Regression Neural Networks - GRNN*) que são apenas usadas para regressão. Destacam-se ainda, as RNAs lineares e as de Kohonen.

Redes MLP são as redes neurais artificiais mais comumente utilizadas⁸⁹. Estas redes utilizam uma função PSP linear, ou seja, elas executam somas de suas entradas multiplicadas por pesos. Permitem o uso de diversas funções de ativação, geralmente não lineares. Podem ter mais de 128 camadas, mas geralmente três ou quatro são suficientes. O treinamento pode ser realizado usando qualquer um dos seguintes algoritmos: gradiente conjugado, quase-Newton, Levenberg-Marquardt, backpropagation, quickpropagation e regra delta.

MLP e RBF são aproximadores universais, por isto, constata-se que sempre existe uma rede RBF capaz de imitar precisamente uma MLP ou vice-versa. Redes RBF contêm uma camada de entrada, uma camada escondida de unidades radiais (funções gaussianas) e uma camada de saída de unidades lineares. Estas duas redes diferem entre si em cinco aspectos importantes⁹⁰:

- 1) Redes neurais RBF são treinadas mais rapidamente, entretanto estas redes geralmente são maiores que as redes MLP, executando portanto mais lentamente;
- 2) Uma rede RBF, na sua forma mais básica, tem uma única camada oculta, enquanto que uma rede MLP pode ter uma ou mais camadas ocultas;
- 3) A camada oculta de uma rede RBF é não linear e a de saída é linear, entretanto, as camadas ocultas e de saída de uma MLP usada como classificadores de padrões são

⁸⁹ E. O. de Cerqueira, J. C. de Andrade, R. J. Poppi, *Quím. Nova.* 24, 864-873 (2001).

normalmente todas não lineares. Deve-se salientar, entretanto, que quando a MLP é utilizada para resolver problemas de regressão não-linear, uma camada linear para a saída é normalmente a escolha preferida;

4) Diferentes PSPs. Uma rede RBF calcula a norma (distância) euclidiana entre o vetor de entrada e o centro daquela unidade, enquanto uma rede MLP calcula o produto interno do vetor de entrada com o vetor de peso sináptico daquela unidade;

5) As redes MLPs constroem aproximações globais de um mapeamento de entrada-saída não-linear. Por outro lado, as redes RBF utilizando não linearidades localizadas com decaimento exponencial constroem aproximações locais para mapeamento de entrada-saída não-lineares.

Do exposto conclui-se que para a aproximação de um mapeamento de entrada-saída não-linear, uma rede MLP geralmente requer um número menor de parâmetros que a rede RBF para o mesmo grau de precisão. Entretanto, necessita-se cautela e rigor, já que arquitetura e tipo de rede depende da natureza do problema e quantidade de dados disponíveis, logo deve-se realizar uma varredura testando diversas redes neurais e verificando os seus desempenhos com relação ao problema de interesse.

Redes lineares têm apenas duas camadas (uma entrada e uma saída), e a camada de saída tem função de ativação e PSP lineares, de modo que a maior parte dos problemas não podem ser solucionados ou bem solucionados usando este tipo de rede neural, entretanto alguns problemas podem e por isso deve-se testar a resolução de problemas usando técnicas mais simples, gerando padrão de comparação para demonstrar necessidade daquelas mais complexas. Estas redes lineares podem ser treinadas para executar, por exemplo, análise de componentes principais.

Redes Bayesian treinam muito rapidamente (quase instantaneamente), mas por serem comumente grandes, executam lentamente.

Redes Kohonen têm uma camada de entrada e uma camada de saída de unidades radiais e são principalmente utilizadas em problemas de classificação ou detecção de *clusters* ou agrupamentos. Estas redes neurais são treinadas usando algoritmos de aprendizado não supervisionado, não requerendo variáveis de saída no conjunto de treinamento.

⁹⁰ S. Haykin, *Redes Neurais. Princípios e Prática*, 2. ed., Bookman, Porto Alegre (2001).

2.6.4. O Método de Retropropagação (*backpropagation*)

O objetivo do treinamento pelo método *backpropagation* consiste em ajustar os pesos (w 's) e *biases* (θ 's) da rede, modificando-os até que uma tolerância entre as saídas da rede “S”, e as saídas desejadas “t”, para N_p padrões, seja atendida, ou seja, minimizar a função erro, com respeito aos pesos e *biases*.

O número de camadas escondidas, bem como o número de neurônios nestas camadas deve ser determinado numa etapa preliminar da fase de treinamento. O método de retropropagação é descrito a seguir para uma rede com apenas uma camada escondida usada por De Souza Jr., 1993⁹¹, apresentada na Figura (2.12). O método poder ser generalizado, apenas com pequenas modificações para a arquitetura de interesse.

A função objetivo a ser minimizada será:

$$E = \sum_{p=1}^{N_p} E_p \tag{2.14}$$

onde E_p é o erro quadrático em relação a cada padrão processado pela rede:

$$E_p = \frac{1}{2} \sum_{l=1}^{n_L} (S_{pl,L} - t_{pl})^2 \tag{2.15}$$

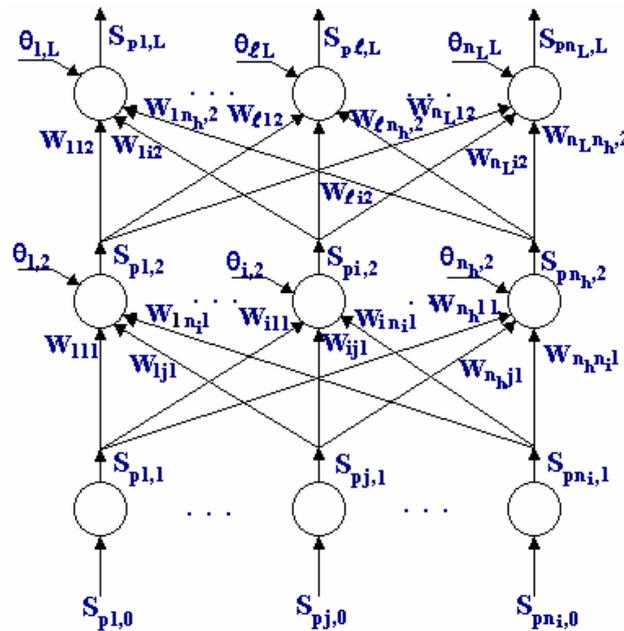


Figura 2.12: Rede neural *feedforward* com 3 camadas.

⁹¹ M. B. De Souza Jr., “Redes Neurais Multicamadas Aplicadas a Modelagem e Controle de Processos Químicos”. Tese de Doutorado. COPPE/PEQ/UFRJ. Rio de Janeiro (1993).

O método do *gradiente decrescente*, o qual é classificado como um método multidimensional indireto, usa a informação do gradiente para calcular a direção de busca, que consiste no negativo do gradiente (daí ser chamado, gradiente *decrescente*), com η fixo, onde η é a “taxa de aprendizagem”.

A condição necessária para um extremo é dada por:

$$\nabla E(W) = 0 \quad (2.16)$$

Assim para a k -ésima iteração:

$$\Delta W_{k+1} = -\mathbf{h}\nabla E(W_k) \quad (2.17)$$

A taxa de aprendizagem, η , é uma constante positiva que controla a taxa na qual os novos pesos são ajustados. Geralmente tem magnitude menor que 1.

Os passos do treinamento *backpropagation* são compostos de duas fases:

Fase 1 – A entrada é apresentada e propagada para frente através da rede, até computar os sinais de saída que são comparados com as saídas alvos, obtendo-se um erro quadrático E_p .

Fase 2 – O erro obtido na fase 1 é propagado para trás e os pesos vão sendo ajustados.

Antes de começar o treinamento da rede, deve-se fornecer valores aleatórios para os pesos entre as camadas e para os *biases*. A seleção deve ser mais cuidadosa, quanto mais complexa for a estrutura da rede neural. Comumente, esses fatores são ajustados usando uma distribuição Gaussiana.

Cada fase é, a seguir apresentada detalhadamente.

FASE 1:

Para cada padrão p , variando de 1 até N_p padrões, tem-se um conjunto de entradas $S_{pj,0}$ ($j = 1, 2, \dots, n_i$), onde n_i é o número de neurônios na camada de entrada, e um conjunto de saídas desejadas t_{pj} ($j = 1, 2, \dots, n_L$), onde n_L é o número de neurônios na camada de saída.

A saída da primeira de camada de neurônios, tem-se $S_{pj,1}$ (equação 2.18), onde p representa um padrão apresentado, o índice 1 representa a saída a partir da primeira camada e j varia de 1 até n_i (cada neurônio da camada de entrada).

$$S_{pj,1} = f(S_{pj,0}) \quad j = 1, \dots, n_i \quad (2.18)$$

f pode ser dada por uma das funções (2.6)-(2.13), sendo que a mais amplamente utilizada é a função sigmoïdal (equação 2.9).

A saída da camada de entrada junto com os pesos e *biases*, produz as seguinte ativações:

$$S_{pj,2} = f(\mathbf{I}_{pj,2}) \quad (2.19)$$

onde :

$$\mathbf{I}_{pj,2} = \left[\sum_{i=1}^{n_i} w_{ji1} S_{pi,1} \right] + \mathbf{q}_{j,2} \quad j = 1, 2, \dots, n_h.$$

tal que f é também dada pela equação escolhida anteriormente entre as equações (2.6)-(2.13)

Finalmente, à saída da rede, tem-se:

$$S_{pj,L} = f(\mathbf{I}_{pj,L}) \quad (2.20)$$

onde:

$$\mathbf{I}_{pj,L} = \left[\sum_{i=1}^{n_h} w_{ji2} S_{pi,2} \right] + \mathbf{q}_{j,L} \quad (2.21)$$

sendo que o índice L indica a camada de saída.

As saídas da rede são comparadas com as saídas desejadas e calcula-se o erro quadrático que é propagado para trás na segunda fase do treinamento.

FASE 2:

Os pesos podem ser atualizados após a apresentação de cada padrão (aprendizagem por padrão) ou após a apresentação de todos os padrões (aprendizagem por época ou batelada). A seguir é descrita a aprendizagem por época.

Cálculo de $\nabla E(W)$, onde W é o vetor contendo os pesos e *biases*.

A derivada parcial do erro em relação aos pesos w_{ji2} , entre a camada escondida e a camada de saída, é calculada pela regra da cadeia do cálculo elementar:

$$\frac{\partial E}{\partial w_{ji2}} = \sum_{p=1}^{N_p} \frac{\partial E_p}{\partial S_{pj,L}} \frac{\partial S_{pj,L}}{\partial \mathbf{I}_{pj,L}} \frac{\partial \mathbf{I}_{pj,L}}{\partial w_{ji2}} = \sum_{p=1}^{N_p} (S_{pj,L} - t_{pj}) f'(\mathbf{I}_{pj,L}) S_{pi,2} \quad (2.22)$$

onde $f'(\mathbf{I}_{pj,L}) = \frac{\partial S_{pj,L}}{\partial \mathbf{I}_{pj,L}}$

Analogamente, a derivada em relação aos *biases* $\mathbf{q}_{j,L}$ da última camada:

$$\frac{\partial E}{\partial \mathbf{q}_{j,L}} = \sum_{p=1}^{N_p} \frac{\partial E_p}{\partial S_{pj,L}} \frac{\partial S_{pj,L}}{\partial \mathbf{I}_{pj,L}} \frac{\partial \mathbf{I}_{pj,L}}{\partial \mathbf{q}_{j,L}} = \sum_{p=1}^{N_p} (S_{pj,L} - t_{pj}) f'(\mathbf{I}_{pj,L}) \quad (2.23)$$

Para os pesos w_{ji1} entre a camada de entrada e a camada escondida:

$$\frac{\partial E}{\partial w_{ji1}} = \sum_{p=1}^{N_p} \sum_{l=1}^{n_L} \frac{\partial E_p}{\partial S_{pl,L}} \frac{\partial S_{pl,L}}{\partial \mathbf{I}_{pl,L}} \frac{\partial \mathbf{I}_{pl,L}}{S_{pj,2}} \frac{\partial S_{pj,2}}{\partial \mathbf{I}_{pj,2}} \frac{\partial \mathbf{I}_{pj,2}}{\partial w_{ji1}} \quad (2.24)$$

$$\frac{\partial E}{\partial w_{ji1}} = \sum_{p=1}^{N_p} \left(\sum_{l=1}^{n_L} (S_{pl,L} - t_{pl}) f'(\mathbf{I}_{pl,L}) w_{lj2} \right) f'(\mathbf{I}_{pj,2}) S_{pi,1} \quad (2.25)$$

onde os w_{ji2} utilizados são os da iteração anterior.

Para os *biases* $\mathbf{q}_{j,2}$ da camada escondida:

$$\frac{\partial E}{\partial \mathbf{q}_{j,2}} = \sum_{p=1}^{N_p} \sum_{l=1}^{n_L} \frac{\partial E_p}{\partial S_{pl,L}} \frac{\partial S_{pl,L}}{\partial \mathbf{I}_{pl,L}} \frac{\partial \mathbf{I}_{pl,L}}{S_{pj,2}} \frac{\partial S_{pj,2}}{\partial \mathbf{I}_{pj,2}} \frac{\partial \mathbf{I}_{pj,2}}{\partial \mathbf{q}_{j,2}} \quad (2.26)$$

$$\frac{\partial E}{\partial \mathbf{q}_{j,2}} = \sum_{p=1}^{N_p} \left(\sum_{l=1}^{n_L} (S_{pl,L} - t_{pl}) f'(\mathbf{I}_{pl,L}) w_{lj2} \right) f'(\mathbf{I}_{pj,2}) \quad (2.27)$$

onde os w_{lj2} utilizados são os da iteração anterior.

Os passos das duas fases são repetidos até que o erro satisfaça uma dada tolerância. Enquanto houver diferença (o delta do nome da regra) entre as saídas da rede e os valores desejados para os padrões E_p , essa diferença vai dar um sinal de volta pela rede e usado para ajustar os pesos.

Existem variantes do método para o qual é permitido variar a taxa de aprendizagem, somando-se ao lado direito da equação (2.17) um termo de momento (ϵ). O coeficiente de momento é um peso extra adicionado para acelerar a taxa em que os fatores são ajustados. Dessa forma, a taxa de aprendizagem e o termo de momento são parâmetros muito importantes que controlam a eficiência com que o algoritmo treina a rede.

A equação (2.17) fica:

$$\Delta W_{k+1} = -\mathbf{h} \nabla E(W_k) + \epsilon \Delta W_k \quad (2.28)$$

Baughman & Liu (1995)⁹², comentam que o conjunto de valores para a taxa de aprendizado e o coeficiente de momento não são os melhores em todas as etapas do treinamento e sugere que o método de aprendizado deve ser capaz de gradualmente reduzir estes valores para próximo de zero.

⁹² D. R. Baughman, Y. A. Liu, *Neural Networks in Bioprocessing and Chemical Engineering*. Academic Press, Inc. San Diego, Califórnia (1995).

De Souza Jr. (1993)⁹³, compara diversas modificações, entre elas uma formulação baseada no método de Newton e na Hessiana analítica.

O método do gradiente com direções conjugadas de Leonard & Kramer (1990)⁹⁴, será adotado neste trabalho. Nele η e ϵ são ajustados dinamicamente. A direção de busca s_{k+1} na iteração $k+1$ é gerada recursivamente como segue:

$$s_0 = -\nabla E(W_0) \quad (2.29)$$

$$s_{k+1} = -\nabla E(W_{k+1}) + \epsilon s_k \quad (2.30)$$

$$\epsilon = [\nabla E(W_{k+1})^T \nabla E(W_{k+1})] / [\nabla E(W_k)^T \nabla E(W_k)] \quad (2.31)$$

onde o vetor dos pesos W_{k+1} foi determinado usando-se uma busca unidimensional (interpolação quadrática) para η :

$$W_{k+1} = W_k + \eta s_k \quad (2.32)$$

O método utilizado elimina as oscilações do gradiente decrescente com passo fixo.

2.7. Análise Topológica de Redes de Ligações de Hidrogênio

As propriedades topológicas⁹⁵ das redes de ligações de hidrogênio foram determinadas calculando-se o coeficiente de agregação ou *clustering* C , o comprimento do caminho L , o grau de distribuição $P(k)$ e o espectro do grafo $\rho(\lambda)$. O coeficiente de *clustering*, C_i , em um nó, sítio ou vértice (molécula de água) i na rede corresponde ao número de conexões ou ligações de hidrogênio, E_i , entre os vizinhos do sítio i dividido pelo número de possíveis conexões ou ligações entre eles, $\binom{k_i}{2}$,

$$C_i = E_i / \binom{k_i}{2} = \frac{2E_i}{k_i(k_i - 1)} \quad (2.33)$$

O coeficiente de *clustering* de uma rede é igual à média sobre todos os coeficientes C_i 's, ou seja, $C = \langle C_i \rangle$. Este pode ser formalizado em termos da teoria de grafos⁹⁶

⁹³ M. B. De Souza Jr., “Redes Neurais Multicamadas Aplicadas a Modelagem e Controle de Processos Químicos”. Tese de Doutorado. COPPE/PEQ/UF RJ. Rio de Janeiro (1993).

⁹⁴ J. Leonard, M. A. Kramer, *Computers Chem. Engng.*, 14, 337-341 (1990).

⁹⁵ R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* 74, 47 (2002).

⁹⁶ R. F. Cancho, C. Janssen, R. V. Solé, *Physical Review E* 64, 046119 (2001).

considerando o conjunto de conexões \mathbf{x}_{ij} ($i, j = 1, 2, \dots, N$), onde $\mathbf{x}_{ij} = 1$ se a conexão existe e $\mathbf{x}_{ij} = 0$ se a conexão não é formada. Tomando, $k_i = \{j | \mathbf{x}_{ij} = 1\}$ como o conjunto dos vizinhos mais próximos ao sítio i , então o número de conexões entre eles é dado por

$$E_i = \sum_{j=1}^N \mathbf{x}_{ij} \left[\sum_{j < k; k \in k_i} \mathbf{x}_{jk} \right] \quad (2.34)$$

e o coeficiente de *clustering* em um sítio i pode ser calculado usando a equação 2.33.

Para um grafo randômico, o qual consiste em uma rede cujas conexões são distribuídas aleatoriamente com probabilidade p , o coeficiente de *clustering* pode ser calculado, como,

$$C_{rand} = p = \frac{\langle k \rangle}{N} \quad (2.35)$$

onde N é o número de sítios ou nós da rede e $\langle k \rangle$ é o grau médio (*average degree*) ou conexões, ou *links* ou ligações de hidrogênio da rede. Tomando $l_{min}(i, j)$ como o comprimento do caminho mínimo (número de conexões mínimo entre dois sítios i e j em uma mesma rede), o comprimento médio⁹⁶ de um dado sítio pode ser calculado como,

$$l(i) = \frac{1}{N} \sum_{j=1}^N l_{min}(i, j) \quad (2.36)$$

O comprimento médio da rede corresponde à média sobre todos os sítios do agregado: $L = \langle l(i) \rangle$.

Para um grafo randômico o comprimento médio é dado por,

$$L_{rand} \approx \frac{\ln N}{\ln \langle k \rangle} \quad (2.37)$$

Um padrão de *small-world* é caracterizado por: $C \gg C_{rand}$ e $L \approx L_{rand}$.⁹⁷

Como nem todos os sítios em uma rede tem o mesmo número de conexões, o grau de distribuição $P(k)$ pode ser definido como a frequência (normalizada) dos sítios tendo k conexões. Em um grafo randômico as conexões são substituídas aleatoriamente, tal que a maior parte dos sítios tem aproximadamente o mesmo grau, próximo do grau médio $\langle k \rangle$ da rede. Especificamente⁹⁵, para um grafo randômico $P(k)$ é uma distribuição binomial, a qual pode ser substituída por uma distribuição de *Poisson* com pico na média $P(\langle k \rangle)$, quando N é grande. Para uma rede com padrão de *small-world* descrita pelo modelo de Watts-

⁹⁷ D. J. Watts, S. H. Strogatz, *Nature* 393, 440 (1998); R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* 74, 47 (2002).

Strogatz⁹⁸, a forma do grau de distribuição é semelhante àquela dos grafos randômicos. Todavia, para algumas redes complexas com comportamento de *small-world*, a forma do grau de distribuição desvia significativamente da distribuição de *Poisson* e são geralmente descritas como⁹⁹: (i) redes *scale-free*, com lei de potência: $P(k) \approx k^{-g}$; (ii) redes *broad-scale*, $P(k) \approx k^{-g} f(k/k^*)$ onde k^* é o raio de corte para a função de corte f da lei de potência; (iii) distribuição *single-scale* (exponencial ou Gaussiana).

Uma matriz de adjacência $A(G)$ com $N \times N$ elementos A_{ij} pode representar um grafo G com N sítios, quando os elementos satisfazem as seguintes condições: $A_{ij} = A_{ji} = 1$ se os sítios i e j estão conectados, e 0 se os sítios não estão conectados¹⁰⁰. A diagonalização da matriz de adjacência $A(G)$ produz um conjunto de N autovalores $\{\lambda_j\}$ que formam o espectro do grafo G e origina a densidade espectral, $\mathbf{r}(\mathbf{I})$,⁴⁰ a qual se aproxima de uma função continua quando $N \rightarrow \infty$,

$$\rho(\lambda) = \frac{1}{N} \sum_{j=1}^N \delta(\lambda - \lambda_j) \quad (2.38)$$

A importância da análise desta propriedade advém do fato de que a densidade espectral pode estar diretamente relacionada aos aspectos topológicos do grafo, pois o seu k -ésimo momento pode ser escrito como o número de caminhos que retornam ao mesmo sítio no grafo, sendo expresso matematicamente como,

$$\frac{1}{N} \sum_{j=1}^N (\lambda_j)^k = \frac{1}{N} \sum A_{i_1, i_2} A_{i_2, i_3} \dots A_{i_k, i_1} \quad (2.39)$$

Deve-se notar ainda que estes caminhos podem conter sítios que já foram visitados.

Para um grafo randômico com N sítios e conexões aleatórias com probabilidade p , a densidade espectral converge para distribuição semi-circular expressa como¹⁰⁰:

$$\mathbf{r}(\mathbf{I}) = \begin{cases} \frac{\sqrt{4Np(1-p) - \mathbf{I}^2}}{2pNp(1-p)} & \text{se } |\mathbf{I}| < 2\sqrt{Np(1-p)} \\ 0 & \text{caso contrário} \end{cases} \quad (2.40)$$

O maior (principal) autovalor é isolado do total do espectro e decresce com pN .

Neste trabalho um programa escrito em linguagem Fortran foi elaborado para executar a leitura das coordenadas cartesianas das moléculas de água a partir de caixas de simulação, determinar as ligações de hidrogênio existentes (conexões) e então executar a análise topológica das redes de ligações de hidrogênio.

⁹⁸ D. J. Watts, S. H. Strogatz, *Nature* 393, 440 (1998).

⁹⁹ L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* 97, 11149 (2000).

¹⁰⁰ I. J. Farkas, I. Derényi, A.-L. Barabási, T. Vicsek, *Phys. Rev. E* 64, 026704 (2001).

CAPÍTULO 3

RESULTADOS E DISCUSSÕES

3.1. Introdução

Neste capítulo serão apresentados e discutidos os resultados para as simulações computacionais da água, metanol e da mistura binária água - metanol 1:1, usando o método de Monte Carlo. Os modelos de interação TIP4P e TIP5P para a água foram utilizados e para o metanol o modelo OPLS com três sítios. Foram analisadas as propriedades termodinâmicas e estruturais destes líquidos e da mistura em condições de altas pressões e temperaturas e nas condições ambiente. O modelo TIP5P gerou melhores resultados para a determinação das funções de distribuição radial e por isso foi escolhido para a etapa seguinte, que consistiu na realização da estatística e análise topológica destes líquidos puros e na mistura em uma condição extrema e nas condições ambiente. O modelo TIP5P também foi utilizado para gerar dados de energia de interação para a água líquida com o objetivo de desenvolver um potencial de interação intermolecular baseado em Redes Neurais Artificiais (RNAs). Para tal usou-se o programa *Statistica Neural Networks* e o programa RNA¹⁰¹ para geração de redes perceptron multicamadas em linguagem Fortran.

Utilizando RNAs a descrição do potencial independe do número de sítios na molécula ou modelo da molécula, podendo ser de interesse para simulação de moléculas com grande número de átomos. Também numa etapa inicial foram obtidos modelos usando a técnica de regressão de mínimos quadrados parciais (PLS), o qual apresentou um desvio absoluto maior nos primeiros testes. Visto que a superfície de energia potencial tem uma não linearidade muito acentuada, optamos por testar a seguir apenas os modelos baseados em RNAs.

A demonstração da viabilidade da técnica para o cálculo do potencial de interação para a molécula de água sugeriu a realização de cálculos *ab initio* com o objetivo de gerar dados para a simulação de líquidos iônicos. Neste caso o número de sítios nas espécies cátion e ânion é muito maior que para a água, tornando a descrição do potencial de interação ainda mais importante para a viabilidade das simulações.

¹⁰¹ V. M. Leite Santos, Laboratório de Análise e Controle de Processos – DEQ, UFPE (2000).

3.2. Simulação Computacional da Água com os Modelos TIP4P e TIP5P

Simulações de Monte Carlo foram executadas com os modelos de potencial TIP4P e TIP5P em diferentes condições próximas da crítica e também em condição ambiente. As simulações computacionais foram realizadas com o programa DIADORIM¹⁰², mantendo-se o número de moléculas (N), a temperatura (T) e a pressão (p) constantes e controladas.

3.2.1. Propriedades Termodinâmicas a 25°C e 1 atm

Para cada simulação utilizou-se caixas cúbicas com 512 moléculas, raio de corte igual a 9 Å, tentativa de variação de volume a cada 2000 configurações e razão de aceitação próxima de 40%, em concordância com trabalhos da literatura¹⁰³. Isto permitiu comparações das propriedades termodinâmicas calculadas com aquelas disponíveis na literatura.

A simulação da água foi realizada para diferentes condições, sendo que para a condição ambiente, cuja etapa de equilíbrio necessitou de 6×10^6 configurações e aquisição de no mínimo 8×10^6 , as propriedades termodinâmicas da água estão apresentadas na Tabela 3.1 e foram comparadas com aquelas obtidas por de Mahoney & Jorgensen (2000)¹⁰³ para TIP5P e por Jorgensen & Jenson (1998)¹⁰⁴ para TIP4P.

Tabela 3.1: Propriedades termodinâmicas calculadas e experimentais para a água líquida nas condições ambiente (1 atm e 25°C). Os valores entre parênteses são os erros percentuais com relação aos dados experimentais.

Propriedade	TIP4P ³	TIP5P ³	EXP. ⁴	TIP4P	TIP5P
$\rho(\text{g cm}^{-3})$	$1,001 \pm 0,001$	$0,999 \pm 0,001$	0,997	1,011 (1,4)	1,005 (0,8)
$-E/N (\text{kcal mol}^{-1})$	$10,06 \pm 0,01$	$9,87 \pm 0,01$	9,92	10,15 (2,3)	9,96 (0,4)
$C_p(\text{cal mol}^{-1} \text{K}^{-1})$	$20,4 \pm 0,7$	$29,1 \pm 0,8$	18,0	14,9 (17,2)	22,5 (25,0)
$10^6 \kappa (\text{atm}^{-1})$	60 ± 5	41 ± 2	45,8	48,9 (6,8)	49,2 (7,4)
$10^5 \alpha (\text{K}^{-1})$	44 ± 8	63 ± 6	25,7	58,1 (126)	55,9 (117)

¹⁰² L. C. Gomide Freitas. Laboratório de química computacional da universidade federal de São Carlos. UFSCar (1999).

¹⁰³ M. W. Mahoney, W. L. Jorgensen. *J. Chem. Phys.* 112, 8910 (2000).

¹⁰⁴ W. L. Jorgensen, C. Jenson. *J. Comput. Chem.* 19, 1179 (1998).

As comparações entre as simulações foram possíveis devido a manutenção dos parâmetros de simulação análogos, incluindo uma razão de aceitação próxima de 40%. A manutenção da taxa de aceitação em torno de 40% foi realizada através do ajuste dos valores máximos para os movimentos de rotação (ângulo) e translação (distância). Estes valores são apresentados na Tabela 3.2.

A capacidade calorífica foi obtida por flutuação da energia, analogamente ao trabalho de Mahoney & Jorgensen (2000)¹⁰³, que gerou 300×10^6 configurações. Neste trabalho gerou-se 150×10^6 configurações. A análise da Tabela 3.1 mostra que o modelo TIP5P forneceu melhores resultados para o cálculo da densidade e os resultados para as demais propriedades são comparáveis àqueles obtidos por Mahoney & Jorgensen, e aos resultados experimentais. Além disso, a comparação entre os simuladores sugere que o protocolo utilizado é apropriado, pois as propriedades calculadas são similares dentre dos erros estatísticos.

3.2.2. Propriedades Termodinâmicas em condições extremas

Para as condições próximas da crítica, os protocolos de simulação tiveram que ser alterados consideravelmente quando comparados com o protocolo utilizado para a condição ambiente. Os mínimos de configurações utilizadas para a equilibração e aquisição foram 8×10^6 e 10×10^6 , respectivamente. Os outros parâmetros do protocolo de simulação estão apresentados na Tabela 3.2.

Tabela 3.2: Protocolos de simulação incluindo razão de aceitação A (%), ângulo máximo de rotação θ_{rot} ($^\circ$), deslocamento máximo de translação $R_{\text{máx}}$ (?), razão de aceitação das variações de volume A_v (%), máxima variação de volume $V_{\text{máx}}$ e dimensão final da caixa cúbica.

p (bar)	T ($^\circ\text{C}$)	Modelo	A (%)	θ_{rot} ($^\circ$)	$R_{\text{máx}}$ (?)	A_v (%)	$V_{\text{máx}}$	a (?)
1	25	TIP5P	~40	15,5	0,155	~40	186	24,77
		TIP4P	~43	14,0	0,14	47	120	24,64
100	150	TIP5P	~40	24,0	0,24	~40-50	290	26,54
		TIP4P	~43	18,0	0,18	~46	170	25,60
1100	300	TIP5P	~40	34,5	0,35	~40	485	27,66
		TIP4P	~50	20,0	0,20	~45	400	27,05
1300	400	TIP5P	~40	43,0	0,43	~39	575	29,35
		TIP4P	~60	24,0	0,24	~70	360	26,54

p (bar)	T (°C)	Modelo	A (%)	θ_{rot} (°)	$R_{\text{máx}}$ (?)	A_v (%)	$V_{\text{máx}}$	a (?)
1900	150	TIP5P	~40	22,5	0,225	~45	275	25,02
		TIP4P	~40	19,0	0,19	~55	160	24,86
1970	300	TIP5P	~39	32,5	0,325	~40	410	26,48
		TIP4P	~40	26,0	0,26	~42	400	26,02
2800	300	TIP5P	~40	31,0	0,31	~50	345	25,81
		TIP4P	~39	29,0	0,29	~45	450	25,52
3400	400	TIP5P	~40	36,0	0,36 0,30	~42	460	26,46
		TIP4P	~40	30,0		~43	520	26,34
500	300	TIP5P	~42	37,0	0,37 0,30	~45	540	29,78
		TIP4P	~45	30,0		~52	430	27,80
500	400	TIP5P	~43	57,0	0,57	~60	1100	34,84
		TIP4P	~45	48,0	0,48	~62	650	32,75

Os resultados acima mostram que em condições de altas pressões e temperaturas existe grande dificuldade de controle sobre as razões de aceitação, que foi feita durante a equilibração de forma manual a cada 1 milhão de configurações. A dificuldade foi particularmente maior para ajustar as aceitações para variações de volume.

As propriedades termodinâmicas obtidas nas simulações são apresentadas na Tabela 3.3.

Tabela 3.3: Algumas propriedades termodinâmicas obtidas nas simulações de Monte Carlo da água usando os modelos TIP4P e TIP5P.

p (bar)	T (°C)	-E/N (kcal mol ⁻¹)		C_p (cal mol ⁻¹ K ⁻¹)		$10^6 \kappa$ (atm ⁻¹)		$10^5 \alpha$ (K ⁻¹)	
		TIP4P	TIP5P	TIP4P	TIP5P	TIP4P	TIP5P	TIP4P	TIP5P
100	150	8,33	7,61	14,27	16,84	112,8	176,2	183,8	238,0
1100	300	6,66	5,93	12,69	11,90	171,8	172,7	183,3	160,7
1300	400	6,75	4,90	12,07	8,81	120,5	297,8	153,3	163,6
1900	150	8,65	8,07	9,59	16,83	35,7	59,1	49,14	124,7
1970	300	7,00	6,30	10,50	10,18	65,92	118,5	94,45	132,2
2800	300	7,21	6,51	8,39	8,81	53,5	69,3	62,19	76,16
3400	400	6,36	5,81	7,02	27,14	60,15	130,0	57,8	233,4
500	300	6,23	5,34	16,18	18,60	349,2	693,3	304,8	470,6
500	400	4,305	3,61	26,91	14,58	2986	2447	996,5	540,6

A densidade é a uma propriedade para a qual dispõe-se de resultados experimentais e é apresentada na Tabela 3.4 com os seus respectivos erros na última coluna. Analisando a dependência com a pressão, verifica-se que os modelos foram capazes de prever o aumento da densidade com o aumento da pressão. O modelo TIP4P, gerou melhor predição da densidade, visto que o erro relativo a este modelo foi aproximadamente a metade daquele relativo ao modelo TIP5P (Tabela 3.4).

Tabela 3.4: Densidade (g cm^{-3}) da água em condições extremas. Os valores entre parênteses são os erros percentuais com relação aos dados experimentais.

p (bar)	T ($^{\circ}\text{C}$)			
Dependência com a pressão		TIP5P	TIP4P	Experimental ¹⁰⁵
2800	300	0,87438 (5,1)	0,90252 (2,0)	0,92092
1970	300	0,81481 (7,9)	0,85448 (3,5)	0,88504
1100	300	0,73293 (11,8)	0,77221 (7,1)	0,83122
500	300	0,60180 (22,6)	0,69931 (10,0)	0,77740
Dependência com a temperatura		TIP5P	TIP4P	Experimental
500	300	0,60180 (22,6)	0,69931 (10,0)	0,77740
500	400	0,35600 (38,6)	0,42193 (27,3)	0,58006
Outras condições extremas		TIP5P	TIP4P	Experimental
100	150	0,85758 (6,9)	0,89540 (2,6)	0,92092
1300	400	0,61265 (16,4)	0,79613 (8,7)	0,73255
1900	150	0,97558 (2,3)	0,99272 (0,6)	0,99866
3400	400	0,81993 (6,1)	0,83892 (3,9)	0,87308

A dependência da densidade com a temperatura está apresentada na Figura 3.1 para pressão de 1 atm e variando-se a temperatura em passos de 30°C , desde a temperatura ambiente até 300°C .

¹⁰⁵ A. K. Soper, F. Bruni, M. A. Ricci *J. Chem. Phys.* 106, 247-254 (1997).

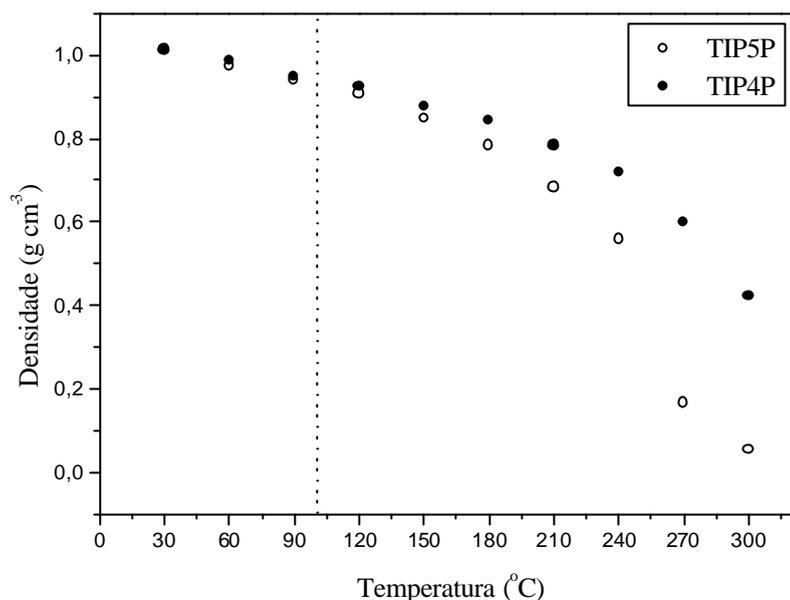


Figura 3.1: Variação da densidade da água com o aumento da temperatura usando os modelos TIP5P e TIP4P.

A análise da Figura 3.1 mostra resultados semelhantes para ambos os modelos até temperaturas próximas a 100°C. Cabe ressaltar que pelo fato de estarmos usando um *ensemble* NpT, os resultados após 100°C não caracterizam transição de fase ou a fase gasosa, pois para tal seria necessária simulação com *ensemble* gran-canônico. Apesar de não terem significado físico, estes resultados apresentam indicações da estabilidade do potencial de interação, que no caso do modelo TIP5P apresenta uma descontinuidade entre 240 e 270°C. Estes resultados podem ser significativos para as simulações em condições extremas, como é o caso da baixa densidade obtida para a água a 300°C e 100 bar com a utilização do modelo TIP5P, observado na Tabela 3.2.

3.2.3. Funções de Distribuição Radial

Um dos principais indicativos da qualidade de um modelo de interação intermolecular é a reprodução das funções de distribuição radial (*radial distribution functions* – RDFs) em líquidos. As RDFs calculadas e experimentais¹⁰⁶ para os pares: g(O-H), g(H-H) e g(O-O) em condições ambiente (25°C e 1 atm) são ilustradas nas Figuras 3.2, 3.3 e 3.4.

¹⁰⁶ A. K. Soper, F. Bruni, M. A. Ricci, *J. Chem. Phys.* 106, 247-254 (1997).

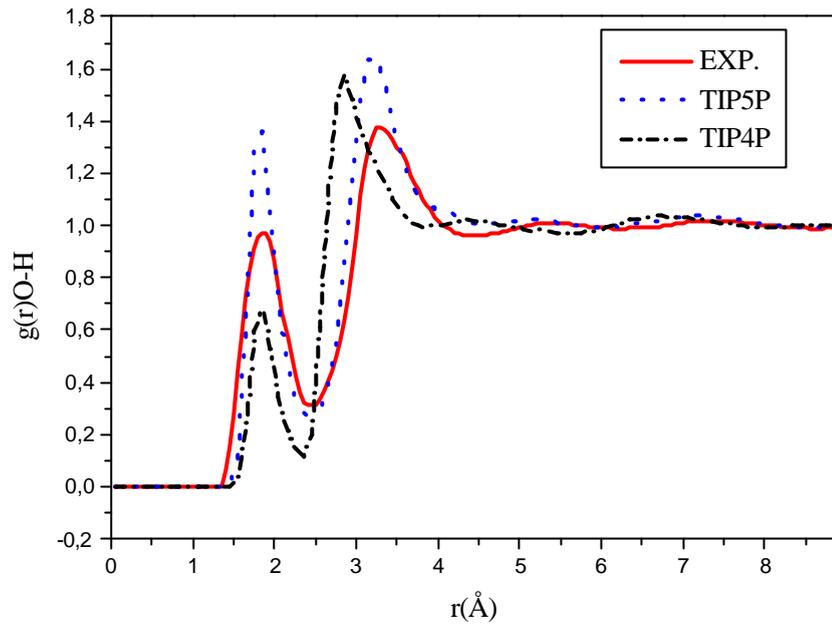


Figura 3.2: RDFs O-H para a água a 1 atm 25°C.

Os dois modelos predizem bem o primeiro pico (1,85 \AA) para a RDF O-H (Figura 3.2), mas para o segundo pico o modelo TIP4P (2,85 \AA) apresentou valor menor que o experimental (3,3 \AA), já o modelo TIP5P (3,25 \AA) manteve desempenho satisfatório.

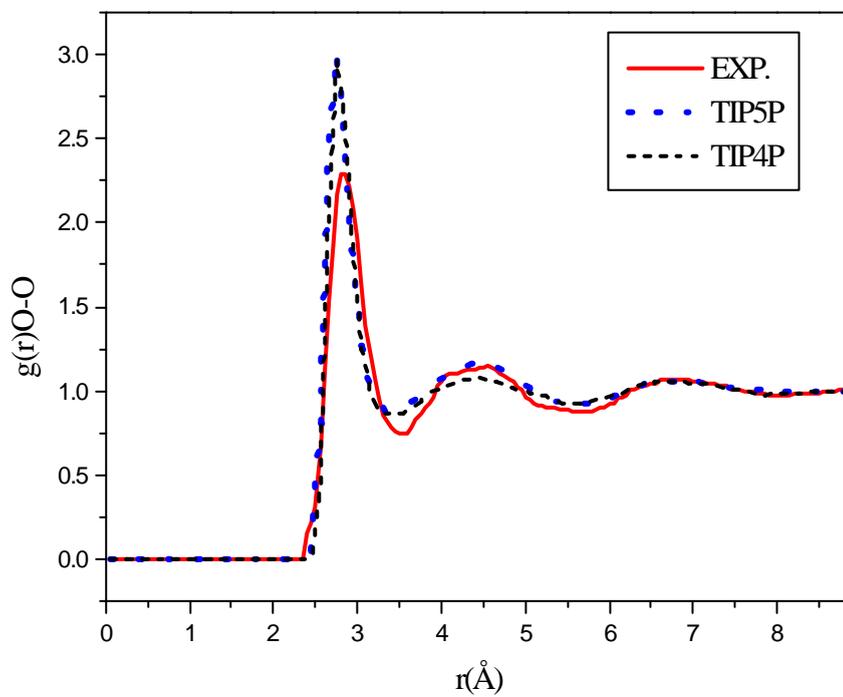


Figura 3.3: RDFs O-O para a água a 1 atm 25°C.

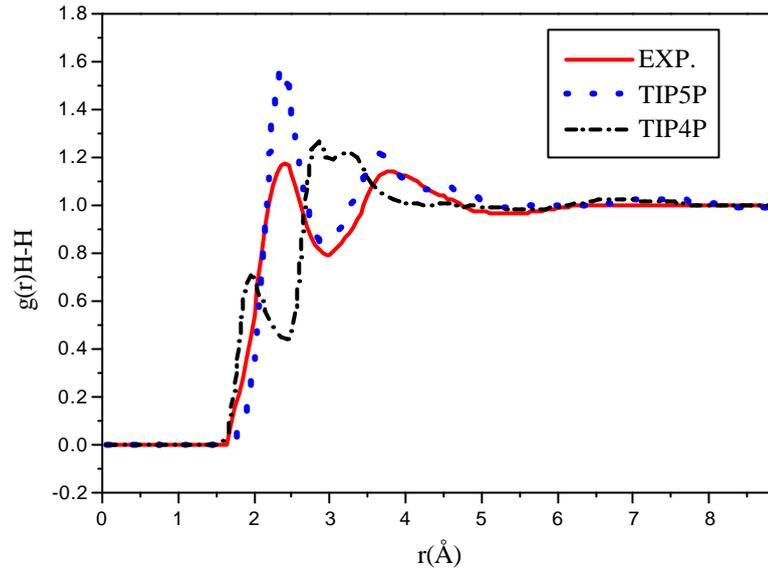


Figura 3.4: RDFs H-H para a água a 1 atm 25°C.

As RDFs H-H (Figura 3.4) obtidas com o modelo TIP5P e experimentais estão em boa concordância, contudo o modelo TIP4P gerou RDFs com estruturas muito distintas das dos resultados experimentais e do modelo TIP5P.

A avaliação do desempenho dos modelos foi feita para algumas condições próximas da crítica, cujos dados experimentais estão disponíveis para as funções de distribuição radial. Os resultados ilustrados nas Figuras 3.5 e 3.6 foram obtidos para uma temperatura de 400°C e pressão de 500 bar.

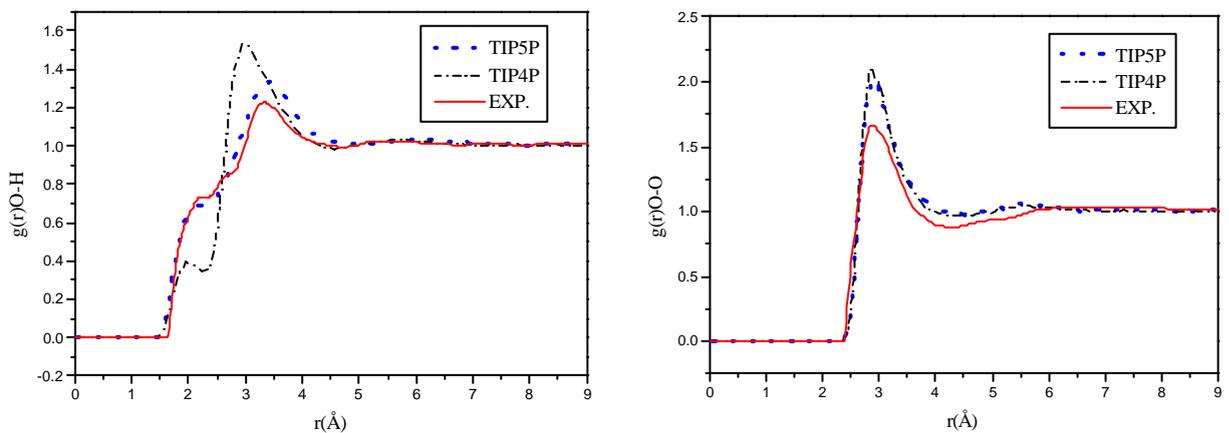


Figura 3.5: RDFs O-H e O-O para a água a 400°C e 500 bar.

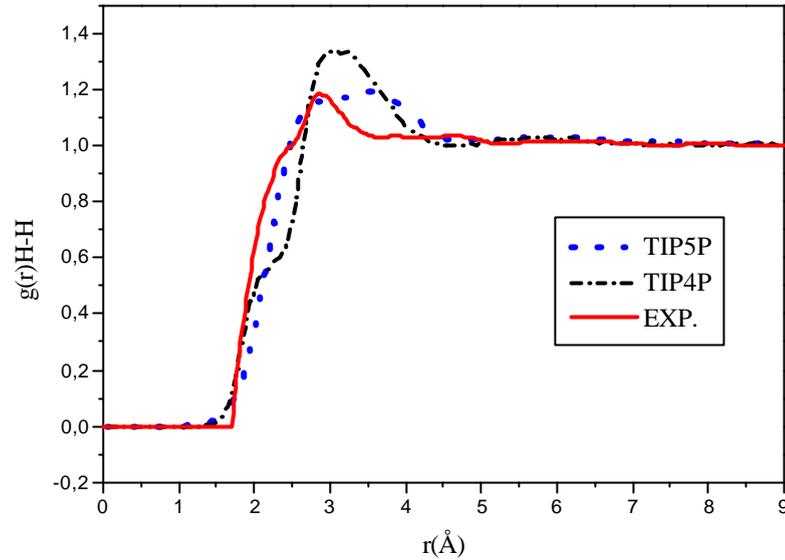


Figura 3.6: RDFs H-H para a água a 400°C e 500 bar.

A Figura 3.5 apresenta as RDFs O-O que mostra uma superposição quase perfeita dos modelos TIP4P e TIP5P, com pequenas diferenças da RDF experimental. Contudo, as RDFs O-H (Figura 3.5) e H-H (Figura 3.6) são muito distintas para os modelos TIP4P e TIP5P, sendo este último o mais adequado quando comparado com os dados experimentais.

Portanto, as funções de distribuição radial geradas nas simulações, tomando-se em consideração todas as condições de temperatura e pressão investigadas, mostram que o modelo TIP5P apresentou melhor desempenho quanto a obtenção de propriedades estruturais.

Dependência com a Pressão

A dependência com a pressão foi verificada simulando-se três pontos à 300°C, temperatura próxima, mas abaixo da crítica. Os três pontos correspondem às pressões de 500, 1100 e 1500 bar (todas acima da pressão crítica: 221 bar). As Figuras 3.7, 3.8 e 3.9 ilustram as RDFs O-H, O-O e H-H, respectivamente.

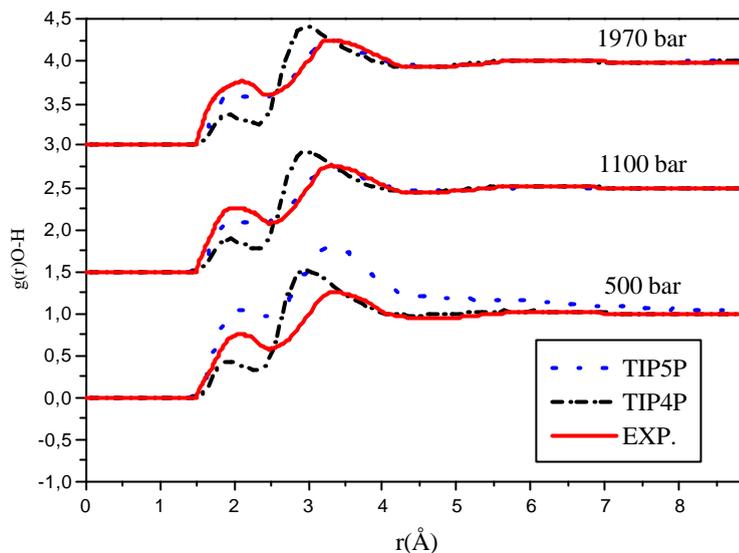


Figura 3.7: RDFs O-H para a água a 300°C e pressões iguais a 500, 1100 e 1970 bar.

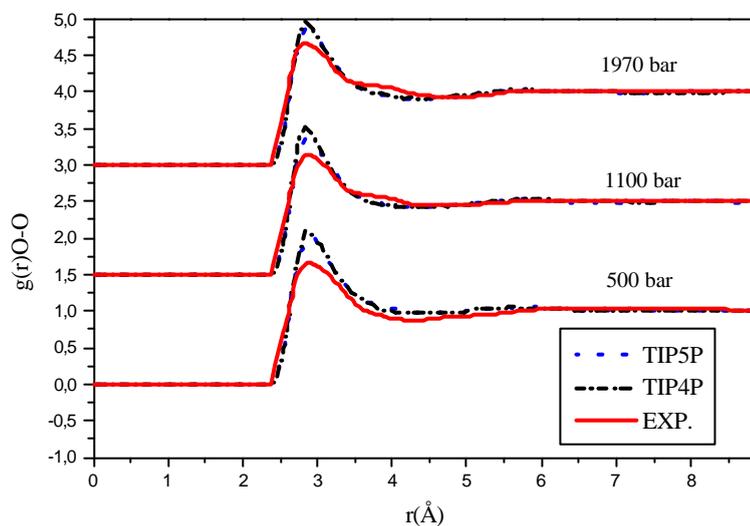


Figura 3.8: RDFs O-O para a água a 300°C e pressões iguais a 500, 1100 e 1970 bar.

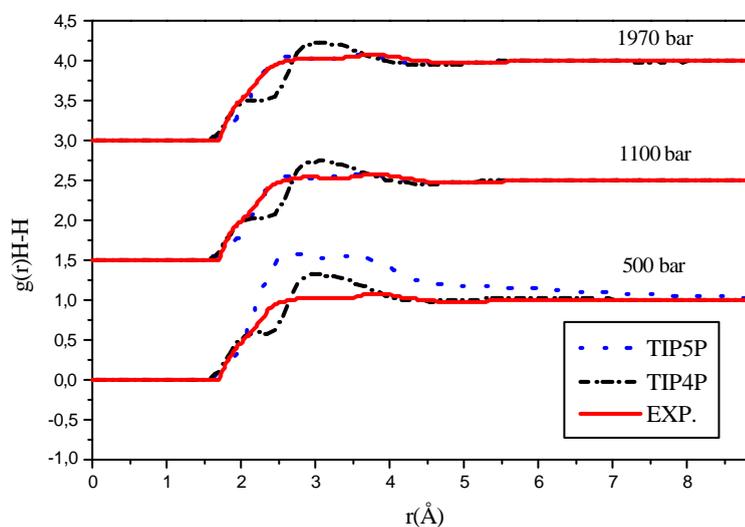


Figura 3.9: RDFs H-H para a água a 300°C e pressões iguais a 500, 1100 e 1970 bar.

A análise da Figura 3.8 indica que ambos os modelos reproduzem as RDFs O-O satisfatoriamente, contudo, para as RDFs O-H e H-H, o modelo TIP5P mostrou-se mais adequado, principalmente para a descrição do segundo pico. Além disso, o modelo TIP4P prediz um ombro mais saliente para as RDFs H-H, que aqueles presentes nos resultados experimentais e modelo TIP5P.

Dependência com a Temperatura

Analisou-se a dependência com a temperatura simulando-se dois pontos à 48 bar e temperaturas iguais a 50 e 125°C. As RDFs H-H obtidas a partir dos modelos TIP4P e TIP5P são apresentadas na Figura 3.10.

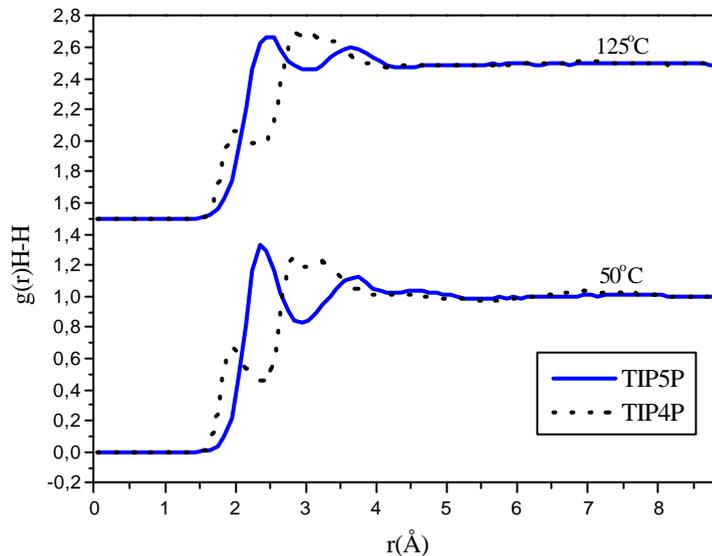


Figura 3.10: RDFs H-H para a água a 48 bar e temperaturas 50 e 125°C.

As diferenças são qualitativa e quantitativamente significativas, e as comparações com os dados experimentais serão fundamentais para estabelecer a confiabilidade dos modelos, contudo, para ambos os modelos observa-se uma esperada redução na estrutura com o aumento da temperatura, com suavização da separação entre os primeiro e segundo picos.

3.2.4. Estatística das Ligações de Hidrogênio

Sabe-se que à temperatura ambiente, a água apresenta propriedades peculiares, provenientes principalmente da presença de fortes interações do tipo ligações de hidrogênio. Em temperaturas e pressões elevadas, a densidade da água diminui e também

sua tensão superficial. A análise da presença e a estimativa das forças das ligações de hidrogênio naquelas condições pode fornecer explicações para estas propriedades. Os resultados que serão apresentados a seguir referem-se às simulações com o modelo TIP5P da água e as restrições para formação de ligações de hidrogênio utilizadas por Chialvo & Cummings (1996)¹⁰⁷. Segundo este critério verificamos que a distribuição angular das ligações de hidrogênio tem o comportamento apresentado na Figura 3.11.

A estatística foi realizada sobre uma caixa equilibrada com 512 moléculas.

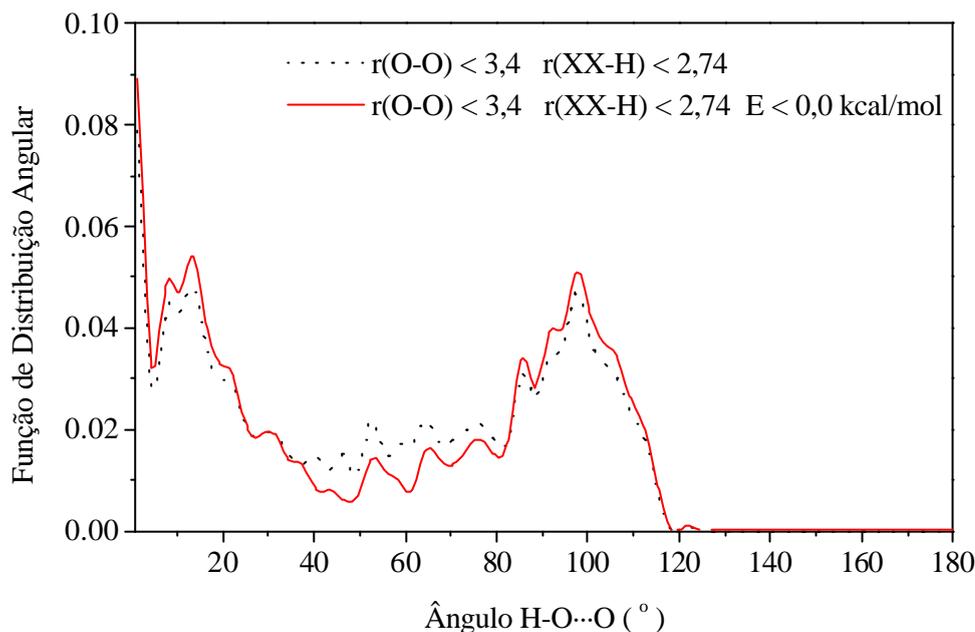


Figura 3.11: Distribuição angular das ligações de hidrogênio O-H...O da água em concordância com os critérios geométricos e energéticos.

Os ângulos menores, entre 15 e 35°, são consistentes com as restrições de distâncias, assim como os ângulos maiores, próximos de 100°. A restrição energética promove a exclusão de ângulos intermediários, entre 40 e 90°, aceitos pelo critério geométrico. Isto nos informa acerca da credibilidade dos critérios adotados, dentro da diversidade de critérios para a quantificação de uma ligação de hidrogênio.

O gráfico apresentado na Figura 3.12 indica a distribuição de moléculas por número de ligações de hidrogênio nas condições ambiente e os histogramas das Figuras 3.13.a-i apresentam este estudo para algumas condições extremas.

¹⁰⁷ A. A. Chialvo, P. T. Cummings, *J. Phys. Chem.* 100, 1309-1316 (1996).

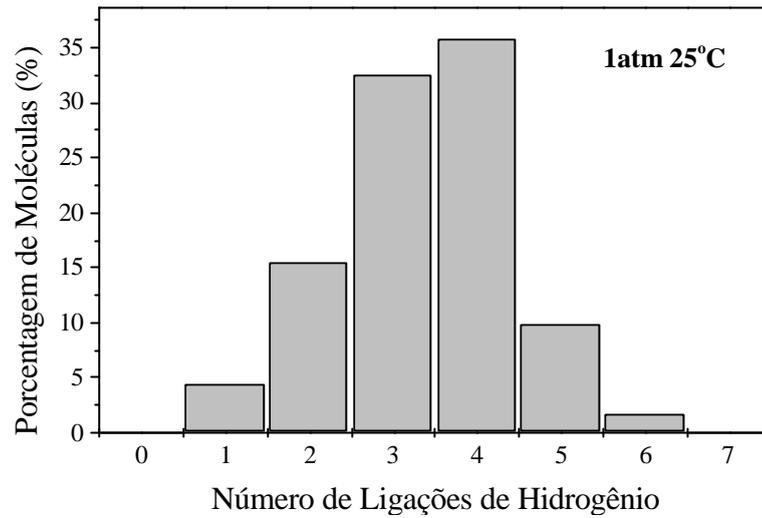
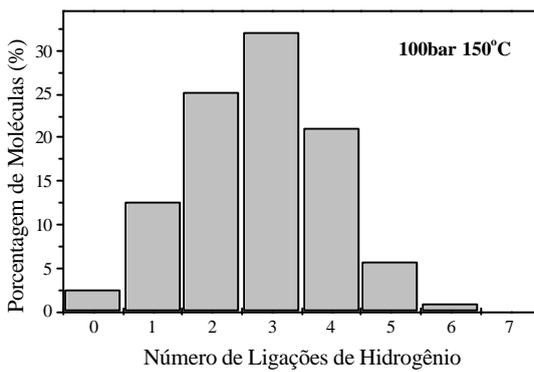
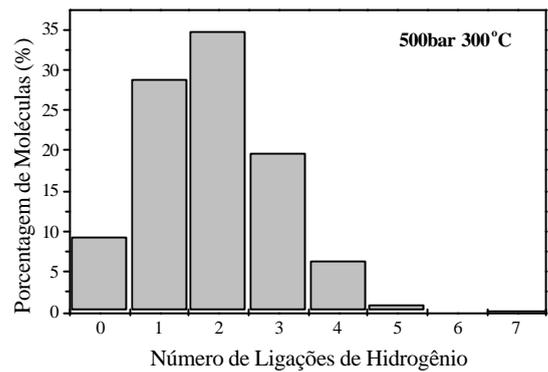


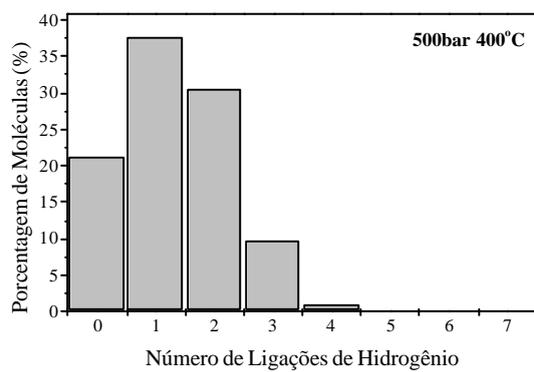
Figura 3.12: Porcentagem de moléculas que formam de uma a sete ligações de hidrogênio. Condições ambiente. Total de ligações de hidrogênio: 861.



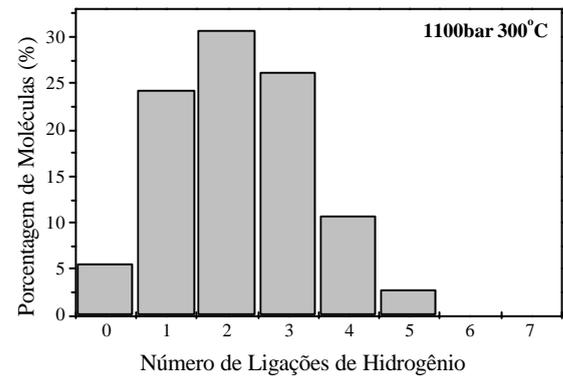
3.13.a. $p = 100 \text{ bar}$ e $T = 150^\circ\text{C}$



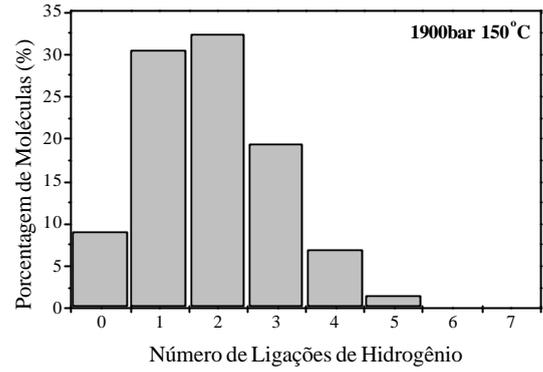
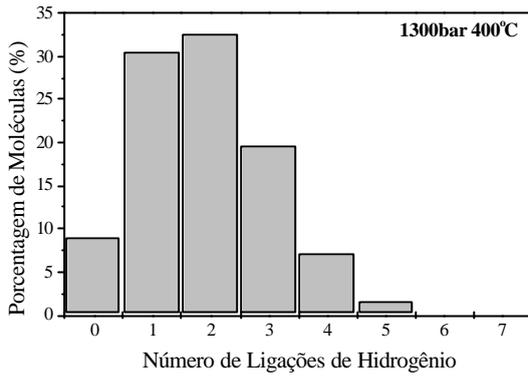
3.13.b. $p = 500 \text{ bar}$ e $T = 300^\circ\text{C}$



3.13.c. $p = 500 \text{ bar}$ e $T = 400^\circ\text{C}$

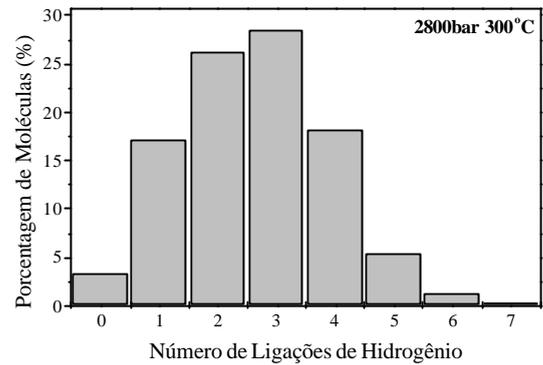
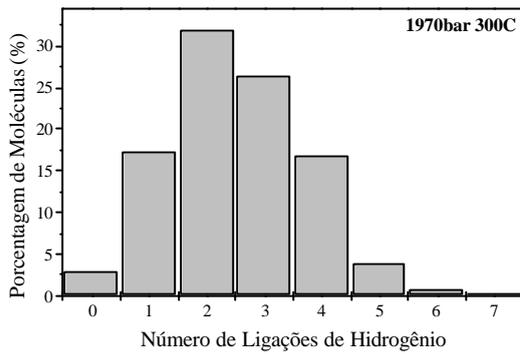


3.13.d. $p = 1100 \text{ bar}$ e $T = 300^\circ\text{C}$



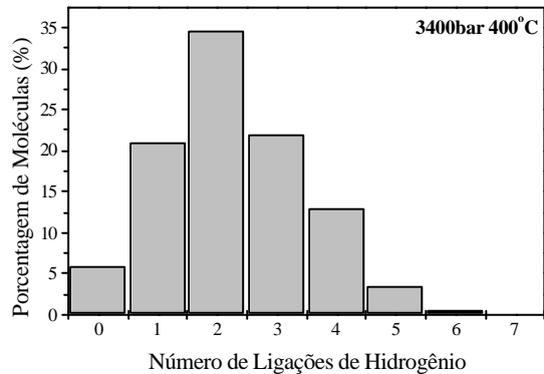
3.13.e. $p = 1300$ bar e $T = 400^\circ\text{C}$

3.13.f. $p = 1900$ bar e $T = 150^\circ\text{C}$



3.13.g. $p = 1970$ bar e $T = 300^\circ\text{C}$

3.13.h. $p = 2800$ bar e $T = 300^\circ\text{C}$



3.13.i. $p = 3400$ bar e $T = 400^\circ\text{C}$

Figura 3.13.a-i: Porcentagem de moléculas que formam de uma a sete ligações de hidrogênio. Diferentes condições de temperatura e pressão.

A Tabela 3.5 apresenta o número de ligações de hidrogênio para cada condição de temperatura e pressão.

Tabela 3.5: Número médio de ligações de hidrogênio sobre 500 configurações para cada condição de temperatura e pressão simulada.

p (bar)	T (°C)	Restrições	Número de ligações de hidrogênio
1	25	Geométrica	956
		Geométrica + Energética	861
100	150	Geométrica	802
		Geométrica + Energética	708
1100	300	Geométrica	661
		Geométrica + Energética	565
1300	400	Geométrica	557
		Geométrica + Energética	486
1900	150	Geométrica	893
		Geométrica + Energética	752
2800	300	Geométrica	834
		Geométrica + Energética	672
3400	400	Geométrica	708
		Geométrica + Energética	582
500	300	Geométrica	548
		Geométrica + Energética	484
500	400	Geométrica	378
		Geométrica + Energética	336

Analisando a Tabela 3.5 verifica-se que o maior número de ligações de hidrogênio estão presentes em condições ambiente. O aumento da pressão aumenta as interações por ligações de hidrogênio. Estes dados podem ser visualizados graficamente na Figura 3.14 para a temperatura de 300°C. O comportamento verificado é crescente até a pressão de ~3500 bar, a partir da qual verifica-se saturação do número de ligações de hidrogênio.

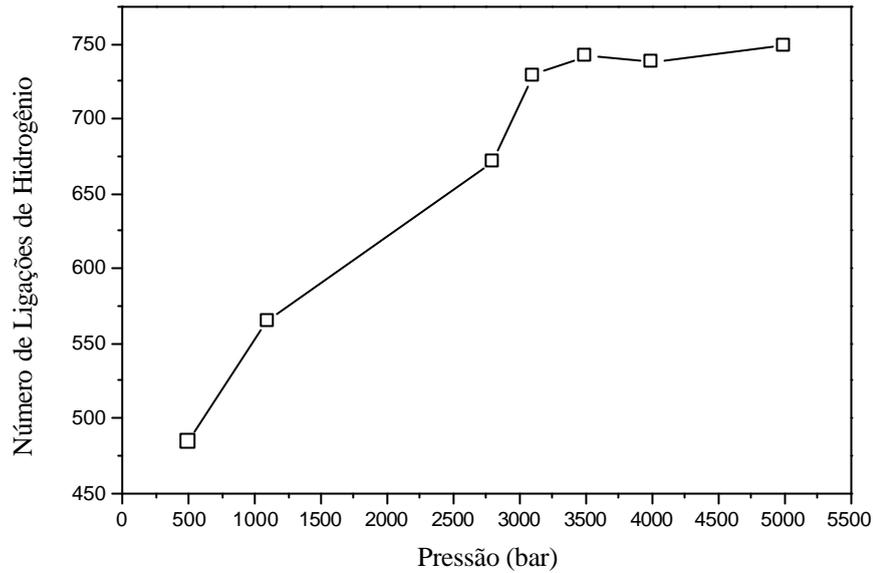


Figura 3.14: Dependência do número de ligações de hidrogênio com a pressão para $T = 300^{\circ}\text{C}$.

O número de ligações de hidrogênio diminui com o aumento da temperatura (Tabela 3.5). Este análise foi também realizada para temperaturas acima da ambiente em intervalos de 30°C e pressão de 1 atm ilustrada na Figura 3.15. Como pode-se observar, usando o modelo TIP5P da água para a simulação, a dependência do número de ligações de hidrogênio é decrescente.

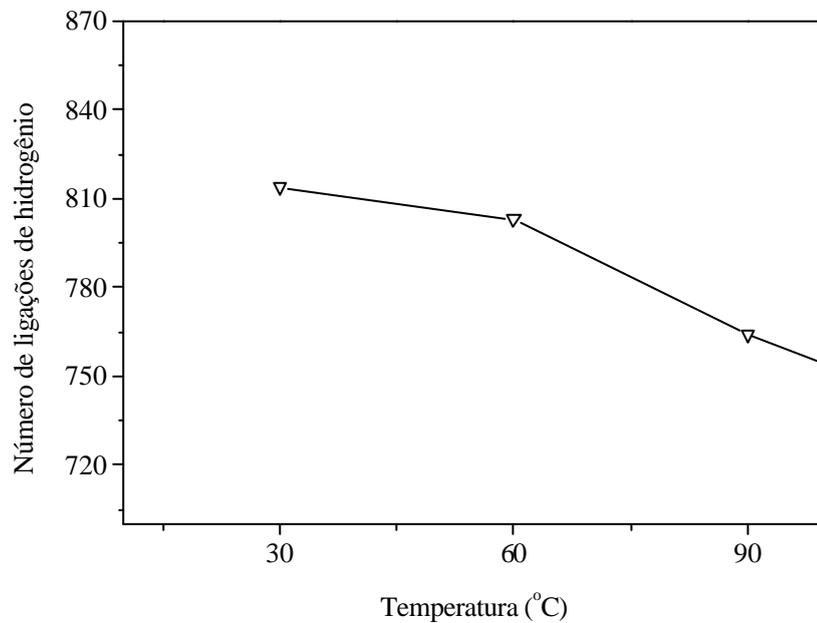
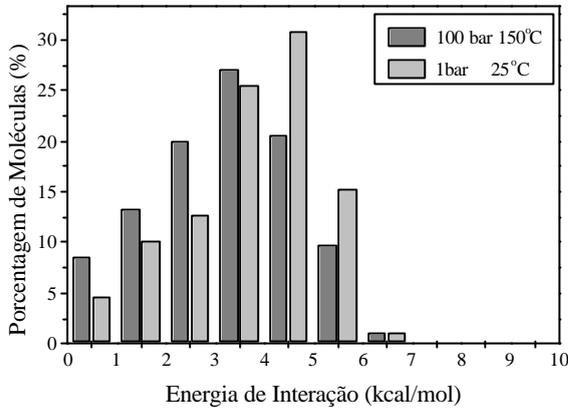
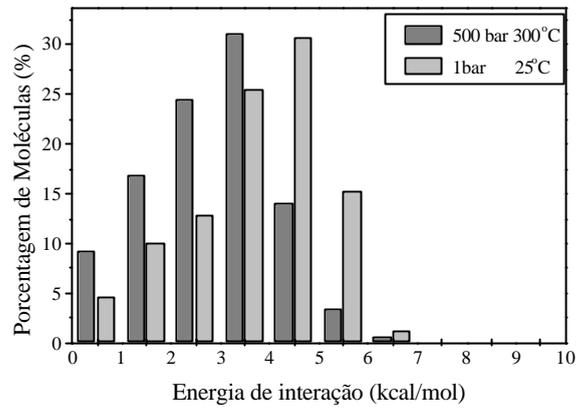


Figura 3.15: Dependência do número de ligações de hidrogênio com a temperatura para $p = 1 \text{ atm}$.

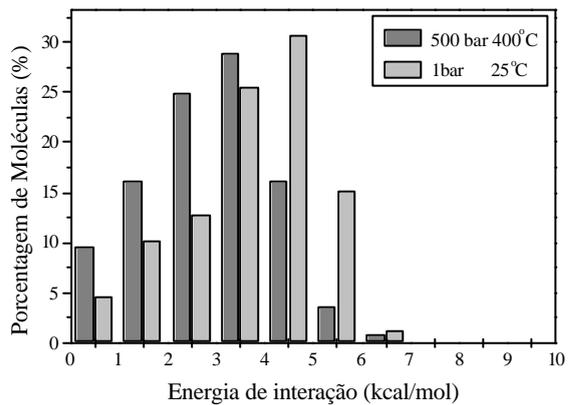
A distribuição das energias das ligações de hidrogênio apresentadas na Tabela 3.5 estão ilustradas nas Figuras 3.16.a-3.16.i, onde se compara o comportamento nas condições ambiente e extremas. Nestes gráficos representa-se a fração de moléculas em função das energias de interação das ligações de hidrogênio entre 0 e 10,0 kcal mol⁻¹, em passos de 1,0 kcal mol⁻¹.



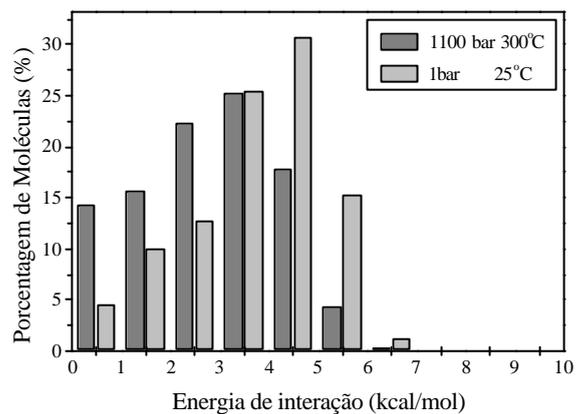
3.16.a. p = 100 bar e T = 150°C



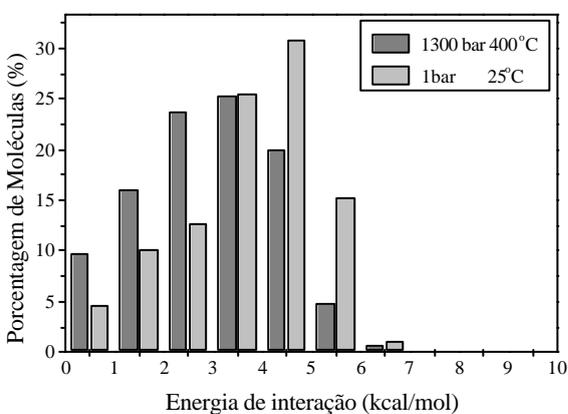
3.16.b. p = 500 bar e T = 300°C



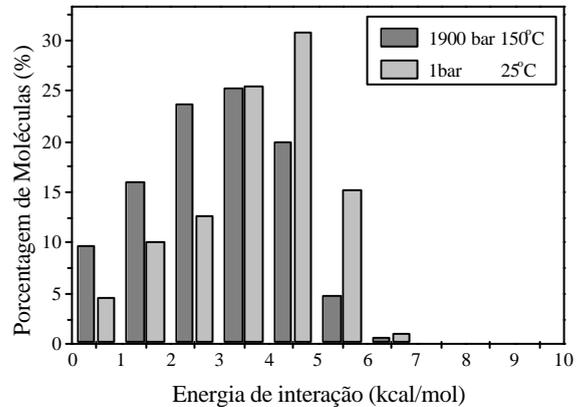
3.16.c. p = 500 bar e T = 400°C



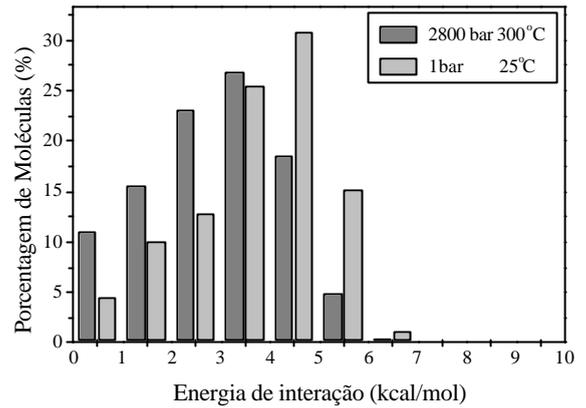
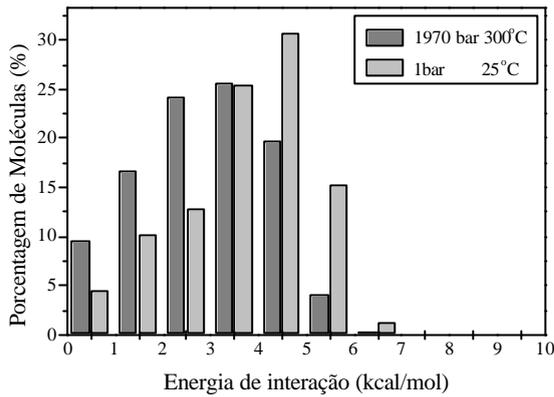
3.16.d. p = 1100 bar e T = 300°C



3.16.e. p = 1300 bar e T = 400°C

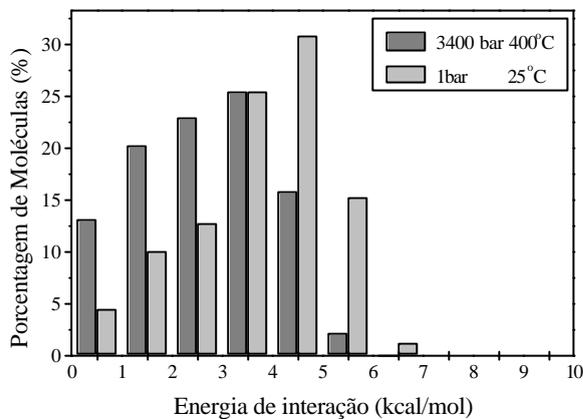


3.16.f. p = 1900 bar e T = 150°C



3.16.g. p = 1970 bar e T = 300°C

3.16.h. p = 2800 bar e T = 300°C



3.16.i. p = 3400 bar e T = 400°C

Figura 3.16.a-i: Porcentagem de moléculas que formam ligações de hidrogênio em função das energias de interação de ligação de hidrogênio.

As energias médias das interações com ligações de hidrogênio para cada uma das condições analisadas e a energia de coesão do líquido por molécula são apresentadas na Tabela 3.6. A energia de coesão corresponde a energia de interação de cada molécula com as demais.

Pode-se concluir a partir das Figuras 3.16 que em altas temperaturas e pressões, além de se formar um número menor de ligações de hidrogênio, a maior parte destas ligações são mais fracas (com média entre 2,7 e 3,2 kcal mol⁻¹) que aquelas formadas à temperatura ambiente (com média igual a 3,7 kcal mol⁻¹). Este comportamento se reflete nas energias de coesão do líquido.

Tabela 3.6: Energia média das interações com ligações de hidrogênio e energia de coesão para cada condição de temperatura e pressão simulada.

p (bar)	T (°C)	Energia média das interações com ligações de hidrogênio (kcal mol ⁻¹)	Energia de coesão por molécula (kcal mol ⁻¹)
1	25	-3,7	-9,96
100	150	-3,2	-7,61
500	300	-2,9	-5,34
500	400	-2,9	-3,61
1100	300	-2,8	-5,93
1300	400	-3,0	-4,90
1900	150	-3,0	-8,07
1970	300	-2,9	-6,30
2800	300	-2,9	-6,51
3400	400	-2,7	-5,82

3.2.5 Análise Topológica das Redes de Ligações de Hidrogênio em Água

A determinação do número de ligações de hidrogênio em altas temperaturas indica uma diminuição da existência destas forças de interação. Desse modo é razoável esperar que nestas condições ocorra a formação de agregados isolados e espaços vazios. Preliminarmente, isso foi verificado pela visualização das caixas de simulação após equilíbrio, como ilustradas nas Figuras 3.17 - 3.18.

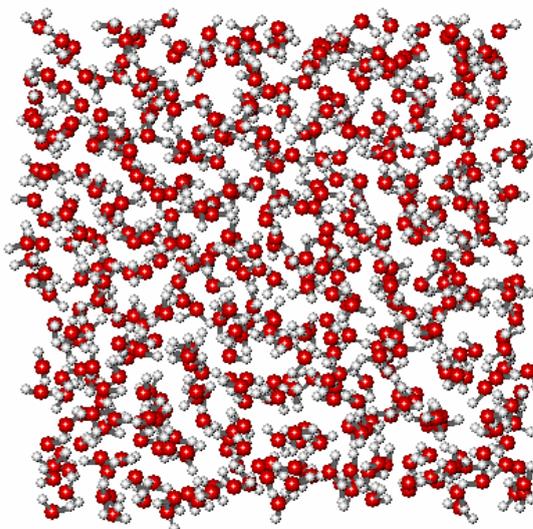


Figura 3.17: Caixa de simulação com 512 moléculas após equilíbrio à $T = 25^{\circ}\text{C}$ e $p = 1$ atm. O volume da caixa é $(24,75 \text{ \AA})^3$.

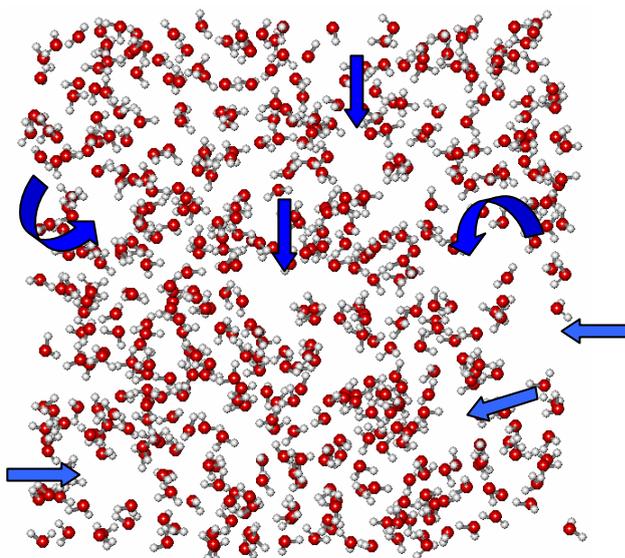


Figura 3.18: Caixa de simulação com 512 moléculas após equilíbrio. $p = 500$ bar e $T = 400^\circ\text{C}$. O volume da caixa é $(34,84 \text{ \AA})^3$.

A análise da Figura 3.18 indica a presença de espaços vazios e agregados (setas). Já a Figura 3.17 anterior mostra uma distribuição homogênea.

Deve-se ressaltar que a análise acima é apenas qualitativa. Para uma quantificação das propriedades topológicas destes agregados verificou-se a presença ou não de padrões de *small-world* através do cálculo das propriedades topológicas locais: coeficiente de agregação ou *clustering* (C), comprimento do caminho médio (L) e grau de distribuição ($P(k)$) e a propriedade global: densidade espectral média. As definições destas propriedades estão apresentadas no item 2.7.

A Tabela 3.7 apresenta os resultados para o coeficiente de agregação, comprimento do caminho médio e o número de conexões médio calculados para 500 estruturas de redes de ligações de hidrogênio nas condições ambiente e em uma condição supercrítica (500 bar 400°C) usando caixas de simulação com 300 e 512 moléculas de água, ou seja, $N = 300$ e $N = 512$.

Observa-se que tanto os parâmetros de caracterização locais (L e C) das redes como a conectividade média ($\langle k \rangle$) são independentes do número de moléculas (entre 300 e 512) de água presente na caixa de simulação (Tabela 3.7). À temperatura ambiente, as redes de ligação de hidrogênio da água líquida tem $L \gg L_{rand}$ e $C \gg C_{rand}$, o que sugere que o comprimento do caminho da rede aumenta e também a sua agregação. Estas são condições características de um agregado gigante ocorrendo percolação¹⁰⁸.

¹⁰⁸ R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* 74, 47 (2002).

Tabela 3.7: Comprimento do caminho médio (L and L_{rand}), coeficiente de agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio.

	25°C e 1 bar		400°C e 500 bar	
	$N = 512$	$N = 300$	$N = 512$	$N = 300$
L	20,36	19,65	3,33	3,30
L_{rand}	5,23	4,90	23,78	31,28
L/L_{rand}	3,9	4,0	0,14	0,11
C	0,5456	0,5302	0,2114	0,2057
$C_{rand} = p$	0,0064	0,0106	0,0025	0,0041
C/C_{rand}	85,2	50,0	84,5	50,2
$\langle k \rangle$	3,3	3,2	1,3	1,2

Na condição supercrítica, as redes apresentam tanto menores comprimentos, como também o número de conexões (conectividade média $\langle k \rangle$), de modo que $L_{rand} \gg L$, ao passo que o coeficiente de agregação é alto, e portanto, indicando uma significativa possibilidade da presença de padrões de *small-world*¹⁰⁹ nas redes de ligações de hidrogênio nesta condição.

A análise mostra ainda que a avaliação de apenas um parâmetro local, como o coeficiente de agregação, não é apropriado para decisão sobre a presença ou não de padrões de *small-world*, que neste caso seria um indicativo para ambas as condições.

A conectividade média apresentado na Tabela 3.7 concorda com o número médio de ligações de hidrogênio, $\langle n_{HB} \rangle$, obtido a partir de simulações recentes¹¹⁰. Em particular, tem sido proposto que a conectividade média na condição supercrítica (1,2-1,3) está abaixo do limiar de percolação¹¹¹, $\sim 1,6$, de modo que não se observa a formação de redes contínuas de ligações de hidrogênio, ao contrário da água à temperatura ambiente. Neste caso o valor é consistente com a ocorrência de percolação. A Figura 3.19 ilustra uma rede de ligações de hidrogênio para uma caixa de simulação à temperatura ambiente. Para comparação, a Figure 3.20 ilustra a mesma rede e seu maior agregado na condição supercrítica (500 bar e 400°C).

¹⁰⁹ D. J. Watts, S. H. Strogatz, *Nature* 393, 440 (1998).

¹¹⁰ A. G. Kalinichev, J. D. Bass, *J. Phys. Chem. A* 101, 9720 (1997).

¹¹¹ R. L. Blumberg, H. E. Stanley, A. Geiger, P. Mousbach, *J. Chem. Phys.* 80, 5230 (1984).

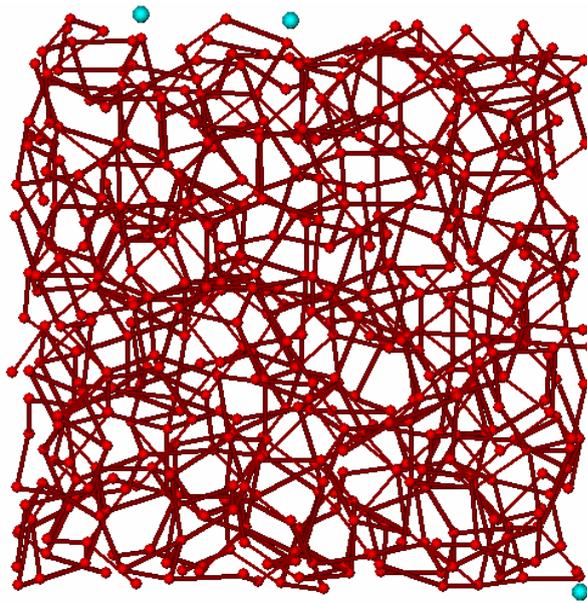


Figure 3.19: Rede de ligações de hidrogênio da água líquida em condições ambiente. Cada círculo (sítio) representa uma molécula de água e as conexões representam as ligações de hidrogênio.

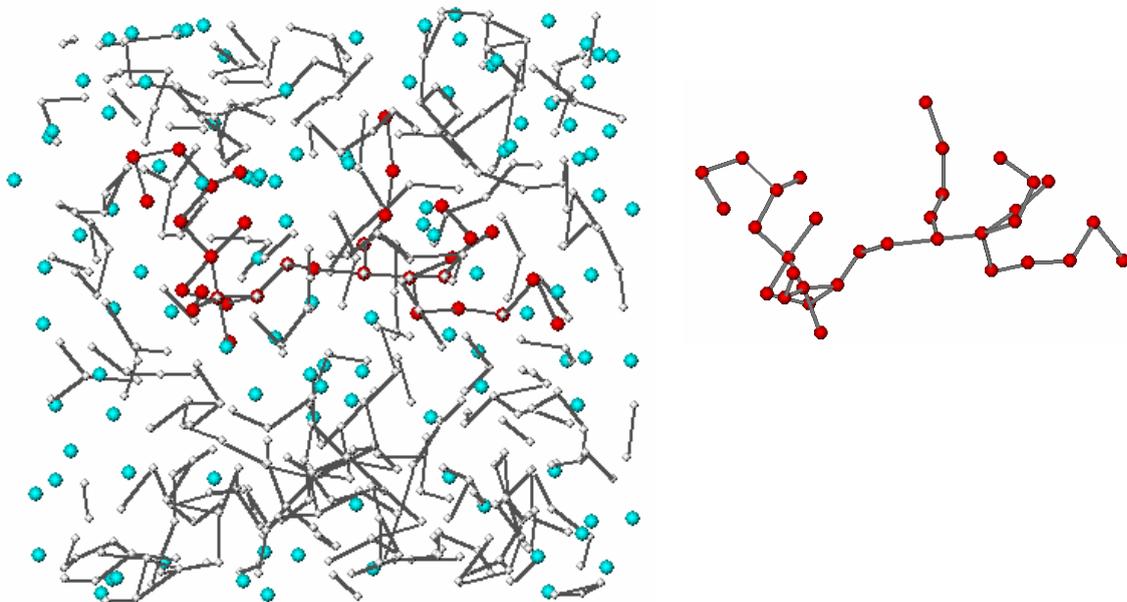


Figure 3.20: Rede de ligações de hidrogênio (esquerda) e o maior agregado da rede (direita) para a água supercrítica. Cada círculo (sítio) representa uma molécula de água e as conexões representam as ligações de hidrogênio.

O grau de distribuição $P(k)$ para as redes de ligações de hidrogênio da água nas condições ambiente e supercrítica são apresentados nas Figuras 3.21 e 3.22, respectivamente. Para um grafo randômico tem-se uma distribuição do tipo binomial. Na

água supercrítica, o comportamento parece similar aqueles de redes larga escala, com a distribuição dada pela lei de potência com função de corte f^{112} . Esta forma da distribuição é esperada quando ocorrem restrições limitando o número de conexões, sendo as mesmas dificilmente formadas. Esta conjectura¹¹² é consistente com os resultados observados para as redes de ligações de hidrogênio da água supercrítica, desde que a formação de uma interação é mais custosa devido à baixa densidade e maiores distâncias intermoleculares, que dessa forma relaciona-se à forma do grau de distribuição com as propriedades termodinâmicas, tal como densidade e entalpia de vaporização. Cabe ressaltar que o grau de distribuição é independente do número de sítios (300 ou 512 moléculas).

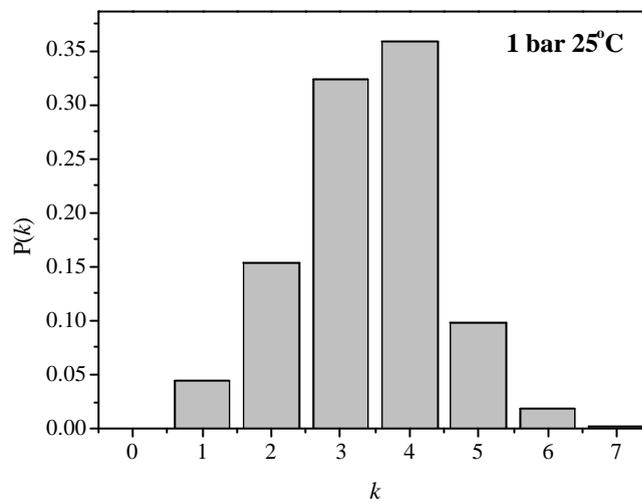


Figura 3.21: Grau de distribuição para rede de ligações de hidrogênio da água líquida nas condições ambiente.

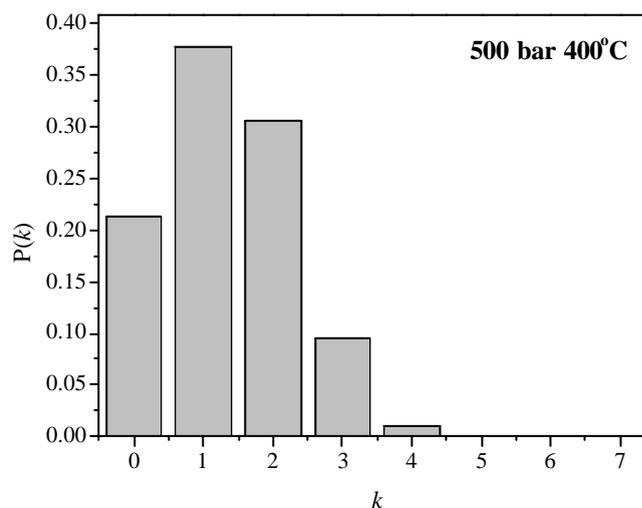


Figura 3.22: Grau de distribuição para rede de ligações de hidrogênio da água supercrítica.

¹¹² L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* 97, 11149 (2000).

Uma análise de agregados usando estatística de ilhas mostra que o cálculo de “ n_s ”, o número de ilhas com “ s ” componentes, resulta em um comportamento claramente distinto para as duas condições analisadas. Para uma melhor visualização gráfica considera-se que para cada “ bin ” tem-se ilhas com número de componentes entre: $2^{bin-1} - (2^{bin} - 1)$. Nas condições ambiente forma-se um grande agregado (contendo em média 99,1% dos componentes) com a existência de uma pequena quantidade de monômeros ou dímeros ($\sim 0,6\%$ em média) e o gráfico tem a forma de blocos opostos de tamanhos pequenos (n_s pequenos), conforme ilustrado na Figura 3.23. Ao passo que na condição crítica analisada observa-se que “ n_s ” decresce com o aumento de “ s ”, ou seja, há um maior número de monômeros, seguido pelo número de dímeros, trímeros, e assim sucessivamente, conforme se observa na Figura 3.24. Nestas condições tem-se em média agregados de tamanho máximo com 26 componentes apenas. Esta observação confirma a existência de comprimentos de caminho curtos verificada na Tabela 3.7. A Tabela 3.8 apresenta o número médio de ilhas de tamanho máximo para as 500 caixas obtidas nas duas condições analisadas.

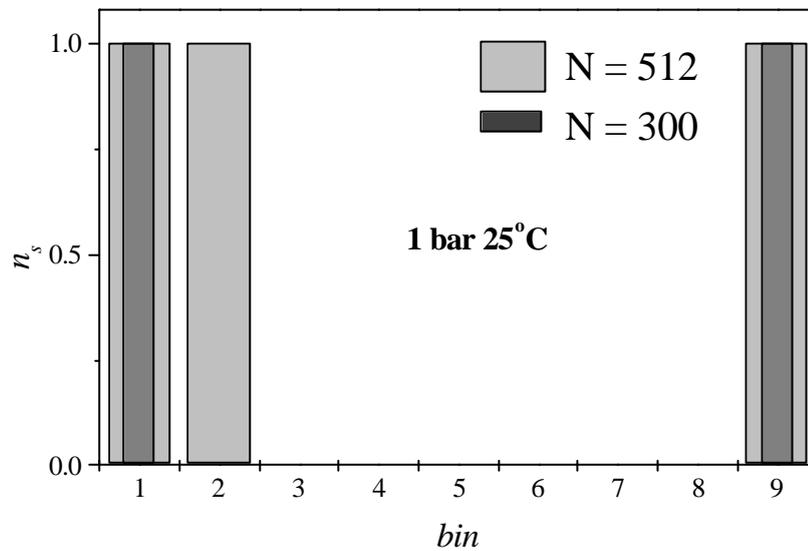


Figura 3.23: Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação da água nas condições ambiente. Para $N = 512$ formam-se apenas três ilhas com 1, 2 e 509 componentes. Para $N = 300$ formam-se duas ilhas com 1 e 299 componentes.

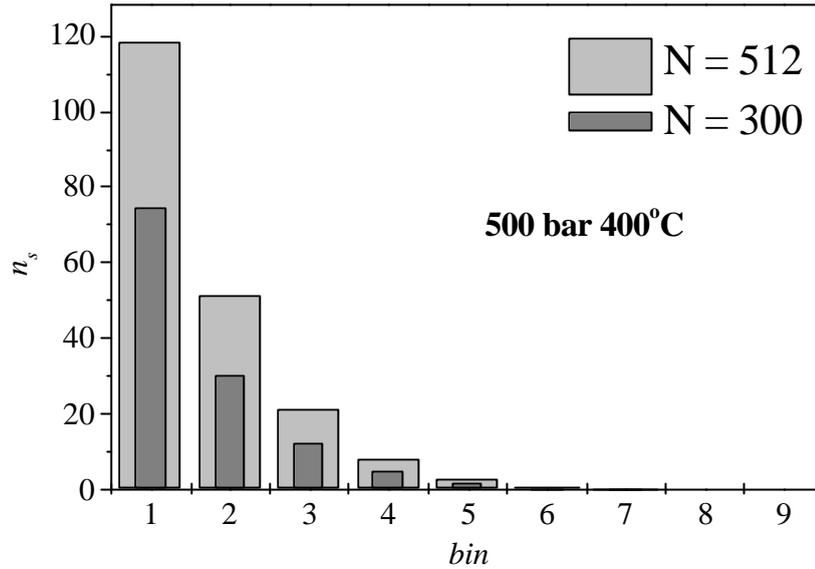


Figura 3.24: Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação da água na condição supercrítica. Para $N = 512$, a maior ilha contém 35 componentes em média (~6,8%) e 118 monômeros de água (~23%). Para $N = 300$, a maior ilha contém 26 componentes em média (~8,7%) e estão presentes 74 monômeros de água (~24%).

Tabela 3.8: Tamanho máximo médio das ilhas para as 500 caixas obtidas nas duas condições analisadas.

p (bar)	T (K)	Ilha de tamanho máximo médio	
		N = 512	N = 300
1	298	508 (99,2%)	297 (99,0%)
500	673	35 (6,8%)	26 (8,7%)

Uma propriedade global de redes complexas é a análise espectral. A densidade espectral reescalada obtida a partir da simulação de 500 redes de ligações de hidrogênio para a água líquida nas condições ambiente e supercrítica são apresentadas nas Figuras 3.25a e 3.25b.

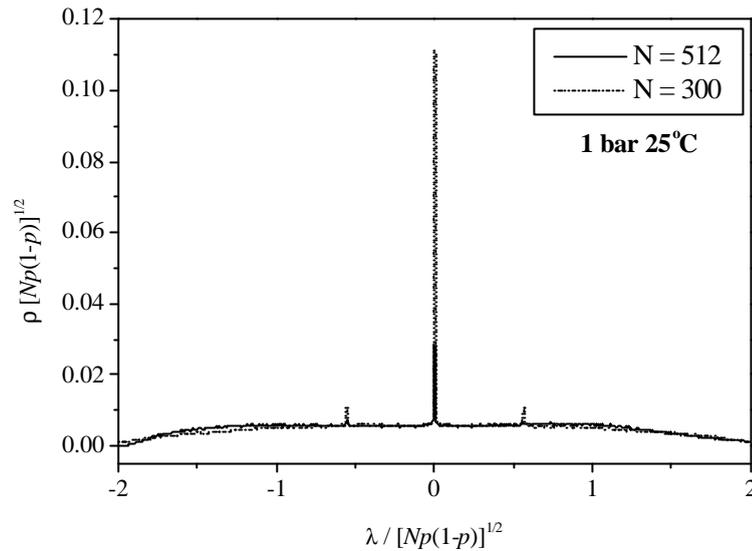


Figura 3.25a: Densidade espectral média reescalada para as redes de ligações de hidrogênio da água nas condições ambiente para caixas de simulação contendo 512 e 300 moléculas.

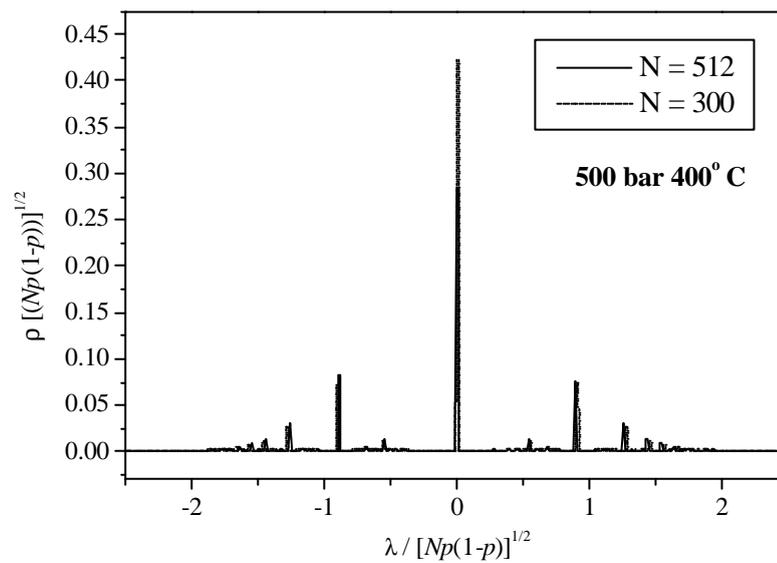


Figura 3.25b: Densidade espectral média reescalada para as redes de ligações de hidrogênio da água na condição extrema 500 bar 400°C, para caixas de simulação contendo 512 moléculas.

Observa-se das Figuras 3.25.a-b que a densidade espectral para a água em condições ambiente é dominada pelo primeiro momento (ver equação 2.39), o que é consistente com a percolação observada, ou seja, a formação de um agregado gigante. Por outro lado, para água supercrítica, a densidade espectral tem contribuições significantes do segundo e terceiro momentos. Isto é consistente com o comportamento de *small-world* das redes de

ligações de hidrogênio nestas condições. Também deve ser observado que quando o número de sítios ou moléculas aumenta de 300 para 512, o momento principal decresce significativamente e densidade espectral tende a um semi-círculo para a condição ambiente e a densidade espectral tem o aspecto similar a de um grafi randômico.

3.3. Ajuste de Superfícies de Energia Potencial para a Água Líquida

Elaborou-se um programa em linguagem FORTRAN para a simulação do dímero da água, utilizando-se o modelo TIP5P. Os dados gerados com este programa foram utilizados para o ajuste de modelos por mínimos quadrados parciais (PLS) e aqueles baseados em redes neurais artificiais (RNAs).

Os dados englobam as regiões ilustradas na Figura 3.26, com as moléculas de água superpostas. Estas configurações de pares foram obtidas através de combinações de rotações e translações no intervalo de distância entre os dois centros de massa das moléculas de água, variados de modo a englobar as principais estruturas do espaço configuracional do dímero. Devido à simetria do sistema, somente uma parte das configurações foi representada.

Foram gerados diferentes conjuntos de treinamento, validação e teste. Variou-se também o número de dados para cada conjunto e optou-se por apresentar os resultados para um dos casos de teste.

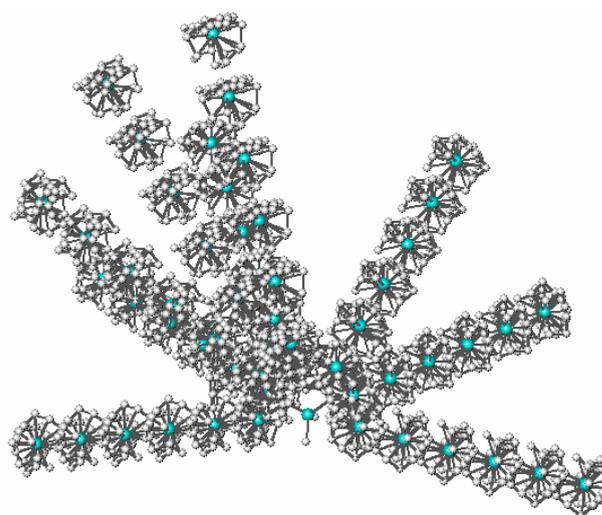


Figura 3.26: Diversas configurações superpostas para o dímero da água (molécula 1 no centro, e a molécula 2 ocupando cada posição superposta ao redor da molécula central).

3.3.1. Análise de Componentes Principais e Análise de Sensibilidade

A análise de componentes principais é uma técnica que permite a diminuição da dimensionalidade dos dados, com menor perda possível da informação sobre o sistema.

Neste trabalho existem vinte e cinco variáveis de distâncias entre os sítios nas moléculas (combinação de 5 sítios em cada molécula), definindo sua configuração espacial para o modelo TIP5P. A entrada para o modelo de potencial de interação depende desta informação. Na verdade a configuração espacial pode ser descrita por nove distâncias (Figura 3.27), numa combinação entre três átomos em cada molécula, visto que os outros 4 são pseudo-átomos (XX, com números 4, 5, 9 e 10) e portanto não fazem parte efetivamente das moléculas. Para reduzir este número de entradas (R16, R17, R18, R26, R27, R28, R36, R37, R38), realizou-se uma análise de componentes principais.

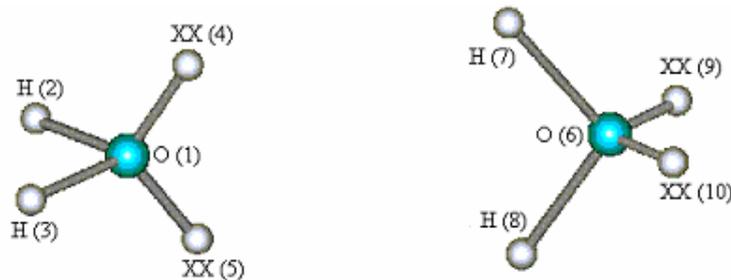


Figura 3.27: Dímero da água. Modelo TIP5P.

Serão apresentados os resultados para dois conjuntos de dados, o primeiro (Figura 3.28), contendo 486 dados, e o segundo (Figura 3.29) para o qual obteve-se uma RNA com desempenho razoável em uma maior faixa de interesse (405 dados).

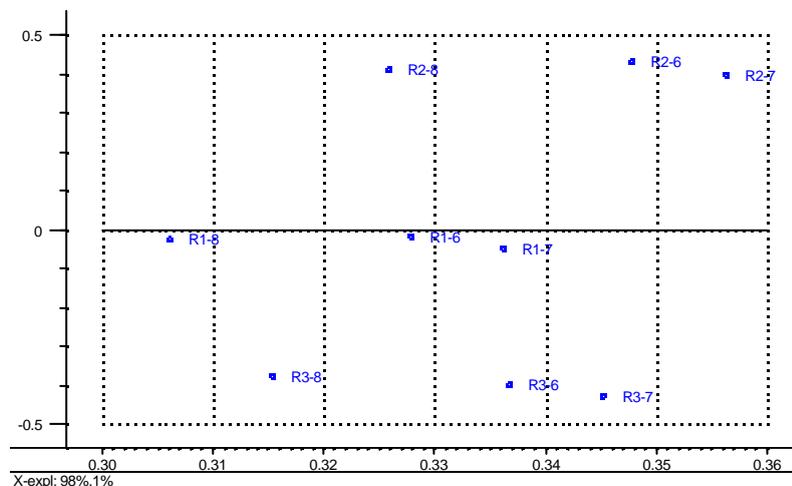


Figura 3.28: Análise de componentes principais para o conjunto de treinamento inicial contendo 486 dados. A abscissa representa a primeira componente e a ordenada a segunda componente principal.

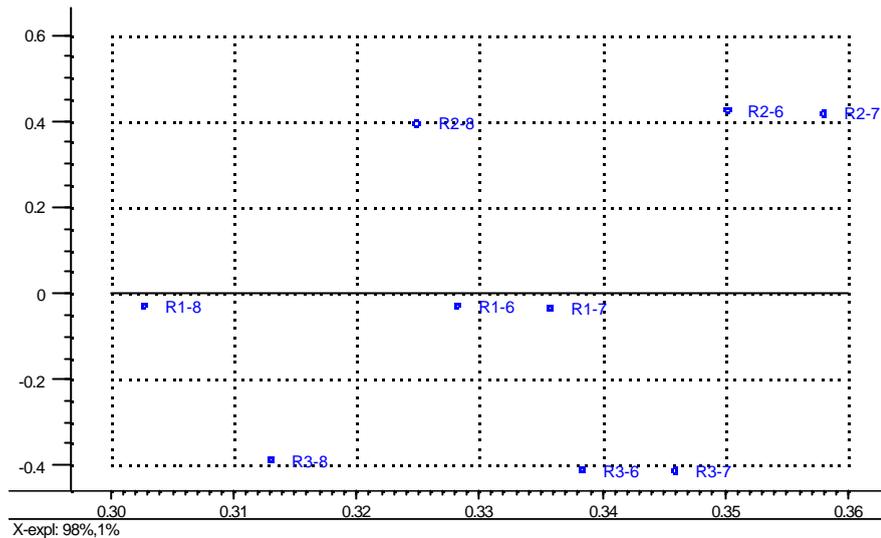


Figura 3.29: Análise de componentes principais para o conjunto de treinamento contendo 405 dados. A abscissa representa a primeira componente e a ordenada a segunda componente principal. X-expl: 98%, 1% significa que a primeira componente captura 98% e a segunda 1% da variância.

Como os dados foram gerados usando a mesma sistemática para os dois conjuntos obtidos, os resultados da análise de componentes principais são semelhantes. Constatou-se ser necessário usar apenas duas componentes principais, com variância igual a 99% (98% capturada pela primeira componente e 1% pela segunda). Baseando-se nos pesos nas componentes principais, reduziu-se as variáveis de entrada para apenas quatro: R16, R26, R27 e R36.

A análise de sensibilidade foi realizada durante o treinamento das RNAs, fornecendo a contribuição relativa de cada variável de entrada no desempenho das RNAs. Isto é realizado testando o modelo de redes neurais com cada variável de entrada separadamente. Os resultados obtidos indicam a seguinte ordem crescente de importância na eficiência dos modelos de RNAs para as quatro variáveis: R27, R36, R16 e R26. O principal parâmetro de checagem é o erro associado a retirada de cada variável, separadamente, sendo um indicativo do peso daquela variável. O indicativo final para exclusão de uma variável é a razão entre o erro citado e o erro quando todas as variáveis são utilizadas (Razão). Se este valor for igual ou menor que a unidade, a variável pode ser descartada sem prejuízo do desempenho do modelo. Esta análise é realizada tanto para o conjunto de treinamento, como para o conjunto de verificação. Os resultados obtidos são apresentados na Tabela 3.9.

Tabela 3.9: Resultados da análise de sensibilidade.

Dados de Treinamento	$(R16)^2$	$(R26)^2$	$(R27)^2$	$(R36)^2$
Erro	8,81	9,13	2,72	8,59
Razão	3,68	3,81	1,14	3,59
Dados de Verificação	$(R16)^2$	$(R26)^2$	$(R27)^2$	$(R36)^2$
Erro	8,39	8,68	2,12	8,20
Razão	4,38	4,53	1,10	4,28

Como podemos observar a variável R27, que corresponde a distância entre átomos de hidrogênio de cada molécula, tem o menor peso na eficiência das RNAs, entretanto como o valor do parâmetro ‘razão’ é um pouco maior que 1, optou-se por incluir esta variável.

3.3.2. Ajuste inicial da energia de interação usando RNAs

Dos 486 dados gerados no espaço configuracional apresentados na Figura 3.26, separou-se 438 dados para o ajuste e 48 dados para teste (Tabela 3.10). Este resultado pode ser comparado com o ajuste usando RNAs (Figura 3.30) e com aquele para o qual a entrada é dada não pelo quadrado da distância, mas pelo inverso deste valor (Figura 3.31).

As restrições para a faixa de aplicação das RNAs treinadas com o conjunto este dados são apresentadas na Tabela 3.11. A capacidade de extrapolação de uma RNA é bastante reduzida com relação a seu potencial de interpolação e por isso deve-se ser cauteloso a este respeito.

A energia apresentada na última linha da Tabela 3.11 (saída da RNA) corresponde a energia de interação de pares E_{ab} , ou seja a energia de interação intermolecular entre a molécula a e a molécula b .

Tabela 3.10: Dados de Teste. As quatro primeiras colunas são as entradas para a RNA e a última coluna apresenta a saída desejada.

Dados de teste	(R16) ²	(R26) ²	(R27) ²	(R36) ²	Energia (TIP5P)
1	5,29	8,90135	8,43242	8,90123	9,50611
2	5,29	8,90135	15,34234	8,90123	25,98592
3	5,29	5,10244	4,9547	10,60935	17,32791
4	5,29	5,10244	5,52488	6,48297	18,7703
5	5,29	5,10244	9,96242	6,48297	25,63651
6	5,29	10,60935	7,70891	5,10229	15,29212
7	5,29	10,60935	9,96221	5,10229	15,57965
8	5,29	10,60935	17,76117	5,10229	18,81279
9	18,49	24,44493	19,59896	22,59896	-1,663555
10	18,49	24,44493	34,71473	24,4447	0,836288
11	18,49	17,34262	12,27702	27,63815	-0,9737
12	18,49	17,34262	17,96265	27,63815	-1,117693
13	18,49	17,34262	26,0314	19,92362	0,593854
14	18,49	27,63815	22,39418	17,34234	-0,9736
15	18,49	27,63815	26,03106	17,34234	-0,9668
16	18,49	27,63815	38,61877	17,34234	-0,3275
17	39,69	47,98852	40,05246	45,28395	-0,512
18	39,69	47,98852	45,5543	45,28395	-0,4557
19	39,69	37,5828	32,08275	52,66695	-0,3274
20	39,69	37,5828	34,0915	52,66695	-0,2448
21	39,69	37,5828	50,10038	52,66695	0,3896704
22	39,69	52,66695	45,07945	37,58239	-0,3274
23	39,69	52,66695	50,09992	37,58239	-0,3738
24	39,69	52,66695	67,47637	37,58239	-0,06326
25	68,89	79,5321	69,20006	75,96893	-0,224
26	68,89	79,5321	75,22402	75,96893	-0,2235
27	68,89	79,5321	97,45949	75,96893	0,0991409
28	68,89	65,82298	63,70874	85,69575	-0,1554
29	68,89	65,82298	82,16937	85,69575	0,1735109
30	68,89	85,69575	75,76473	65,82244	-0,1463
31	68,89	85,69575	82,16877	65,82244	-0,1554
32	68,89	85,69575	104,334	65,82244	-0,0179
33	106,09	119,0757	108,345	103,9802	-0,05813
34	106,09	119,0757	113,7434	114,6539	-0,09962
35	106,09	119,0757	140,8319	114,6539	0,0534807
36	106,09	102,0632	92,50232	121,3122	-0,055845
37	106,09	102,0632	122,2383	126,7245	0,0915926
38	106,09	126,7245	114,45	102,0625	-0,07719
39	106,09	126,7245	122,2376	102,0625	-0,06759
40	106,09	126,7245	149,1916	102,0625	-0,00618
41	151,29	166,6193	151,857	148,5926	-0,033387
42	151,29	166,6193	162,7079	148,5926	-0,03391
43	151,29	166,6193	192,2043	161,3389	0,0320108
44	151,29	146,3033	132,5142	169,29	-0,03264
45	151,29	146,3033	143,0358	169,29	-0,04323
46	151,29	175,7534	161,1353	146,3025	-0,04611
47	151,29	175,7534	170,3065	146,3025	-0,05325
48	151,29	175,7534	202,0492	146,3025	-0,002337

Tabela 3.11: Restrição de máximos e mínimos para as entradas e saída das redes treinadas usando-se 486 dados.

	Valor Mínimo ($?^2$ e $?$)	Valor Máximo ($?^2$ e $?$)
$(R16)^2$ e R16	5,290 e 2,300	151,3 e 12,30
$(R26)^2$ e R26	5,102 e 2,259	175,8 e 13,26
$(R27)^2$ e R27	2,262 e 1,504	202,1 e 14,22
$(R36)^2$ e R36	5,102 e 2,259	175,8 e 13,26
Energia (kcal mol^{-1})	-1,8440	28,9199

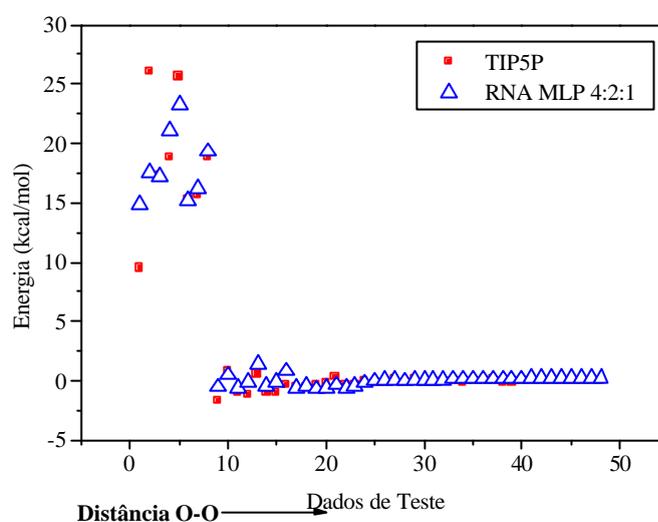


Figura 3.30: Conjunto de teste para predição da energia potencial usando-se o ajuste por RNAs. Entradas para os modelos são os quadrados das distâncias.

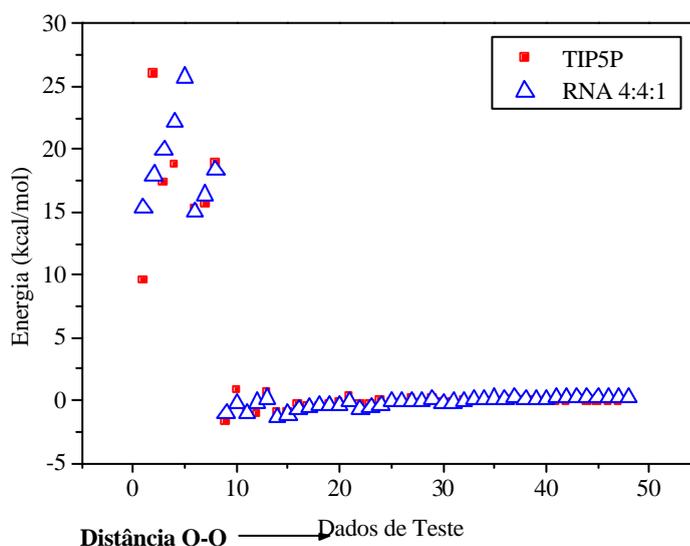


Figura 3.31: Conjunto de teste para predição da energia potencial usando-se o ajuste por RNAs. Entradas para os modelos são o inverso dos quadrados das distâncias.

Testou-se ainda a aplicação de entradas para as RNAs como inversos das distâncias quadradas, dada a dependência inversa da energia com as distâncias entre os átomos constituintes das moléculas. Entretanto apesar de se utilizar uma rede neural mais complexa (4:4:1) para esta análise (Figura 3.73), não se verificou mudanças significativas.

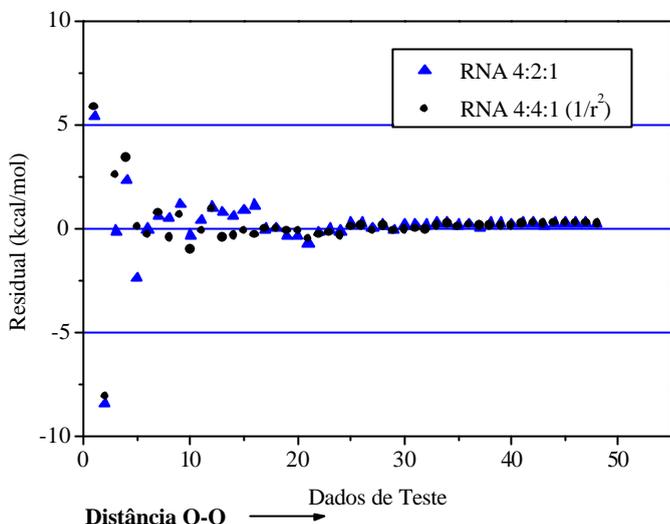


Figura 3.32: Residual (valor predito – valor TIP5P) para as RNAs analisadas.

3.3.3. Ajuste da energia de interação usando RNAs

Os resultados detalhados a seguir referem-se ao conjunto formado por um total de 405 dados gerados no intervalo de distância entre os centros de massa de 2,0 Å até 11,4 Å, distribuídos para os seguintes valores de distância (2,0; 2,4; 5,4; 8,4 e 11,4 Å). Este conjunto engloba uma faixa maior para a aplicação das RNAs com relação ao limite mínimo. O limite máximo apenas precisa englobar o raio de corte ajustado, ou seja, 9 Å. A Tabela 3.12 mostra a ampliação dos limites com relação a Tabela 3.10.

Tabela 3.12.: Restrição de máximos e mínimos para as entradas e saída da redes treinadas usando-se 405 dados.

	Valor Mínimo (?^2)	Valor Máximo (?^2)
$(R16)^2$ e R16	4,000 e 2,000	129,9 e 11,40
$(R26)^2$ e R26	3,956 e 1,989	152,7 e 12,36
$(R27)^2$ e R27	2,236 e 1,495	177,3 e 13,32
$(R36)^2$ e R36	3,956 e 1,989	125,7 e 11,21
Energia de interação (kcal mol^{-1})	-0,9366	133,83

Do conjunto total de 405 dados, 394 representam o conjunto de treinamento, 6 o de validação e 5 o conjunto de teste. Como os dados foram gerados de maneira sistemática, de modo a englobar ou garantir representatividade dos dados, também a distribuição do total em cada conjunto foi sistemática. A RNA escolhida foi então submetida ao desafio de prever a superfície de energia potencial para um novo conjunto com 243 dados, gerado dentro do intervalo da distância entre os centros de massa usado no treinamento, mas em passos distintos: 2,0; 6,0 e 10,0 Å.

Resultados dos Treinamentos

Dentre os métodos de otimização disponíveis para as RNAs do tipo MLP, foi utilizado o método do gradiente conjugado dentro do algoritmo de treinamento de retro-propagação (*backpropagation*). Foram testadas várias funções de transferência, como: linear, sigmoideal e tangente hiperbólica. Para as redes do tipo MLP os melhores resultados foram obtidos usando-se a função logística ou sigmoideal.

O treinamento foi realizado com 4 neurônios na camada de entrada, com dados relativos as distâncias e um neurônio na camada de saída (*single output*), com o valor da energia de interação referente aquela configuração.

Na Tabela 3.13 e Figura 3.33 a seguir estão relacionadas as melhores RNAs ajustadas, com os respectivos erros para os conjuntos de treinamento, validação e teste.

Tabela 3.13: Erro quadrático médio (kcal mol^{-1})² para os conjuntos de treinamento, validação e teste.

Tipos de RNAs	Erros			N° Pesos + <i>Bias</i>
	Treinamento	Validação	Teste	
a. MLP (4:18:13:1)	2,406127	1,452004	1,54061	319 + 32
b. MLP (4:2:1)	1,664958	0,8483889	1,187354	10 + 3
c. MLP (4:18:3:1)	1,706998	0,5714107	1,457506	129 + 22
d. MLP (4:4:1)	1,549138	0,5572845	1,323607	20 + 5
e. RBF (4:34:1)	1,530329	0,9335335	4,2326	170 + 35
f. GRNN (4:394:2:1)	1,299346	1,850632	5,551361	2366 + 397

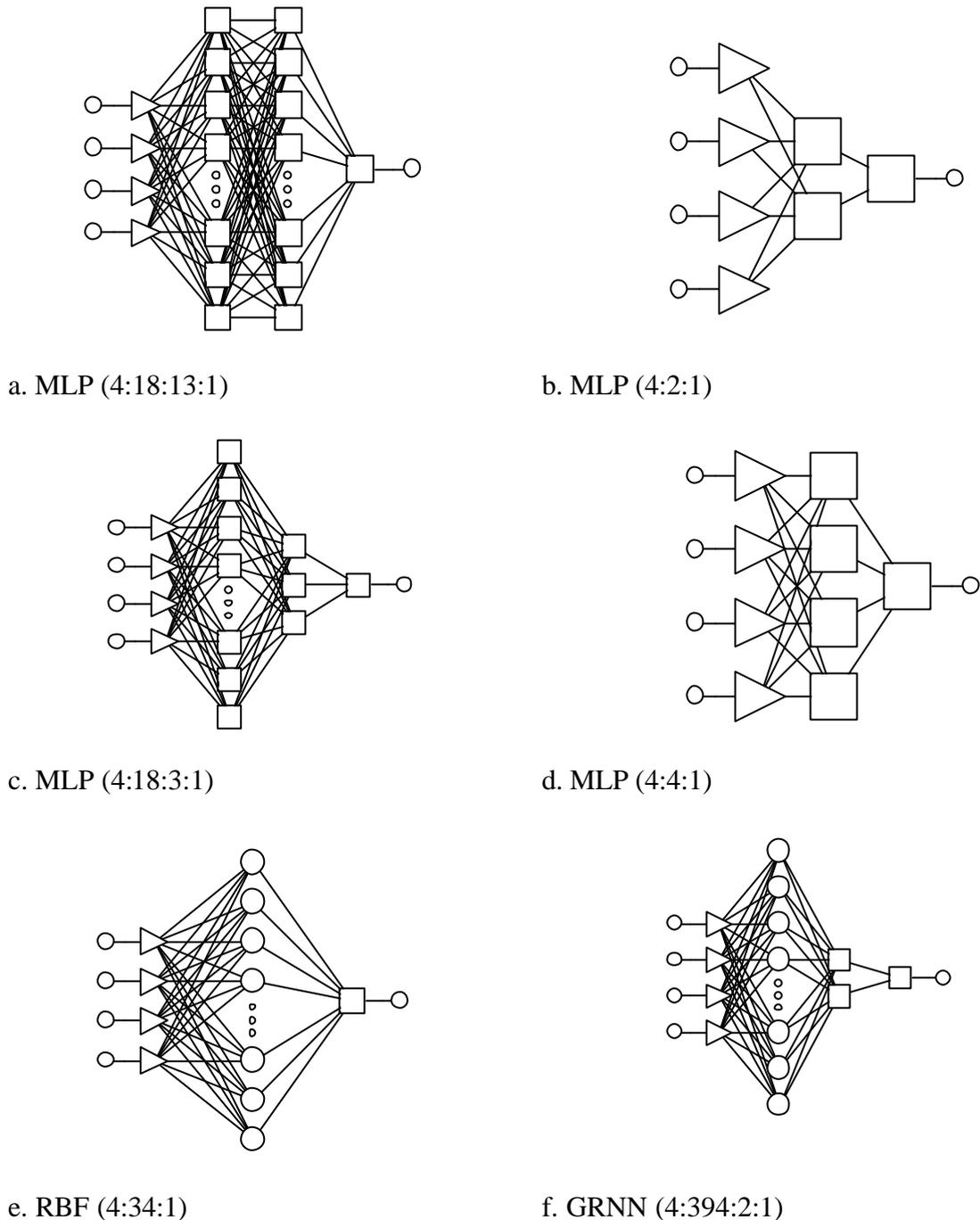


Figura 3.33: Representações esquemáticas de RNAs utilizadas.

A superfície de energia potencial prevista pelo modelo TIP5P para os 405 dados é apresentada na Figura 3.34 e seu comportamento deve ser capturado pelo modelo de rede neural. Examinando os resultados mostrados na Tabela 3.15 (erros) e na Figura 3.33 (complexidade), serão apresentados os resultados para 4 redes escolhidas entre as enumeradas acima. São elas: MLP (4:18:13:1); MLP (4:2:1); MLP (4:18:3:1) e MLP (4:4:1). As superfícies de energia potencial previstas por estas redes neurais são apresentadas na Figura 3.35.

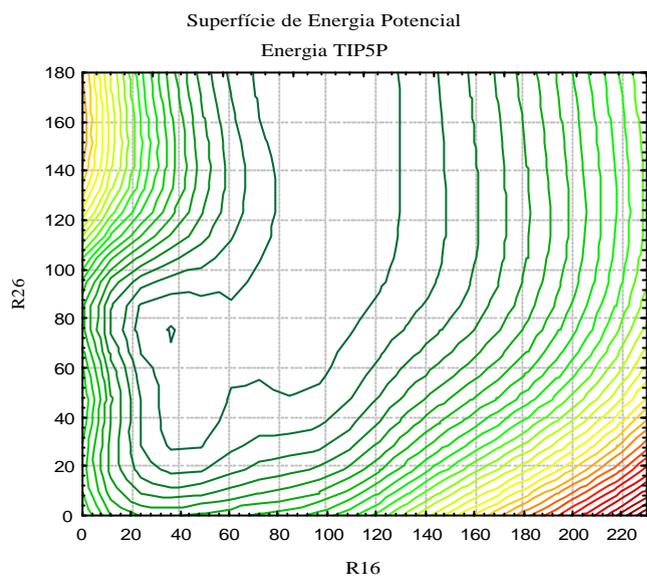


Figura 3.34: Superfície de energia potencial para o dímero da água usando o modelo TIP5P.

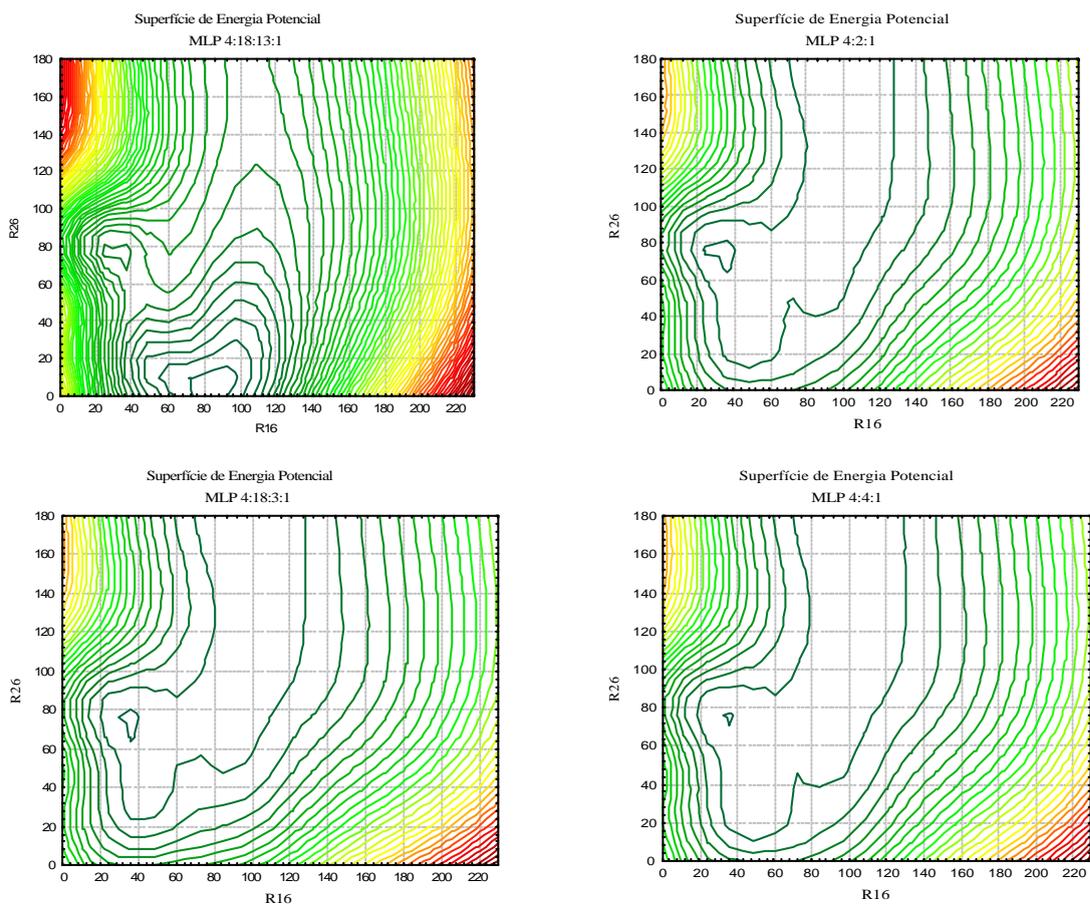


Figura 3.35: Superfícies de energia potencial para o dímero da água, obtidas usando os modelos de redes neurais a serem testados.

Para permitir uma escolha criteriosa da melhor rede neural foram ainda colocados em gráficos os resultados dos residuais em kcal mol^{-1} entre os dados previstos pelas redes e aqueles obtidos usando o modelo TIP5P e os erros relativos absolutos (Figura 3.36a e 3.36b). Além disso podemos analisar as previsões geradas pelas RNAs para os 405 dados na Figura 3.37.

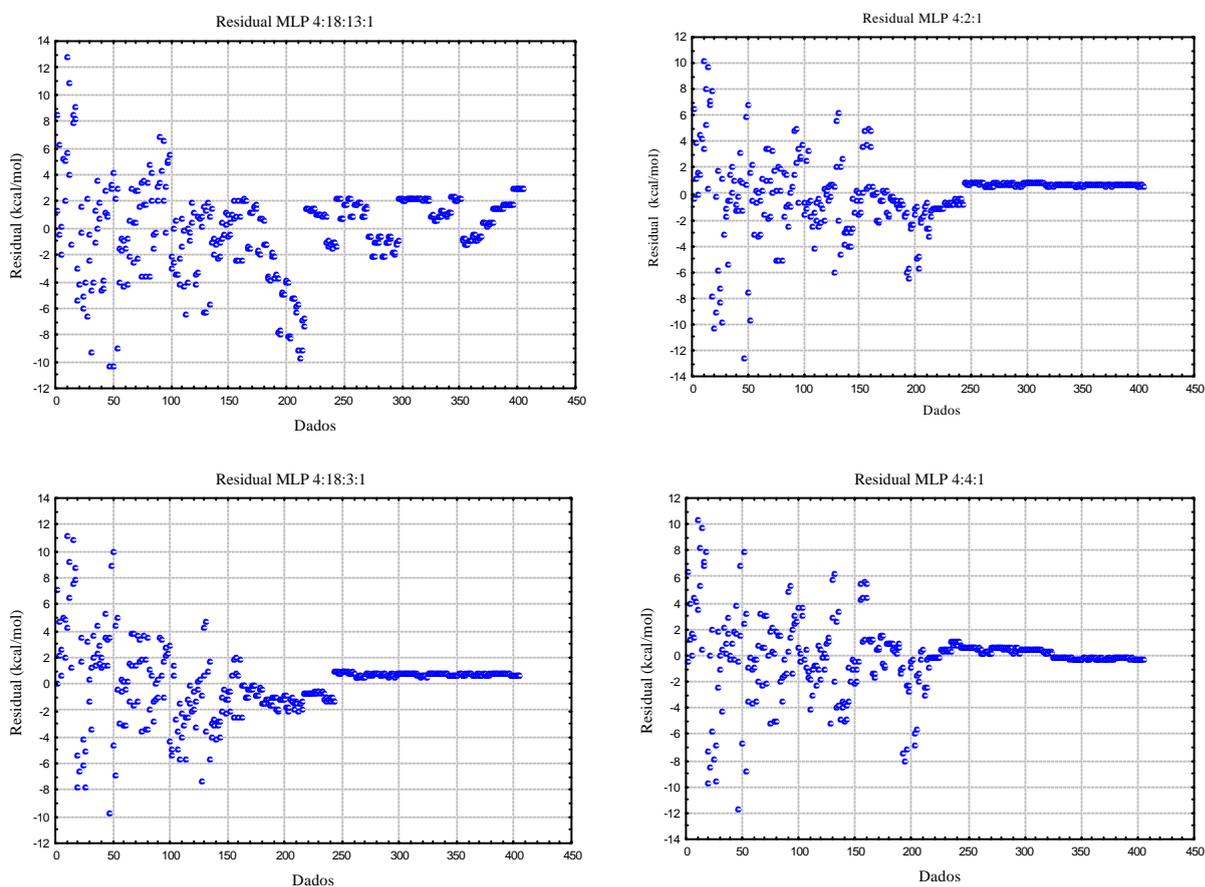


Figura 3.36a: Residuais em kcal mol^{-1} gerados pelos modelos baseados em RNAs.

O erro absoluto médio na região de alta energia obtido por Tai No *et al.* (1997)¹¹³ ($5,904 \text{ kcal mol}^{-1}$) é comparável com o erro máximo visualizado na Figura 3.36a ($< 12 \text{ kcal mol}^{-1}$) para a RNA 4:4:1. A análise do erro é melhor indicativa do desempenho das RNAs quando calcula-se o erro relativo e o logaritmo do módulo erro relativo (Figura 3.36b), que permitiu verificar um alto erro relativo nas regiões de baixa energia, inversamente ao comportamento observado na Figura 3.36a.

¹¹³ K. Tai No, B. H. Chang, S. Y. Kim, M. S. Jhon, H. A. Scheraga, *Chem. Phys. Lett.* 271, 152-156 (1997).

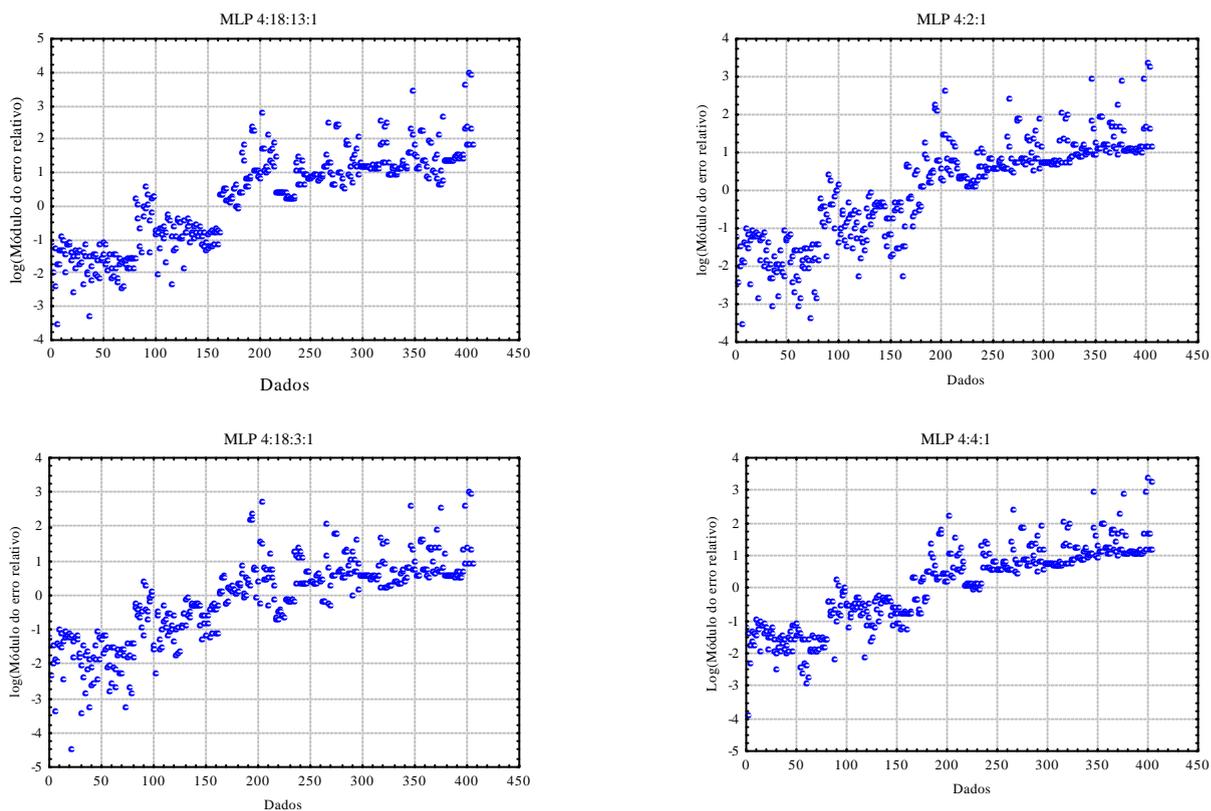


Figura 3.36b: Logaritmo do módulo do erro relativo gerados pelos modelos baseados em RNAs.

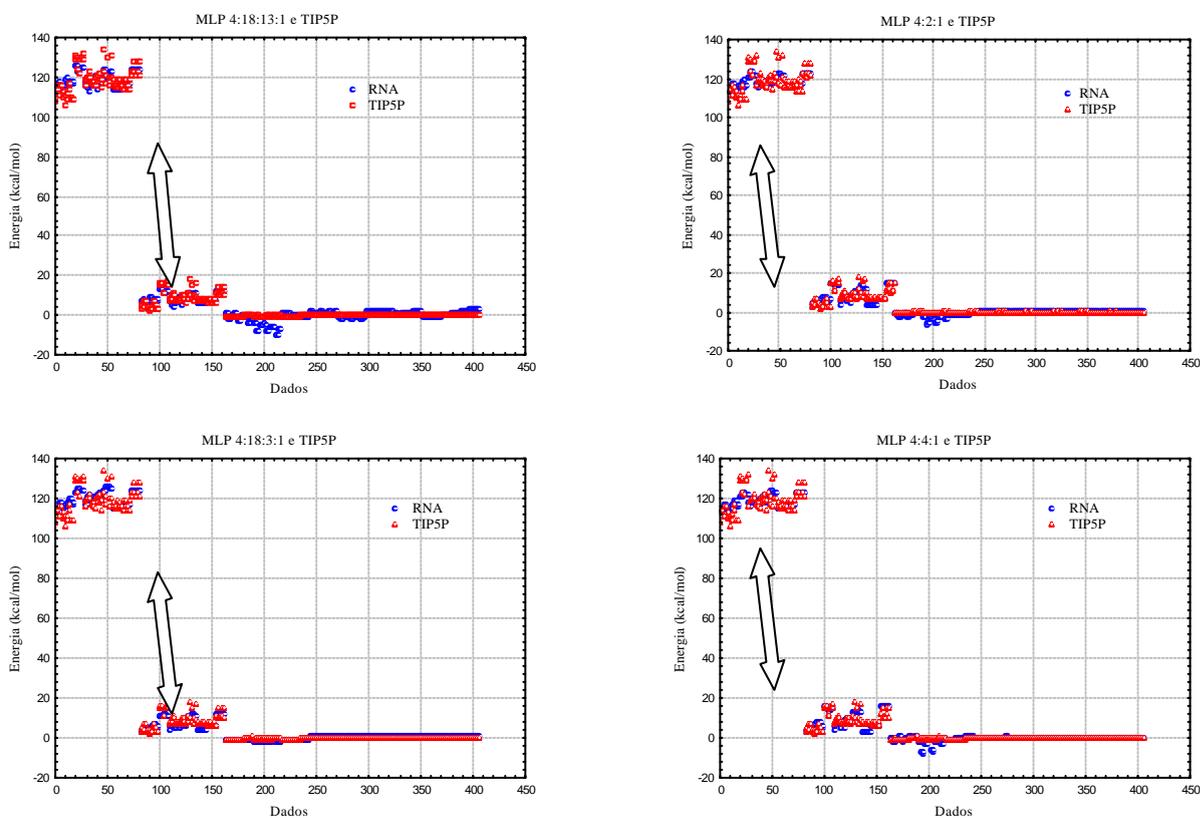


Figura 3.37: Predições da energia para os quatro modelos baseados em RNAs comparados com TIP5P.

Examinando os resultados (complexidade, residual, predição), escolheu-se a RNA 4:4:1 (Figura 3.38). A função de ativação para esta RNA é a sigmoideal, para a qual deve-se restringir a saída da rede entre 0 e 1. Por isso, os dados de treinamento foram escalonados para esta faixa (Tabela 3.14). As conexões (pesos e *bias*) entre os neurônios para esta rede são apresentados nas Tabelas 3.15-3.17. Neste ponto é importante observar que os dados gerados sistematicamente contém uma região de dados de energia não explorada, representada na Figura 3.38 pelas setas duplas.

Tabela 3.14: Escalonamento linear interno para os dados de entrada para a RNA 4:4:1.

	Entrada igual a zero	Entrada diferente de zero
Entrada 1	0,34880544	$R16^2 * 0,00793903 + (-0,03175611)$
Entrada 2	0,33699436	$R26^2 * 0,00672311 + (-0,02659940)$
Entrada 3	0,30875635	$R27^2 * 0,00568829 + (-0,00834465)$
Entrada 4	0,31745484	$R36^2 * 0,00672296 + (-0,02659891)$

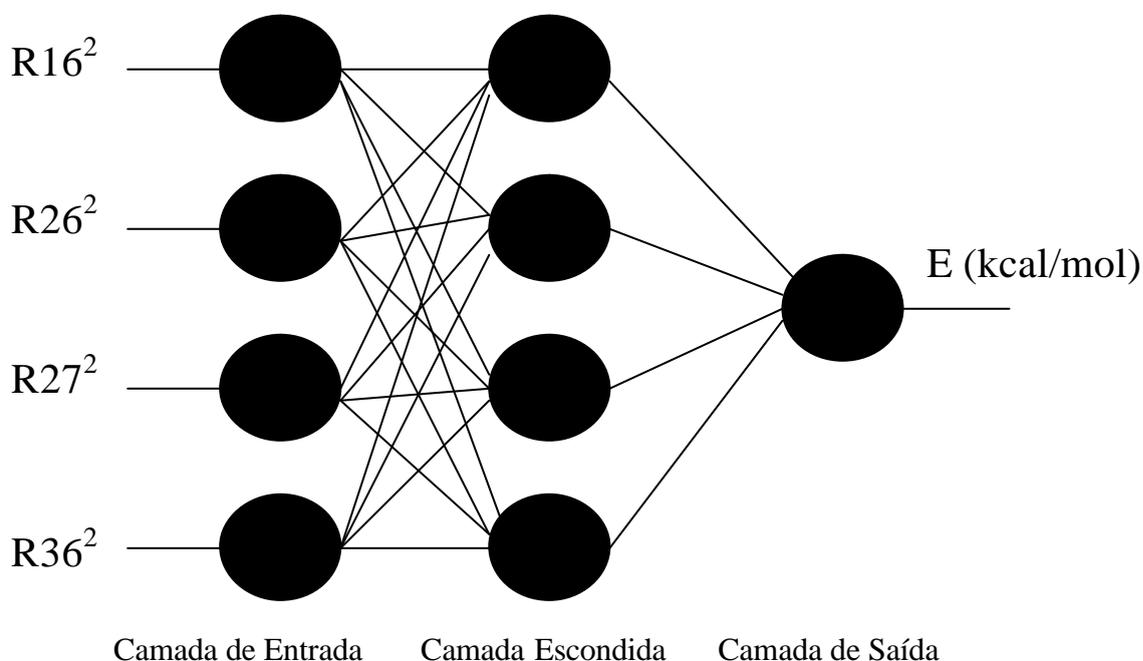


Figura 3.38: Rede neural *feedforward* para predição da superfície de energia potencial.

Tabela 3.15: Pesos (w_{ij}) ajustados para as conexões entre a camada de entrada e a escondida.

i	j	5	6	7	8
1		0,57970238	0,59987568	0,46659641	1,00333488
2		24,08403920	9,90255845	5,43278654	11,30857508
3		4,77611042	10,24936858	6,46708333	12,28130290
4		1,47393221	1,40326288	1,03977157	1,82916667

Tabela 3.16: Pesos (w_{ij}) ajustados para as conexões entre a camada escondida e a camada de saída.

j	i	5	6	7	8
9		-0,19398073	-11,72474124	10,55814720	-1,48131279

Tabela 3.17: Limiares ou *bias* ($?_i$) ajustados para os neurônios das camadas escondida e de saída.

i	5	6	7	8	9
$?_i$	-5,02339148	0,78770309	0,78647756	-2,86340119	-2,84611651

A rede neural escolhida foi então submetida a novos dados gerados para um passo de R16 (quadrado da distâncias entre os centros de massa da molécula) distinto e em total de 234 dados. As superfícies de energia potencial são apresentadas na Figura 3.39 para todos os dados de teste e permite concluir que a rede neural obtida foi também capaz de capturar este novo comportamento.

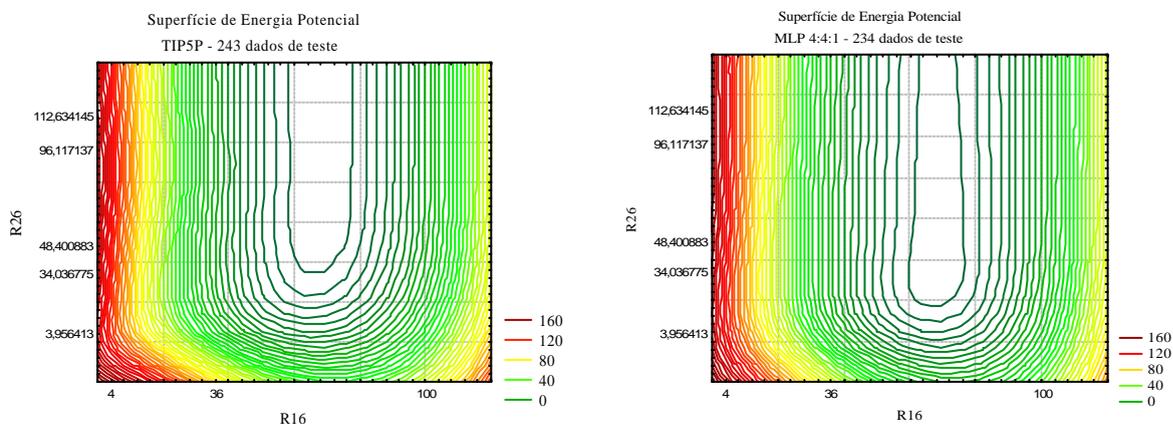


Figura 3.39: Superfícies de energia potencial para o dímero da água, obtidas com o TIP5P e a RNA 4:4:1.

As previsões para três dos dados de teste, variando-se apenas as distâncias entre os centros de massas são apresentadas na Tabela 3.18, juntamente com os residuais (valor predito – valor TIP5P). Os resultados indicam um bom desempenho para a RNA 4:4:1.

Tabela 3.18: Energias e resíduos (kcal mol^{-1}) para três dados de teste variando-se apenas a distância entre os centros de massa.

$(\text{O-O})^2 \text{ \AA}^2$	TIP5P	MLP 4:4:1	Residual
4	108,180881	114,630488	-6,44960698
36	0,68430017	1,77023192	-1,08593175
100	-0,14935279	-0,146921929	-0,00243086

Os tempos para execução da subrotina responsável pelo cálculo do potencial de interação através do modelo TIP5P e através do modelo de RNA com estrutura 4:4:1 estão apresentados na Tabela 3.19. O tempo foi determinado como a média a partir de um milhão de chamadas das subrotinas, tornando o resultado estatisticamente confiável. Os tempos calculados são comparados com resultados da literatura¹¹³, entretanto é importante lembrar que a comparação do tempo é realista para máquinas idênticas e a evolução pela qual os computadores vem passando dificulta ainda mais esta comparação, de modo que o tempo calculado para BPNN deve ser pelo menos 10 vezes menor, dada a evolução das máquinas de 1997 a 2002 (data em que estes resultados foram obtidos), ou seja, 5 anos.

Tabela 3.19.: Demanda computacional para os modelos TIP5P e aquele baseado em RNAs.

	Dímero da água	Estimativas de uma simulação de Monte Carlo ^{a)}	
		Tempo total (h)	Razão
RNA (4:4:1)	11 μs	17,6	1
TIP5P	41 μs	42,0	3,7
BPNN (6:12:12:1) ¹¹³	1,9 ms	3048/10	17,3

^{a)} 10 milhões de configurações com mudança de volume a cada 2000 configurações.

Como podemos observar, uma RNA bastante simples foi capaz de capturar o comportamento do potencial intermolecular da água. Sua execução permite um ganho de

tempo significativo, mesmo para moléculas com somente cinco sítios de interação. O modelo apresentado na literatura, para o qual as entradas utilizadas são seis coordenadas internas, dificulta a inclusão daquele potencial para a realização de simulação computacional, dada alta demanda computacional.

Além disto, o modelo da literatura utiliza 3331 energias *ab initio* MP2 para treinamento da rede neural, valor muito superior ao utilizado no treinamento da RNA 4:4:1 (405).

3.3.4. Aplicação de Modelos de potencial da Água baseados em RNAs em Simulações de Monte Carlo

Simulações de Monte Carlo são continuamente investigadas do ponto de vista computacional, para que o tempo de simulação não inviabilize o estudo de sistemas grandes (elevado número de átomos) e a exatidão seja aprimorada com representações que prevejam com confiabilidade as propriedades dos sistemas simulados. A dependência limitante da demanda computacional é dada pelo cálculo do potencial de interação. Modelos baseados em RNAs poderiam constituir uma opção para este fim, através do ajuste de funções a partir de resultados *ab initio* ou outros ramos da inteligência artificial, como algoritmos genéticos, ou combinação de ambos¹¹⁴. Daí o estudo a que se voltou esta etapa do trabalho de pesquisa, na qual testou-se potenciais de interação baseados em RNAs em simulações de Monte Carlo. Existem trabalhos publicados referentes à obtenção de superfícies de energia potencial coerentes para a água líquida usando RNAs¹¹⁵, entretanto ainda permanece desconhecida a atuação destes modelos em simulações.

Para obtenção do modelo baseado em RNAs capaz de atuar em simulações de Monte Carlo, além da rede neural treinada com dados sistemáticos, gerou-se novos conjuntos de dados, os quais provém de caixas de simulação geradas randomicamente ou equilibradas previamente em simulações de Monte Carlo. Tais dados contém energias de valor maior, ampliando por completo a faixa de atuação da RNA. O objetivo é investigar a dependência das redes neurais com relação ao conjunto de treinamento.

O estudo realizado no item 3.5.3 indicou que o tipo de RNA a ser escolhida para esta tarefa é aquela do tipo MLP. Novos testes ratificaram esta escolha, de modo que os demais tipos de RNAs foram descartadas.

¹¹⁴ S. Hobday, R. Smith, J. BelBruno, *Nuclear Instr. and Meth. in Phys. Research B*. 153: 247-263 (1999).

A RNA 4:4:1 mostrada na Figura 3.38 treinada a partir de dados gerados de forma sistemática (sistemático 1 na Tabela 3.30), cuja saída da rede fornece a energia de interação de pares em kcal mol^{-1} apresentou desempenho desfavorável em simulações de Monte Carlo. A densidade e energia média por molécula divergiram do valor simulado com o modelo TIP5P, bem como as energias monitoradas indicaram que a RNA não foi capaz de extrapolar a faixa para a qual foi treinada. Os dados gerados sistematicamente não explorou uma região considerável de energia (setas duplas na Figura 3.37), de modo que um dos teste seguintes foi incluir um passo menor nesta região formando um novo conjunto de dados, chamado sistemático 2. A Figura 3.40 apresenta este conjunto, para o qual em valores menores da distância quadrada usou-se um passo menor de distância para cálculo da energia.

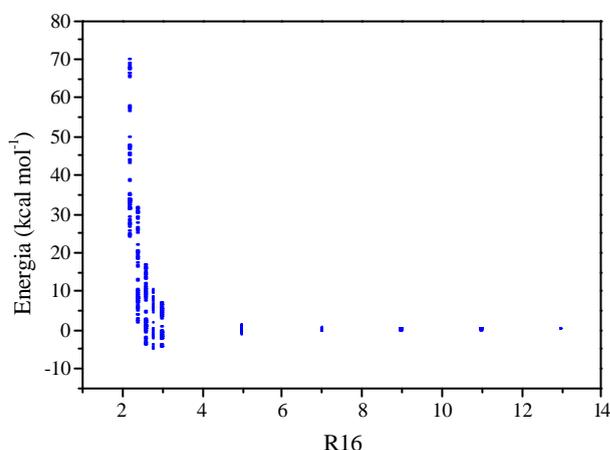


Figura 3.40: Energia do conjunto de dados de treinamento (Sistemático 2) com relação a distância R16 ($O_1 \cdots O_2$: oxigênios da moléculas 1 e 2).

Outros testes foram realizados: geração de novos conjuntos de dados, mudança da saída rede. Para este conjunto de dados contendo uma faixa de energia maior, usou-se a saída da rede neural sugerida no trabalho de No *et al.* (1997) [$f(V) = \log(V - V^{ab,0} + C)$], onde V é a energia de interação de pares, $V^{ab,0}$ representa a energia da conformação mais estável e C é uma constante que tem o papel de localizar a posição de máximo da função $f(V)$ e aumentar a eficiência do treinamento das RNAs. O valor 0,25 foi adotado pelos autores¹¹⁶.

Neste trabalho para evitar o cálculo da energia da configuração mais estável durante as simulações, incorporou-se $V^{ab,0}$ na constante [$f(E) = \log(E + C)$, com $E = V$], de modo

¹¹⁵ K. Tanabe, T. Tamura, H. Uesaka, *Appl. Spectrosc.* 46, 807-810 (1992); B. Curry, D. E. Rumelhart, *Tetrahedron Comput. Methodol.* 3, 213-237 (1990).

que testou-se diversos valores de C para os quais $C > V^{ab,0}$. A alteração foi importante para permitir a inclusão de altas energias no conjunto de treinamento com menor efeito no erro calculado. Entretanto, o desempenho das RNAs nas simulações de Monte Carlo não foi satisfatório, além de requerer um grande número de configurações para equilíbrio.

A Tabela 3.30 mostra um resumo dos resultados obtidos para algumas das RNAs testadas, exibindo o tipo de RNA, a função de saída ($f(E)$) com o valor da constante C , conjunto de treinamento utilizado e o percentual de aceitação para as variações de volume.

O conjunto de treinamento nomeado “Caixa equilibrada” corresponde a um conjunto de treinamento formado por valores de energias de interação calculadas a partir das configurações de pares aleatórios presentes em uma caixa previamente equilibrada em simulações de Monte Carlo usando o modelo TIP5P.

Os dados apresentados na Tabela 3.20 foram adquiridos após pelo menos 13 milhões de configurações, apesar do acompanhamento ter sido realizado a cada 1 milhão de configurações. Quando necessário, apresentou-se valores entre parênteses que representam a quantidade de configurações utilizadas para a leitura da medida apresentada.

Tabela 3.20: Resultados de simulações de Monte Carlo usando RNAs distintas e usando o modelo TIP5P.

RNA	$f(E)$	C	Conjunto de Treinamento	AVV(%) (milhões de configurações)	$(E/N)_{\text{RNA}}$	ρ (g cm ⁻³)
4:4:1	E	-	Sistemático 1	0,0	-950,7	1,01
4:4:1	E	-	Sistemático 2	0,0	-17,66	1,01
4:2:1	$\log(E + C)$	8,0	Caixa equilibrada	0,0	-90,13	1,01
4:4:1	$\log(E + C)$	8,0	Caixa equilibrada	~11,0 (< 4) 0,0 (> 4)	-257,2	1,15
4:3:1	$\log(E + C)$	10,0	Caixa equilibrada	~95,0	3,25	0,88
4:2:1	$\log(E + C)$	6,0	Caixa equilibrada	~30,0 0,0 (> 10)	-182,4	0,95

A energia média por molécula (E/N) para a caixa equilibrada após 190 milhões de configurações é igual a $-9,97$ kcal mol⁻¹, usando o modelo TIP5P. As simulações de Monte Carlo usando o modelo TIP5P permitiram cálculos de correção de longa distância, ao passo

¹¹⁶ K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, H. A. Scheraga, *Chem. Phys. Lett.* 271: 152-156 (1997).

que com RNAs isto não foi feito. Verifica-se que para as seis RNAs apresentadas acima, a energia média por molécula diverge deste valor. Pode ser necessário um número maior de configurações para que a RNA 4:3:1 esteja devidamente equilibrada.

As Tabelas 3.21-22 apresentam resultados mais detalhados obtidos a partir do monitoramento de simulações de Monte Carlo executadas a partir de uma dada configuração (Figura 3.41) previamente equilibrada, usando a constante C igual a 8 (RNA 4:2:1). O protocolo inicial da equilibrção é o mesmo apresentado no item 3.2.

Para variações de volume das primeiras configurações (< 200.000) ocorre aceitação para variação de volume, mesmo que muito baixa. As novas energias calculadas para movimentos rotacionais e translacionais vão diminuindo rapidamente usando o modelo baseado em RNAs e aumentando a partir do cálculo da energia usando o modelo TIP5P, de modo que o comportamento observado na Tabela 3.21 inverte-se e nenhuma variação de volume é aceita para a RNA (4:2:1) em teste. Por isto a energia média por molécula tende a um valor menor que aquele previsto pelo modelo TIP5P (Tabela 3.22).

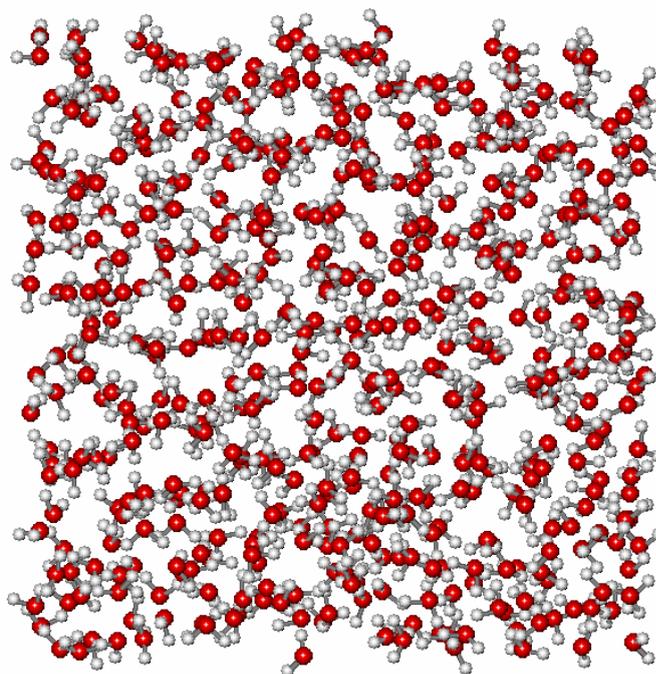


Figura 3.41: Configuração de pares de moléculas de água usadas para o treinamento de RNAs (caixa equilibrada em condição ambiente de temperatura e pressão).

A cada 200 mil configurações coletou-se as densidades do líquido (ρ) e a energia média por molécula (E/N). Verificou-se ainda a aceitação quanto à variação de volume (AVV) do sistema (Tabela 3.22).

Tabela 3.21: Energias totais de interação (kcal mol^{-1}) para uma caixa com 256 (L1) + 256 (L2) moléculas de água, usando o modelo baseado em RNAs e o modelo TIP5P.

	RNA 4:2:1	TIP5P
Interações Líquido 1 - Líquido 1 (L1L1)	-988,633	-1259,74
	-1010,14	-1274,31
	-1029,32	-1278,07
Interações Líquido 2 - Líquido 2 (L2L2)	-991,525	-1316,08
	-999,351	-1338,26
Interações Líquido 1 - Líquido 2 (L1L2)	-2007,29	-2489,72
	-2057,44	-2492,09

Tabela 3.22: Energias médias por molécula (L1L1 + L2L2 + L1L2) para uma caixa com 512 moléculas de água utilizada em simulações de Monte Carlo.

FVV = 180					
ρ_{RNA} (g cm^{-3})	ρ_{TIP5P} (g cm^{-3})	E / N_{RNA} (kcal mol^{-1})	E / N_{TIP5P} (kcal mol^{-1})	AVV_{RNA} (%)	AVV_{TIP5P} (%)
1,0091	1,0023	-21,982	-9,9741	0	45
1,0091	0,9947	-46,819	-10,001	0	42
1,0091	0,9998	-56,748	-9,9913	0	42
1,0091	0,9943	-62,667	-9,9205	0	22
1,0091	0,9889	-65,634	-9,9216	0	30
FVV = 80					
1,0091	0,9996	-68,513	-9,9042	0	62
1,0091	0,9983	-70,331	-9,9245	0	60
1,0091	0,9979	-70,746	-9,9660	0	54
1,0091	0,9956	-72,688	-9,9714	0	68
1,0091	1,0051	-73,545	-9,9938	0	50
1,0091		-73,881		0	
1,0091		-74,205		0	
1,0091		-74,523		0	
1,0091		-74,606		0	
1,0091		-74,707		0	

Alterando-se a frequência de variação do volume de 180 para 80 era de se esperar um aumento na aceitação de variações de volume, assim como ocorreu usando o modelo TIP5P, entretanto não se verificou este resultado usando RNAs, permanecendo $AVV = 0$.

Após 16 milhões de configurações usando o modelo baseado em RNAs a energia média por molécula continua diminuindo (divergindo do valor $-9,9 \text{ kcal mol}^{-1}$, obtido usando o modelo TIP5P), cujo valor atinge $-90,133 \text{ kcal mol}^{-1}$ com 26 milhões de configurações (Figura 3.42). Neste ponto da simulação nenhuma variação de volume foi aceita.

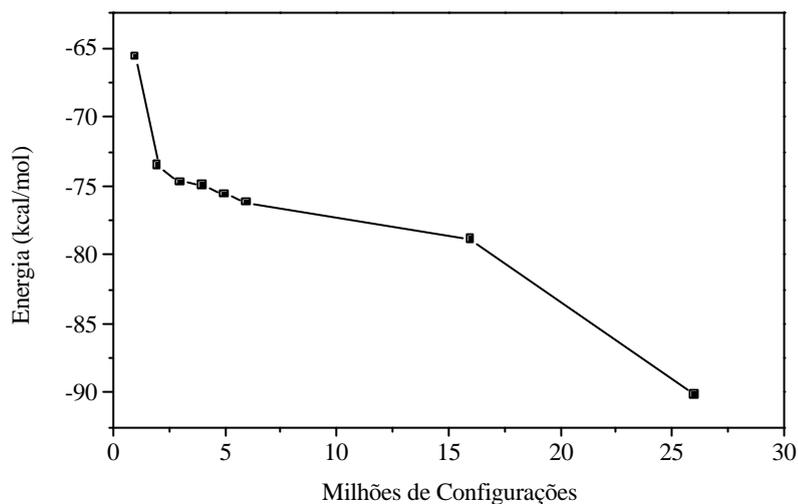


Figura 3.42: Comportamento da energia média por molécula (E/N) durante a simulação de Monte Carlo em funções de número de configurações geradas.

O tempo necessário para execução dos dados permaneceu consideravelmente menor usando RNAs, com tempo 3 vezes menor que usando o modelo TIP5P, apesar do logaritmo acrescentado, já que a estrutura da rede permaneceu de baixa complexidade.

Várias RNAs foram testadas, entretanto até a data atual não se obteve uma rede neural com precisão suficiente para aplicação em simulações de Monte Carlo.

Como pode-se observar, a demanda computacional reduziu em até 3 vezes para a simulação usando RNA, o que estimula o treinamento de novas RNAs que tenham desempenho satisfatório nestas simulações, levando em consideração que a demanda computacional aumenta com o aumento do número de sítios e uma RNA independe deste número. De modo que sistemas mais complexos teriam sua simulação computacional possível.

3.4. Simulação Computacional do Metanol

Simulações de Monte Carlo foram executadas usando o modelo de três sítios desenvolvido por Jorgensen *et al.* (1983)¹¹⁷, que ajustaram os parâmetros às propriedades do metanol líquido (OPLS). As simulações computacionais foram realizadas com o programa DIADORIM¹¹⁸, mantendo-se o número de moléculas (N), a temperatura (T) e a pressão (p) constantes e controladas. Foram analisadas cinco condições extremas além das condições normais de temperatura 25°C e pressão 1 atm.

3.4.1. Propriedades Termodinâmicas

A Figura 3.43 apresenta uma caixa computacional equilibrada nas condições ambiente. As etapas de equilibração e de aquisição de dados foram realizadas com no mínimo 10×10^6 configurações em cada etapa.

Realizou-se a simulação computacional gerando uma caixa cúbica com 512 moléculas de metanol, com raio de corte igual a 15 Å, e mudança de volume a cada 2000 configurações. Os ajustes adotados para movimentos moleculares e mudanças de volume para buscar manter uma razão de aceitação próxima de 40%, são apresentados na Tabela 3.23.

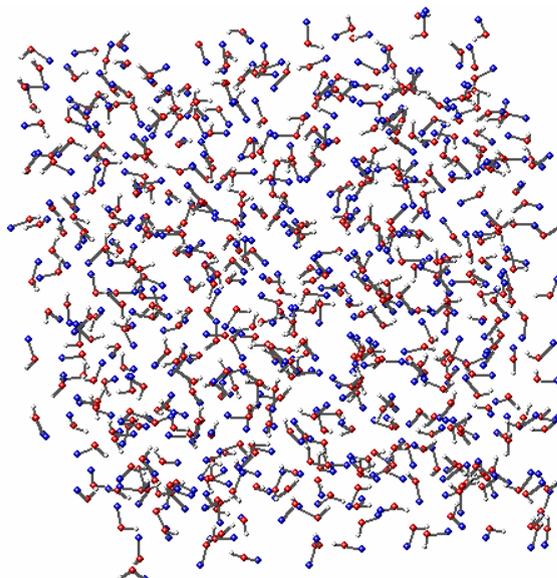


Figura 3.43: Caixa de simulação equilibrada em condições ambiente, contendo 512 moléculas de metanol (três sítios: CH₃, O, H). O volume da caixa é (32,82 Å)³.

¹¹⁷ W. L. Jorgensen, J. Chandrasekhar, J.D. Madura, *J. Chem. Phys.* 79, 926 (1983).

As propriedades termodinâmicas do metanol estão apresentadas na Tabela 3.24.

Tabela 3.23: Protocolos de simulação incluindo razão de aceitação A (%), ângulo máximo de rotação θ_{rot} ($^{\circ}$), deslocamento máximo de translação $R_{\text{máx}}$ (?), razão de aceitação das variações de volume A_v (%), máxima variação de volume $V_{\text{máx}}$ e dimensão final da caixa cúbica.

p (bar)	T ($^{\circ}\text{C}$)	A (%)	θ_{rot} ($^{\circ}$)	$R_{\text{máx}}$ (?)	A_v (%)	$V_{\text{máx}}$	a (?)
1	25	~40	16,5	0,165	~50	500	32,89
500	100	~50	18,0	0,18	~55	600	33,19
500	200	~50	37,0	0,37	~46	600	35,31
500	300	~72	37,0	0,37	~45	700	37,88
500	400	~80	37,0	0,37	~55	700	41,84
500	500	~85	37,0	0,37	~60	700	46,42

Tabela 3.24: Propriedades termodinâmicas obtidas nas simulações de Monte Carlo do metanol. O valor entre parênteses representa o erro em relação ao valor experimental ($0,79 \text{ g cm}^{-3}$)*.

p (bar)	T ($^{\circ}\text{C}$)	ρ (g cm^{-3})	$-E/N$ (kcal mol^{-1})	C_p ($\text{cal mol}^{-1} \text{ K}^{-1}$)	$10^6 \kappa$ (atm^{-1})	$10^5 \alpha$ (K^{-1})
1	25	0,77 (2,5%)*	8,59	14,85	104,9	130,2
500	100	0,74	7,64	14,31	113,5	95,7
500	200	0,63	6,00	13,90	246,8	142,8
500	300	0,50	4,33	11,76	399,0	146,1
500	400	0,37	2,95	9,20	663,7	127,0
500	500	0,27	1,96	3,79	619,8	58,7

Como era esperado, a densidade do metanol líquido diminui com o aumento da temperatura. Verifica-se também a diminuição da energia média por molécula e da capacidade calorífica. A partir dos dados de energia média por molécula a temperatura ambiente verifica-se um valor menor ($8,59 \text{ kcal mol}^{-1}$) que aquele obtido para a água apresentado na Tabela 3.1 ($9,87 \text{ kcal mol}^{-1}$), mostrando a mais forte interação entre

¹¹⁸ L. C. Gomide Freitas. Laboratório de química computacional da universidade federal de São Carlos. UFSCar (1999).

moléculas de água do que entre as moléculas de metanol nas mesmas condições, assim como os resultados obtidos por Wu *et al.* (1992)¹¹⁹.

3.4.2. Funções de Distribuição Radial

As funções de distribuição radial oxigênio-hidrogênio (O-H), oxigênio-oxigênio (O-O), e hidrogênio-hidrogênio (H-H) para as condições ambiente e extremas são apresentadas nas Figuras 3.44-3.46. Os dados à temperatura ambiente estão em acordo com resultados experimentais.

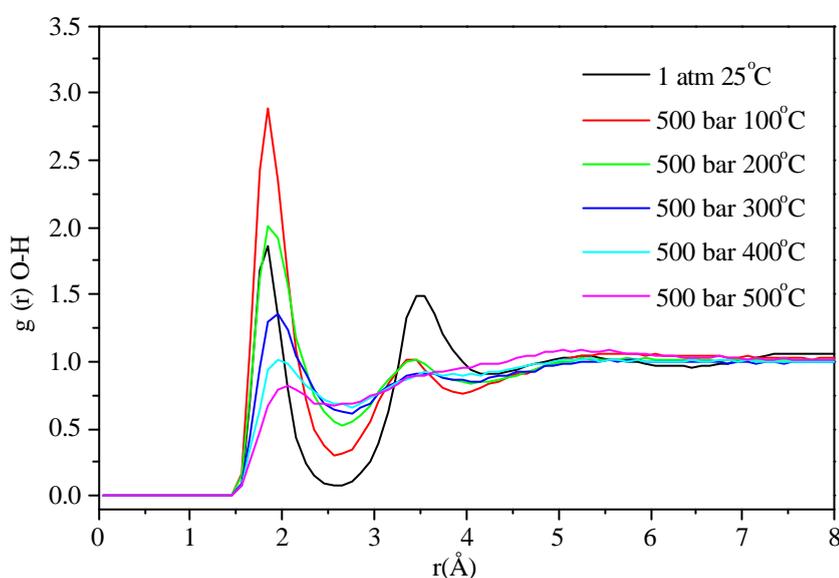


Figura 3.44: RDFs O-H para o metanol em condições ambiente e extremas.

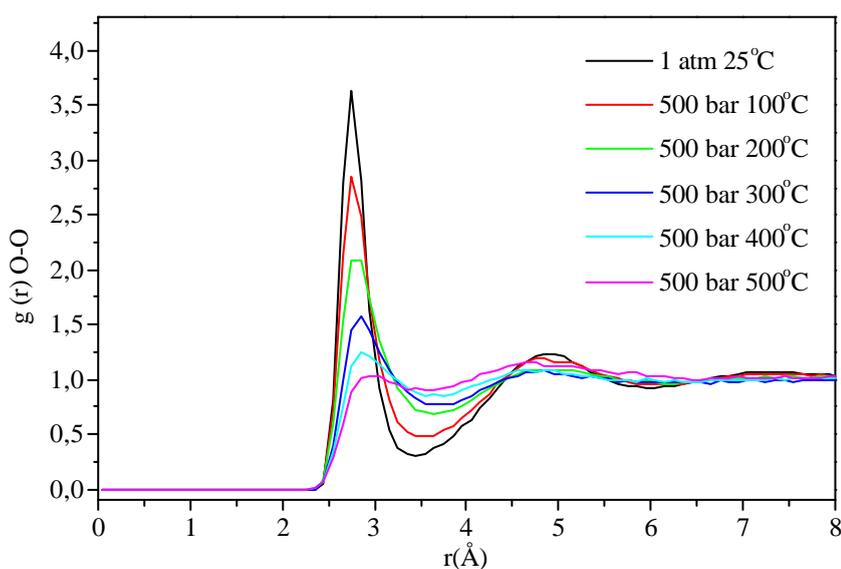


Figura 3.45: RDFs O-O para o metanol em condições ambiente e extremas.

¹¹⁹ X. Wu, Y. Li, J. Lu, T. Teng, *Fluid.Phase Equilibria* 77, 139-156 (1992).

Analisando-se a Figura 3.44 observa-se que o primeiro pico tem alturas diferentes para as condições analisadas sendo reduzido em condições extremas, com temperaturas acima de 200°C sob a pressão de 500 bar. Isto indica que em condições extremas ocorre uma diminuição do número de ligações de hidrogênio. Verifica-se também uma redução do segundo pico, correspondente a segunda camada de solvatação, onde as distâncias O-H não permitem formação de ligações de hidrogênio. Portanto, em altas pressões e temperaturas observa-se uma menor densidade (Tabela 3.24) e um menor ordenamento das moléculas de metanol no líquido puro (Figuras 3.44 -3.46).

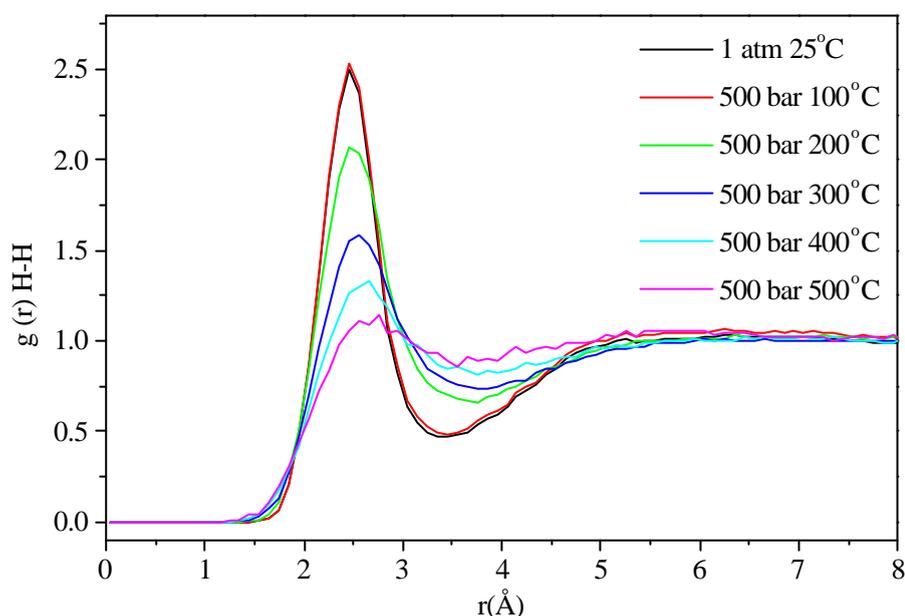


Figura 3.46: RDFs H-H para o metanol em condições ambiente e extremas.

Comparando a RDF O-O do metanol líquido a temperatura ambiente com os resultados experimentais de Yamaguchi *et al.*¹²⁰ na Figura 3.47 abaixo, verifica-se a concordância dos dados obtidos pelas simulações de Monte Carlo.

¹²⁰ T. Yamaguchi, K. Hidaka, A. K. Soper, *Mol. Phys.* 97, 603 (1999).

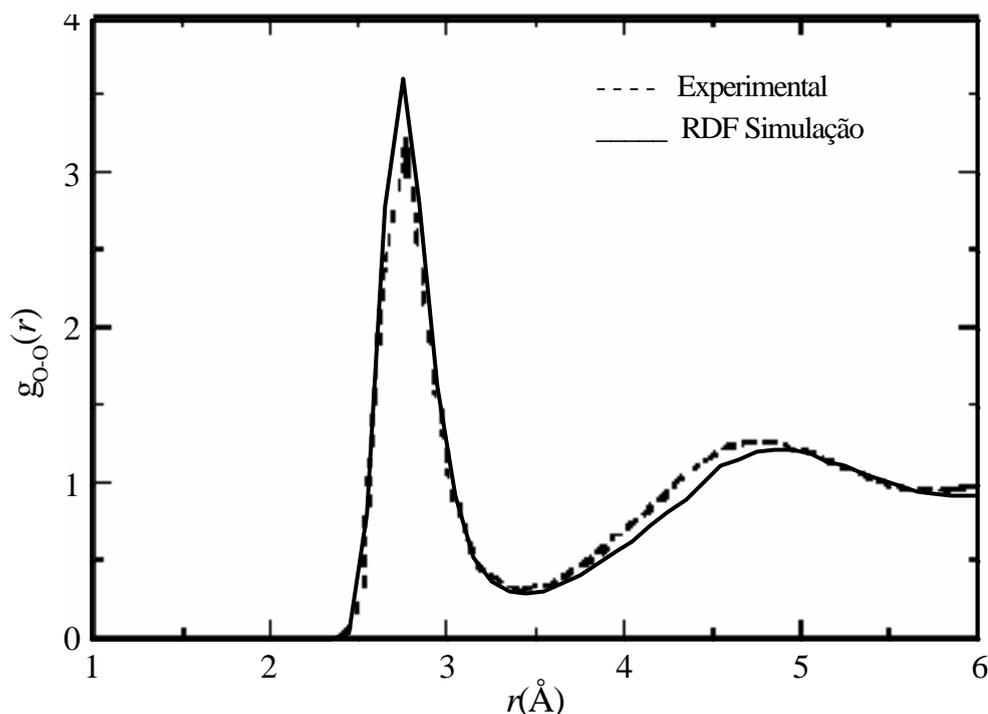


Figura 3.47: RDFs O-O experimental e simulada para o metanol na condição ambiente.

3.4.3. Análise Topológica das Redes de Ligações de Hidrogênio no Metanol

O metanol líquido forma redes complexas de ligações de hidrogênio. A análise da estrutura e estabilidade das redes de ligações de hidrogênio que caracterizam o metanol puro em condições ambiente e supercríticas foi realizada utilizando os critérios geométricos propostos por Pagliai *et al.* (2003)¹²¹ referentes aos comprimentos máximos das ligações (O...H) e (O...O), descritos detalhadamente no item 2.3.7. No mesmo trabalho, os autores demonstraram que para o sistema analisado a restrição angular não era necessária. Para o nosso caso, testamos esta proposta e verificamos também que não há necessidade de impor restrição angular, pois esta é automaticamente satisfeita pelas restrições nas distâncias, como pode ser observado na Figura 3.48.

¹²¹ M. Pagliai, G. Cardini, R. Righini, V. Schettino, *J. Chem. Phys.* 119, 6655-6662 (2003).

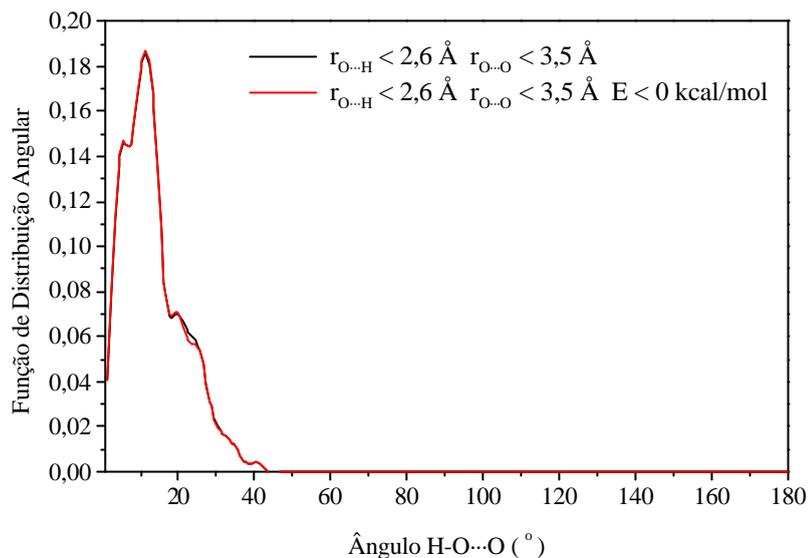


Figura 3.48: Distribuição angular das ligações de hidrogênio em metanol puro nas condições ambiente.

Considerando-se agora a restrição energética ($E < 0 \text{ kcal mol}^{-1}$, linha vermelha da Figura 3.48) nesta caixa de simulação, somente uma configuração $\text{CH}_3\text{OH}\cdots\text{HOCH}_3$ foi rejeitada, na qual o ângulo $\text{H-O} \cdots \text{O}$ é igual a $23,48^\circ$ e a distância $\text{O}\cdots\text{H}$ é $2,2 \text{ \AA}$, ilustrada na Figura 3.49.

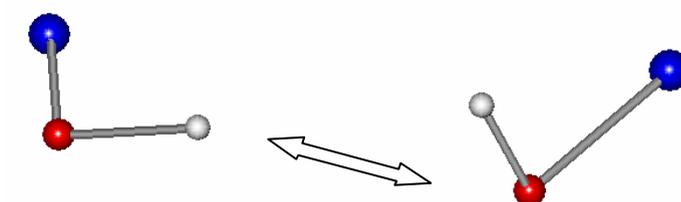


Figura 3.49: Ligação de hidrogênio rejeitada pelo critério energético.

A Figura 3.50 mostra as ligações de hidrogênio formadas à temperatura ambiente e pressão de 1 atm e a Figura 3.51 mostra as ligações de hidrogênio a 500 bar e 400°C para uma dada caixa equilibrada. Nas Figuras os círculos representam as moléculas de metanol e as conexões representam as ligações de hidrogênio.

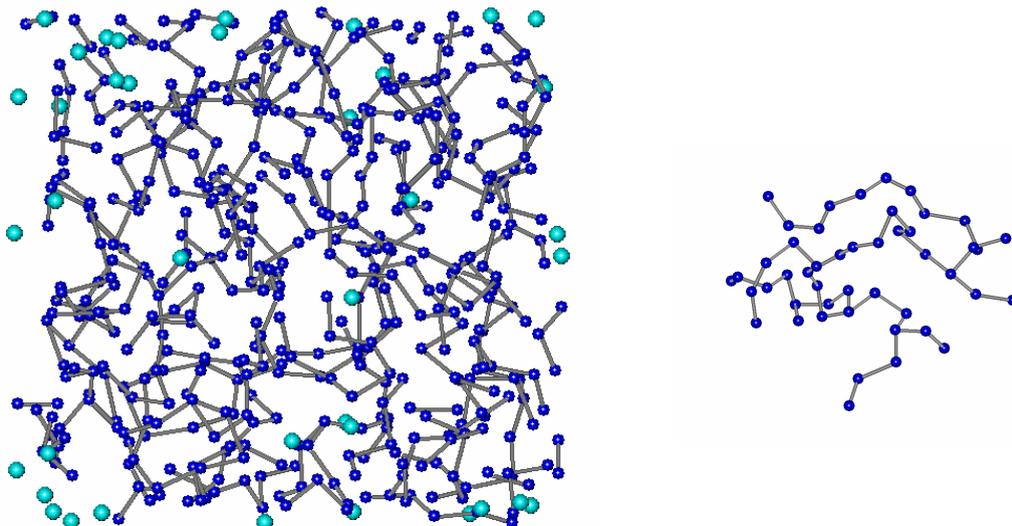


Figura 3.50: Rede de ligações de hidrogênio no metanol puro nas condições ambiente. Maior agregado da rede à direita. Cada círculo representa uma molécula de metanol. O volume da caixa é $(32,82 \text{ \AA})^3$.

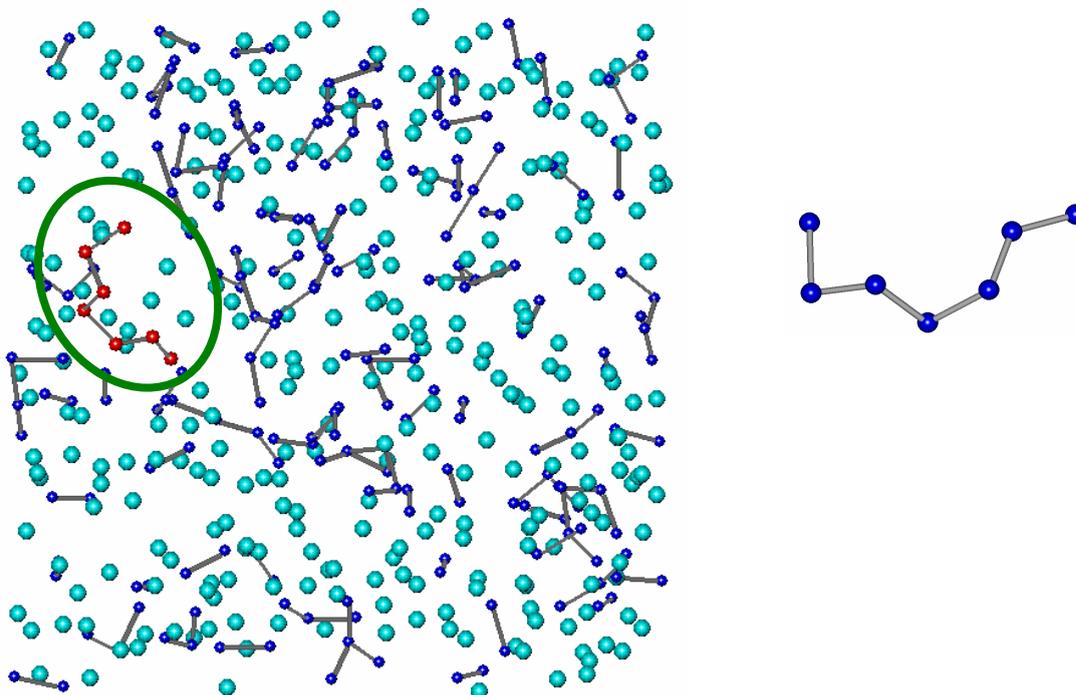


Figura 3.51: Rede de ligações de hidrogênio (esquerda) e o maior agregado da rede (direita) no metanol puro à 500 bar 400°C. Cada círculo representa uma molécula de metanol. O volume da caixa é $(42,85 \text{ \AA})^3$.

Como podemos perceber, a visualização da rede em duas dimensões não nos permite observar detalhes importantes, o que ressalta a aplicação de estudos quantitativos para uma

melhor compreensão da estrutura do líquido, assim como foi realizado para a água. A Tabela 3.25 apresenta os resultados acerca dos parâmetros indicativos da presença de padrões de *small-world*, com dados obtidos para o coeficiente de agregação, comprimento do caminho médio e o número de conexões para 500 estruturas de redes de ligações de hidrogênio nas condições ambiente e na condição supercrítica (500 bar 400°C), assim como avaliada para a água.

A rede de ligações de hidrogênio no metanol líquido em condições ambiente, ao contrário da água nestas condições, não percola no sistema. Ao invés disto, apresenta um indicativo da presença de padrões de *small-world* tanto na condição ambiente como na condição supercrítica ($L < L_{rand}$ e $C \gg C_{rand}$). A razão C/C_{rand} é ainda aproximadamente seis vezes maior que o valor apresentada para a água líquida. Entretanto esta análise não pode ser considerada definitiva do ponto de vista estatístico. Para tal pode-se realizar novos cálculos para o metanol considerando o sistema maior ($N > 512$). O número de moléculas igual a 512 mostrou-se ideal para água, mas os resultados sugerem uma análise específica para definir o N apropriado para o metanol.

Tabela 3.25: Comprimento do caminho médio (L and L_{rand}), coeficiente de agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio no metanol puro.

	1 bar e 25°C	500 bar e 400°C
L	5,19	2,33
L_{rand}	11,76	9,99
L/L_{rand}	0,44	0,23
C	1,6595	0,5355
$C_{rand} = p$	0,0032	0,0010
C/C_{rand}	518,6	535,6
$\langle k \rangle$	1,65	0,536

As Figuras 3.52 e 3.53 mostram o grau de distribuição das ligações de hidrogênio no metanol puro nas condições ambiente e supercrítica para N igual a 512.

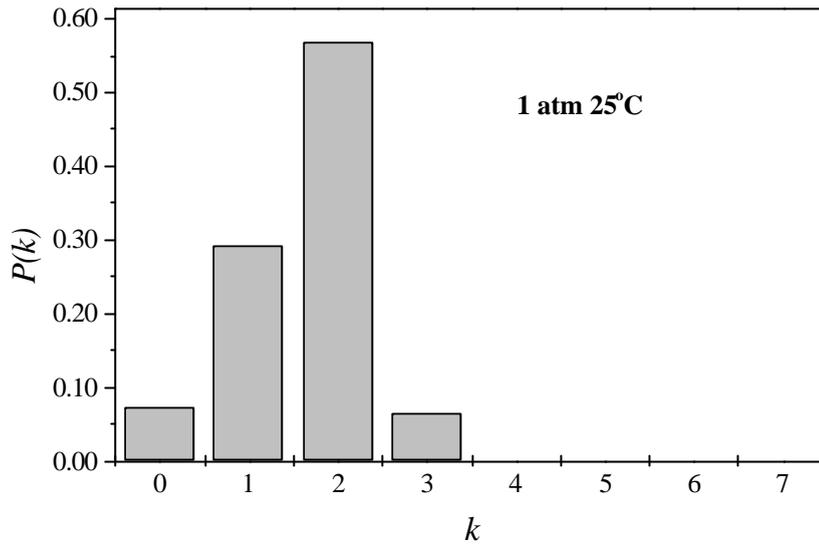


Figura 3.52: Grau de distribuição das ligações de hidrogênio do metanol puro nas condições ambiente.

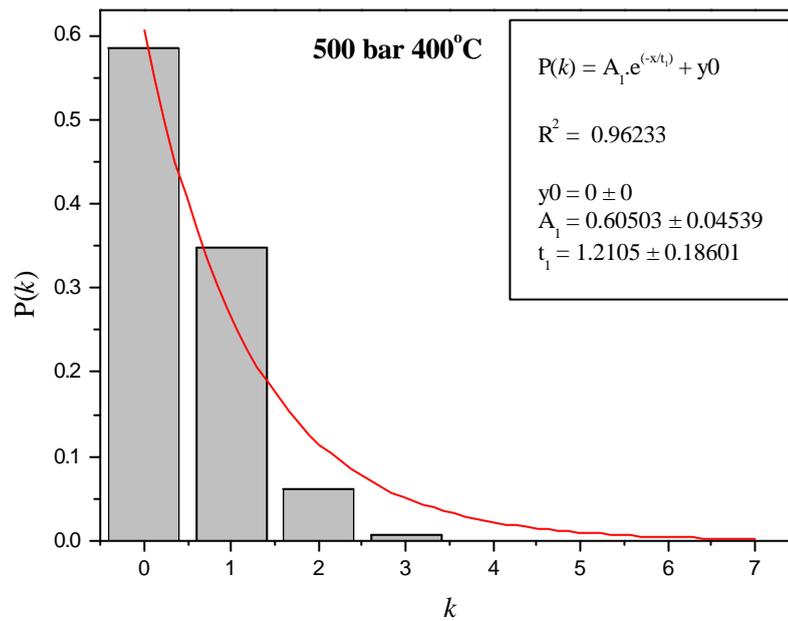


Figura 3.53: Grau de distribuição das ligações de hidrogênio do metanol na condição supercrítica (500 bar 400°C).

Pode-se notar que para o metanol na temperatura ambiente o grau de distribuição apresenta o comportamento observado na Figura 3.52, com média aproximadamente igual 2, ao passo que para o metanol em condição supercrítica a distribuição tem o comportamento de um decaimento exponencial.

A análise de agregados usando estatística de ilhas mostra (considerando n_s igual ao número de ilhas com s componentes) um comportamento claramente distinto para as duas condições analisadas (Figuras 3.54 e 3.55). Como realizado para a água, a visualização

gráfica parte da consideração de que para cada *bin* tem-se ilhas com número de componentes entre: $2^{bin-1} - (2^{bin} - 1)$.

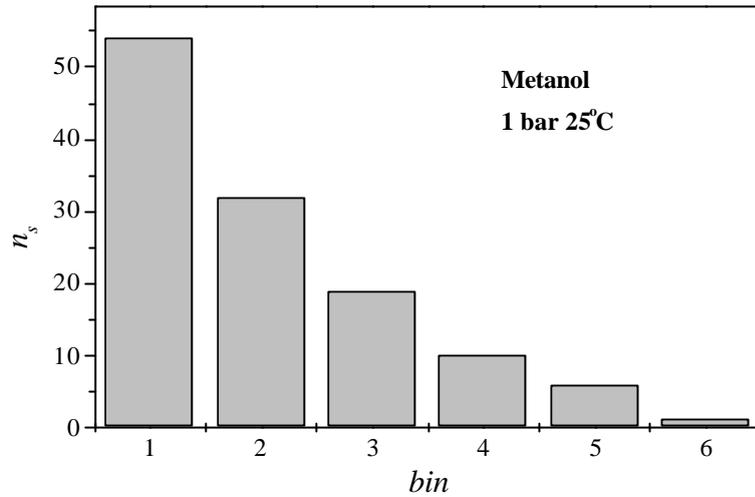


Figura 3.54: Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação do metanol nas condições ambiente.

Na Figura 3.54 verifica-se a existência de 54 monômeros entre as 512 moléculas de metanol. O total de moléculas formando monômeros, dímeros ou trímeros é igual a 118, correspondendo a 46% do total de moléculas. Esta observação indica a formação de uma grande quantidade de pequenos agregados no metanol líquido nas condições ambiente, ao contrário do verificado para a água. Para a condição supercrítica este comportamento da estrutura é acentuado, apresentando 59% de monômeros e 13% de dímeros e trímeros, com a formação de agregados com tamanho máximo igual a sete componentes apenas (Figura 3.51 e Tabela 3.26).

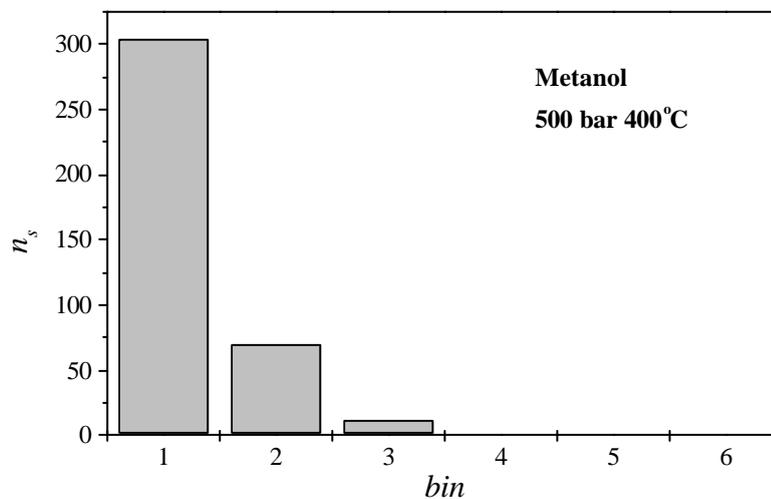


Figura 3.55: Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das 500 caixas de simulação do metanol à 500 bar 400°C.

Tabela 3.26: Tamanho máximo médio das ilhas para as 500 caixas obtidas nas duas condições analisadas.

p (bar)	T (°C)	Ilha de tamanho máximo médio
1	25	41
500	400	7

Os resultados da análise espectral sobre as 500 caixas de simulação equilibradas do metano líquido nas condições ambiente e supercrítica são apresentados nas Figuras 3.56 e 3.57, respectivamente.

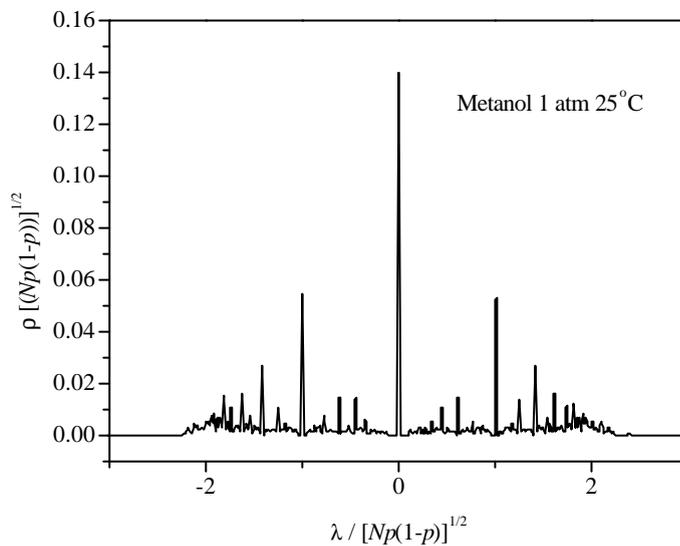


Figura 3.56: Densidade espectral média reescalada para as redes de ligações de hidrogênio do metanol nas condições ambiente.

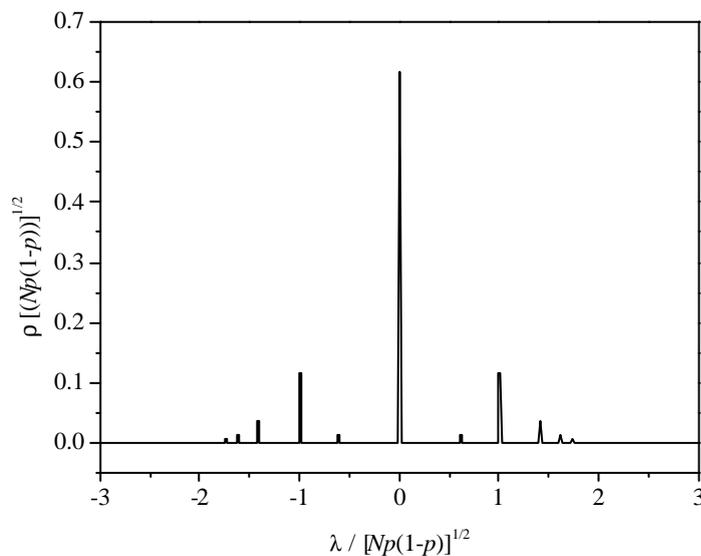


Figura 3.57: Densidade espectral média reescalada para as redes de ligações de hidrogênio do metanol na condição extrema (500 bar 400°C).

Observa-se das Figuras 3.56 e 3.57 que a densidade espectral para o metanol líquido em condições ambiente apresenta contribuições significativas dos segundo, terceiro e quarto momentos, ao contrário do resultado obtido para a água líquida. O metanol a temperatura ambiente não forma um agregado gigante. Para o metanol supercrítico, a densidade espectral tem contribuições significantes do segundo e terceiro momentos.

3.5. Simulação Computacional da Mistura Água-Metanol 1:1

Para a mistura binária, a simulação de Monte Carlo foi realizada usando o modelo TIP5P para a água e o modelo de três sítios de Jorgensen¹²² para o metanol. Utilizou-se 256 moléculas de água e 256 moléculas de metanol, ou seja, uma mistura 1:1. As condições de simulação estão apresentadas na Tabela 3.27. A variação de volume ocorreu a cada 2000 configurações.

Tabela 3.27: Simulação de Monte Carlo para o sistema água-metanol 1:1.

	Condições ambiente	Condições extremas
Número de moléculas de água e de metanol	256	256
Raio de corte (?)	13	15
Nº Configurações mínimo (equilíbrio)	8 x 10 ⁶	12 x 10 ⁶
Nº Configurações mínimo (aquisição)	10 x 10 ⁶	10 x 10 ⁶

3.5.1. Propriedades Termodinâmicas

Algumas propriedades termodinâmicas desta mistura são apresentadas (Tabelas 3.28 e 3.29) juntamente com aquelas obtidas a partir das simulações dos líquidos puros para facilitar a análise comparativa.

Tabela 3.28: Propriedades termodinâmicas nas condições ambiente (1 atm 25°C).

Sistema	ρ (g cm ⁻³)	-E/N (kcal mol ⁻¹)	C _p (cal mol ⁻¹ K ⁻¹)
Água pura	1,005	9,96	22,50
Metanol puro	0,766	8,59	14,85
Água : Metanol 1:1	0,762	9,41	21,35

¹²² W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, *J. Chem. Phys.* 79: 926 (1983).

Verificando a Tabela 3.28 observa-se que a energia média por molécula e a capacidade calorífica da mistura binária são intermediárias aquelas dos líquidos puros. A densidade da mistura manteve-se próxima da densidade do metanol na condição ambiente. As Tabelas 3.29 e 3.30 apresentam as propriedades dos líquidos puros e da mistura em duas condições supercríticas e a Tabela 3.31 apresenta as demais condições simuladas. As energias obtidas mostram que em todas as condições obtêm-se um líquido com energia de interação mais fraca que a interação média entre as moléculas presentes na água líquida, entretanto mais forte que a interação média presente no metanol líquido puro.

Tabela 3.29: Propriedades termodinâmicas à 500 bar e 300°C.

Sistema	ρ (g cm ⁻³)	-E/N (kcal mol ⁻¹)	Cp (cal mol ⁻¹ K ⁻¹)
Água pura (TIP5P)	0,602	5,34	18,60
Metanol puro	0,497	4,33	11,76
Água : Metanol 1:1	0,528	4,89	10,85

Tabela 3.30: Propriedades termodinâmicas à 500 bar e 400°C.

Sistema	ρ (g cm ⁻³)	-E/N (kcal mol ⁻¹)	Cp (cal mol ⁻¹ K ⁻¹)
Água pura (TIP5P)	0,36	3,61	14,58
Metanol puro	0,37	2,95	9,20
Água : Metanol 1:1	0,35	3,17	10,79

Tabela 3.31: Densidade e energia média por molécula da mistura água-metanol nas condições extremas analisadas.

p (bar)	T (°C)	ρ (g cm ⁻³)	-E/N (kcal mol ⁻¹)
500	100	0,815	8,48
500	200	0,685	6,62
500	500	0,239	2,06

3.5.2. Funções de Distribuição Radial

As funções de distribuição radial, em condições ambiente, da água pura e na mistura são apresentadas nas Figuras 3.58-3.60, seguida pelas funções de distribuição radial do metanol puro e na mistura nas Figuras 3.61-3.63.

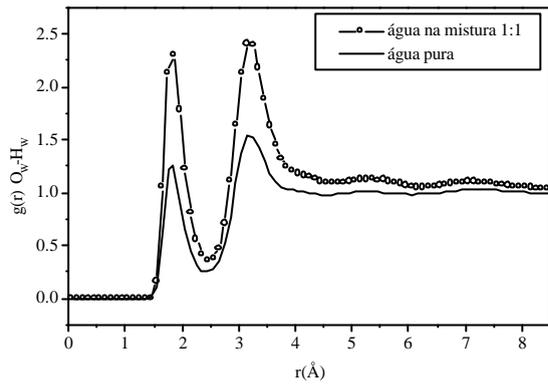


Figura 3.58: RDFs O-H para a água pura e na mistura 1:1 nas condições ambiente.

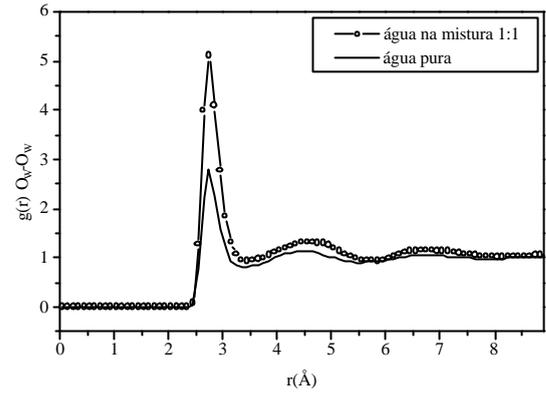


Figura 3.59: RDFs O-O para a água pura e na mistura 1:1 nas condições ambiente.

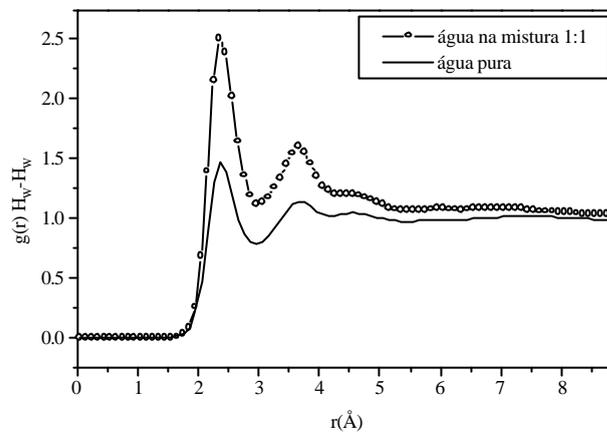


Figura 3.60: RDFs H-H para a água pura e na mistura 1:1 nas condições ambiente.

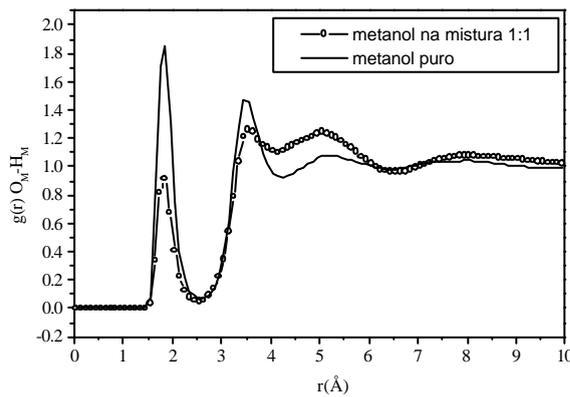


Figura 3.61: RDFs O-H para o metanol puro e na mistura 1:1 nas condições ambiente.

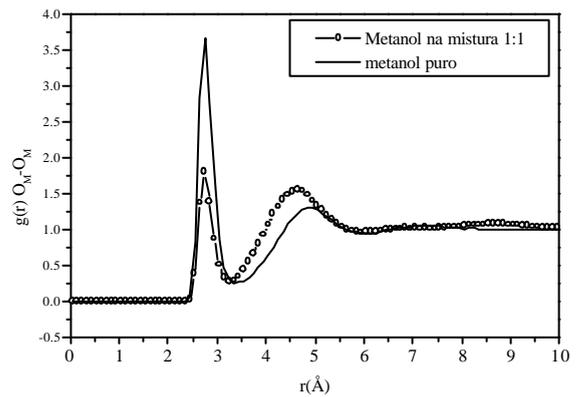


Figura 3.62: RDFs O-O para o metanol puro e na mistura 1:1 nas condições ambiente.

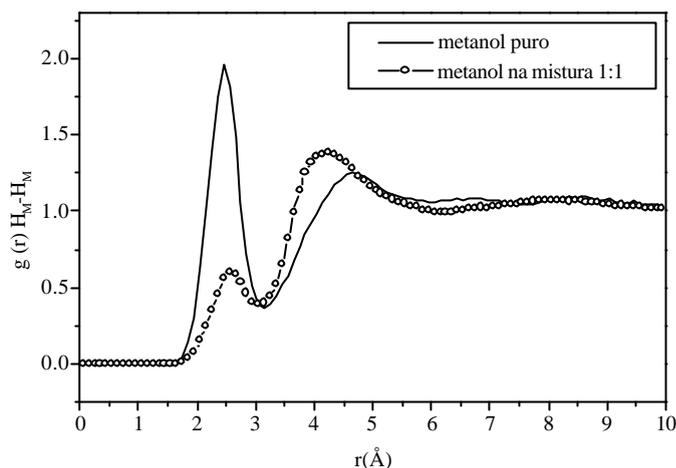


Figura 3.63: RDFs H-H para o metanol puro e na mistura 1:1 nas condições ambiente.

Comparando-se as RDFs obtidas para a água e para o metanol, verifica-se que a altura do primeiro pico aumenta no caso da água e diminui no caso do metanol quando misturados. Portanto, na mistura água-metanol 1:1, as moléculas de água estão mais estruturadas do que no líquido puro e formam um maior número de ligações de hidrogênio entre dímeros, pois a altura do primeiro pico para as RDFs O-H (1,85 ?) é menor que a distância necessária entre os átomos de oxigênio e hidrogênio necessária para a formação da ligação de hidrogênio (2,0-2,4 ?). Ao passo que no caso do metanol observa-se um comportamento menos estruturado das moléculas e diminuição das possíveis interações de ligações de hidrogênio, visto que a altura do primeiro pico é menor. Este resultado concorda com resultados obtidos por X. Wu *et al.* (1992)¹²³.

As RDFs obtidas para as misturas água-metanol em uma das condições críticas analisadas (500 bar e 300°C) são apresentadas nas Figuras 3.64-3.66 para a água e Figuras 3.67-3.69 para o metanol. Estas RDFs são comparadas com aquelas obtidas para os líquidos puros nas mesmas condições extremas. As RDFs entre os átomos da molécula de água e átomos do metanol são apresentadas em seguida, comparando-se aquelas obtidas em condições ambiente com aquelas obtidas à 500 bar e 300°C (Figuras 3.70-3.72).

¹²³ X. Wu, Y. Li, J. Lu, T. Teng, *Fluid Phase Equilibria* 77: 139-156 (1992).

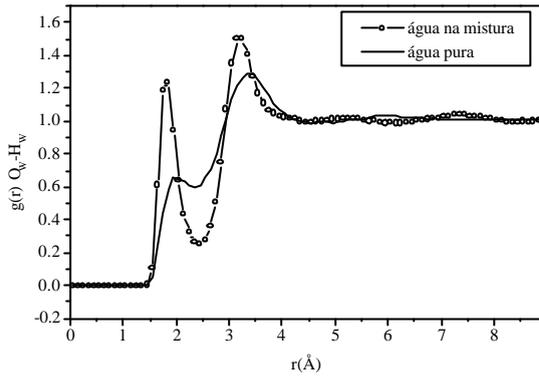


Figura 3.64: RDFs O-H para a água pura e na mistura 1:1 em 500 bar 300°C.

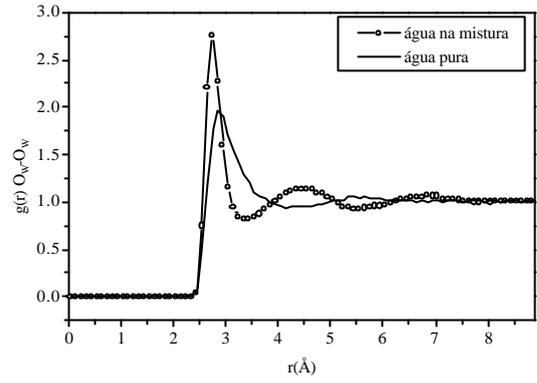


Figura 3.65: RDFs O-O para a água pura e na mistura 1:1 em 500 bar 300°C.

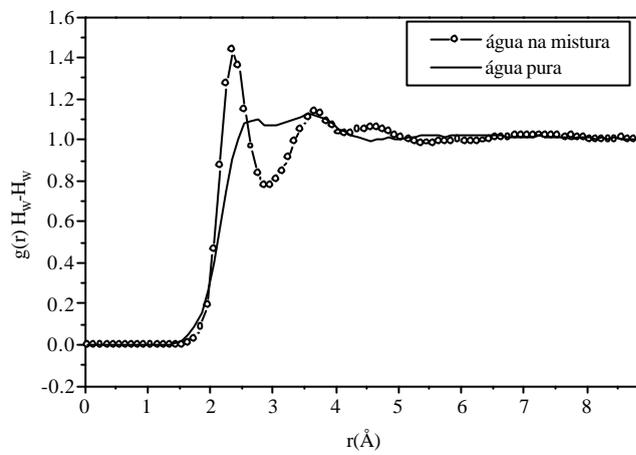


Figura 3.66: Valores das RDFs H-H para a água pura e na mistura 1:1 em 500 bar 300°C.

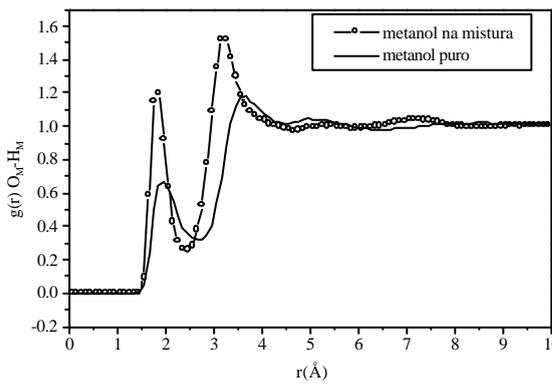


Figura 3.67: RDFs O-H para o metanol puro e na mistura 1:1 em 500 bar 300°C.

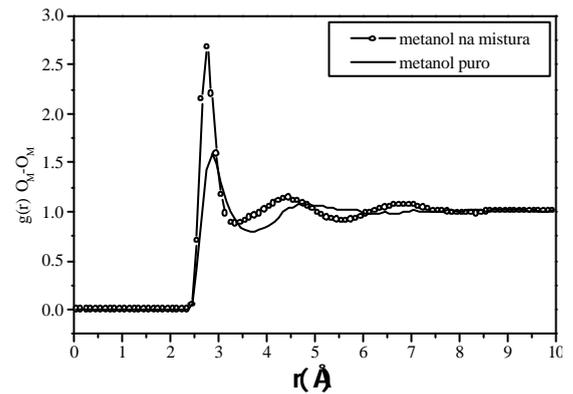


Figura 3.68: RDFs O-O para o metanol puro e na mistura 1:1 em 500 bar 300°C.

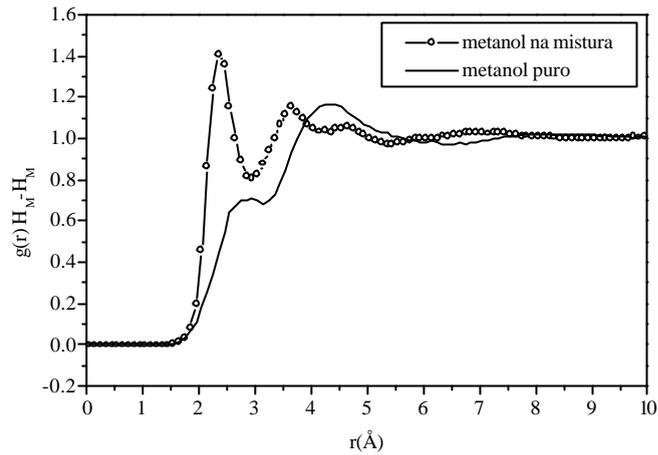


Figura 3.69: RDFs H-H para o metanol puro e na mistura 1:1 em 500 bar 300°C.

Em condições críticas observa-se uma maior estruturação das moléculas de água (água-água) e do metanol (metanol-metanol), quando na mistura 1:1 do que quando puros, ao contrário do que observamos em condições ambiente. Também verifica-se que as moléculas de água e metanol (água-metanol) estão menos estruturadas nestas condições. Estes resultados fornecem indicações para a ocorrência de uma possível separação de fases na mistura em condições extremas de temperatura e pressão (Figuras 3.70-3.72).

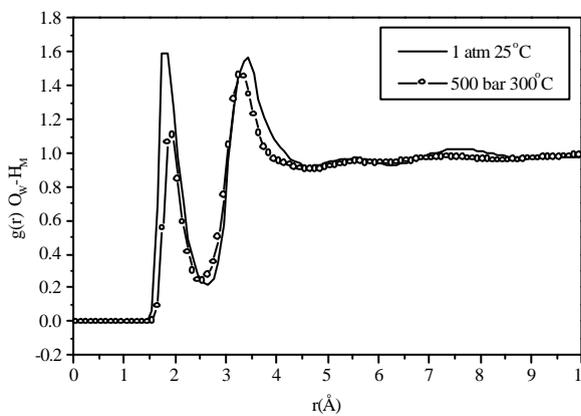


Figura 3.70: RDFs entre o átomo O da água e o átomo H do metanol em condições ambiente e críticas.

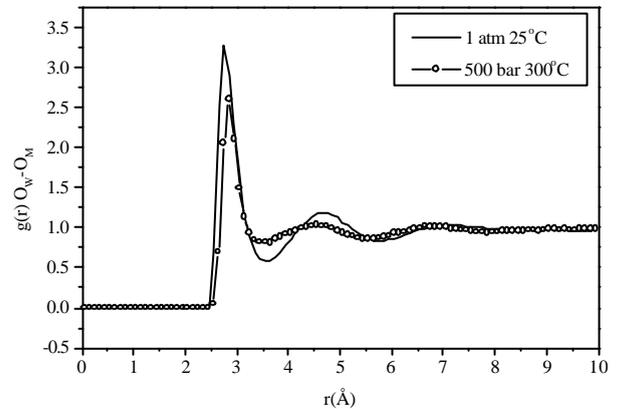


Figura 3.71: RDFs entre o átomo O da água e o átomo O do metanol em condições ambiente e críticas.

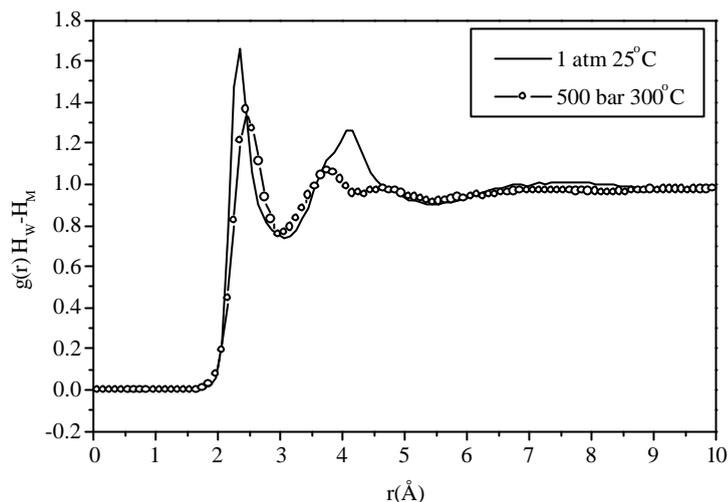


Figura 3.72: RDFs entre o átomo H da água e o átomo H do metanol em condições ambiente e críticas.

Cabe ressaltar que as RDF's obtidas para a condição de 500 bar 400°C têm comportamentos qualitativos e quantitativos semelhantes aqueles apresentados acima.

3.5.3 Análise Topológica das Redes de Ligações de Hidrogênio na Mistura equimolar de Água : Metanol

As ligações de hidrogênio entre as moléculas de água e metanol consideradas estão entre os dois tipos de interação: ocorre a interação entre o hidrogênio da água e o átomo de oxigênio do metanol ou a interação entre o oxigênio da água e o hidrogênio do metanol.

Na Tabela 3.32 estão apresentados os resultados sobre a presença de padrões de *small-world*, com dados obtidos para o coeficiente de agregação, comprimento do caminho médio e o número de conexões para 500 estruturas de redes de ligações de hidrogênio nas condições ambiente e em uma das condições supercríticas (500 bar 400°C).

Tabela 3.32: Comprimento do caminho médio (L and L_{rand}), coeficiente de agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio na mistura água-metanol 1:1.

	1 bar e 25°C	500 bar e 400°C
L	27,59	2,74
L_{rand}	6,81	59,2
L/L_{rand}	2,5330	0,0463
C	1,1425	0,3995
$C_{rand} = p$	0,0049	0,0017
C/C_{rand}	233,2	235,0
$\langle k \rangle$	2,5	0,9

A Figura 3.73 apresenta a mistura água-metanol à líquido à temperatura e pressão ambientes, com o maior agregado formado mostrado à direita. Cada círculo (sítio) vermelho e verde representa uma molécula de água e metanol, respectivamente. As conexões representam as ligações de hidrogênio.

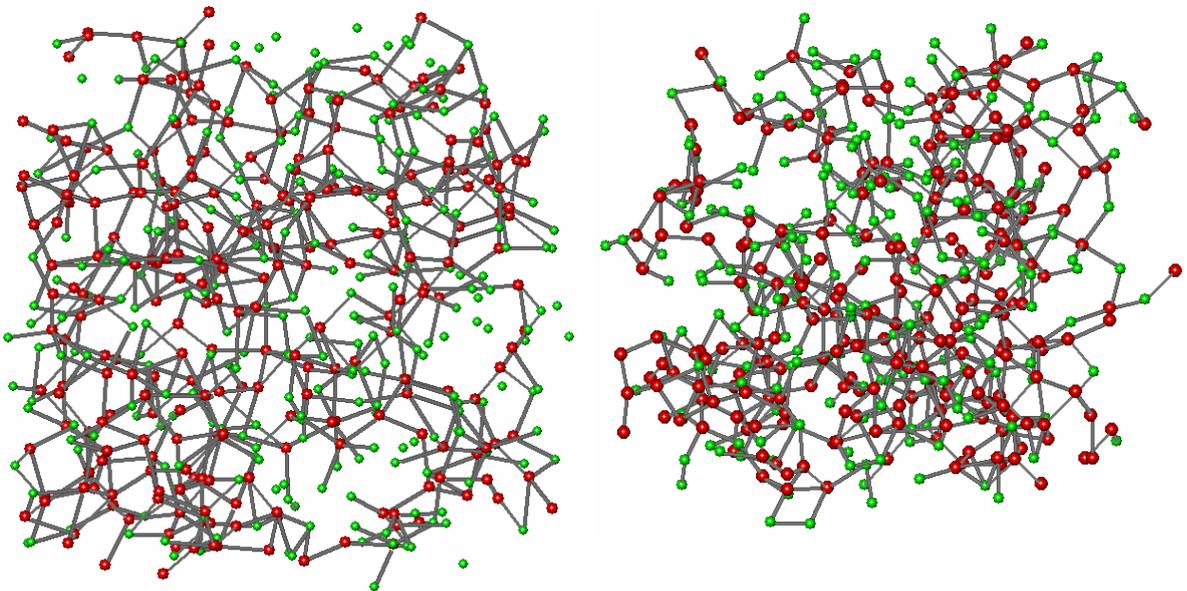


Figure 3.73: Rede de ligações de hidrogênio da mistura água-metanol 1:1 em condição ambiente. O volume da caixa é $(30,10 \text{ \AA})^3$.

O maior agregado à direita na Figura 3.73 indica que ocorre percolação no sistema, de modo que o limiar de percolação para a mistura binária água-metanol 1:1 deve ser igual a um valor compreendido entre 0,9 e 2,5, ou seja, um valor intermediário entre as conectividades médias calculadas em condições ambiente e supercrítica. Nesta condição,

apesar da mistura apresentar uma elevada razão C/C_{rand} verificamos que L é maior que L_{rand} , de modo que nestas condições a mistura água-metanol não apresenta padrão de *small-world*.

Podemos observar que o número de monômeros é significativamente maior para o metanol (círculos verdes). Em condições extremas (Figura 3.74) são formados agregados menores e elevado número de monômeros, dímeros e trímeros.

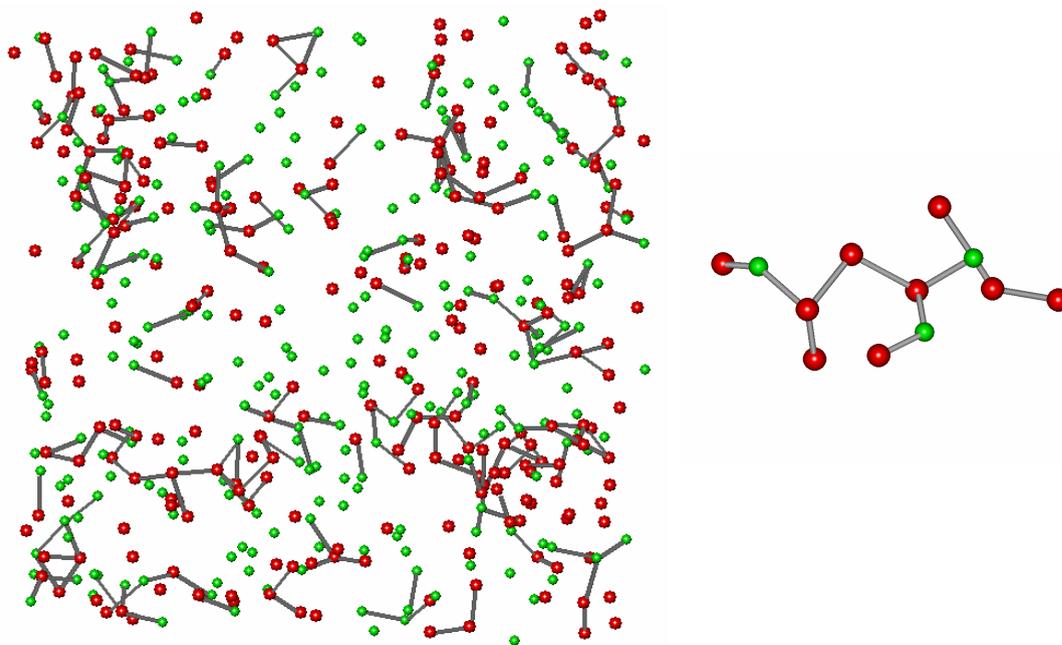


Figure 3.74: Rede de ligações de hidrogênio (esquerda) e o maior agregado da rede (direita) para a mistura água-metanol em condição supercrítica (500 bar 400°C). O volume da caixa é $(40,19 \text{ \AA})^3$.

Na mistura formam-se ligações de hidrogênio água-água, metanol-metanol ou água-metanol. Cada molécula de água presente na mistura em condições ambiente apresenta em média o grau de distribuição das ligações de hidrogênio como mostrado na Figura 3.75 à esquerda. Podemos concluir que nestas condições, em média as moléculas de água formam 3,3 ligações de hidrogênio. Ao passo que cada molécula de metanol (Figura 3.75 - direita) forma 1,8 ligações de hidrogênio em média. Analisando o grau de distribuição com relação a todas as moléculas presentes na mistura binária líquida obtém-se o comportamento do grau de distribuição mostrado na Figura 3.76, obtida também no caso da água líquida pura.

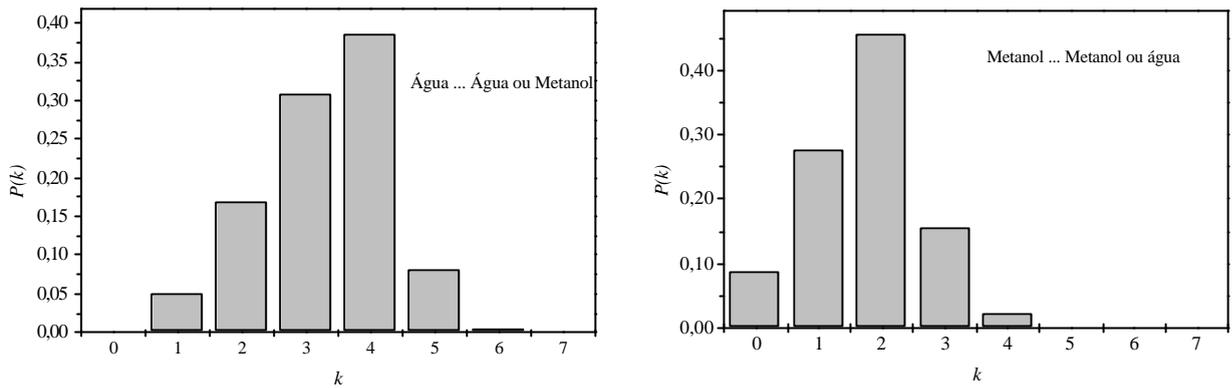


Figura 3.75: Grau de distribuição das ligações de hidrogênio das moléculas de água (esquerda) e metanol (direita) na mistura água-metanol 1:1 nas condições ambiente.

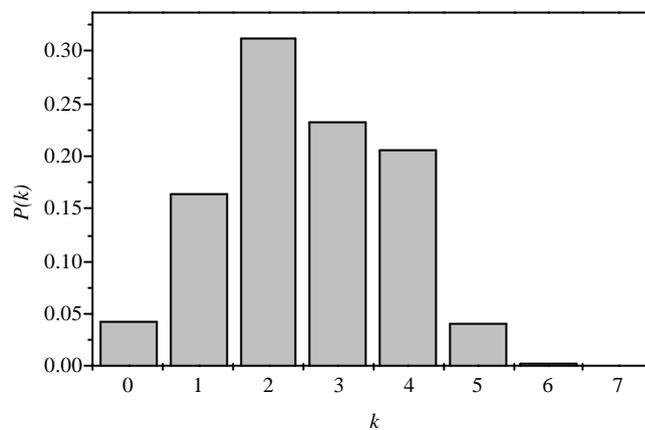


Figura 3.76: Grau de distribuição das ligações de hidrogênio de todas as moléculas presentes na mistura água-metanol 1:1 nas condições ambiente.

O grau de distribuição das ligações de hidrogênio na condição supercrítica é apresentado na Figura 3.77 para a água (esquerda) e metanol (direita). A conectividade média diminui. Analisando o grau de distribuição com relação a todas as moléculas presentes na mistura binária líquida obtém-se o comportamento do grau de distribuição mostrado na Figura 3.78. Esta distribuição é assimétrica, com comportamento semelhante ao decaimento exponencial, a partir de k igual a 1.

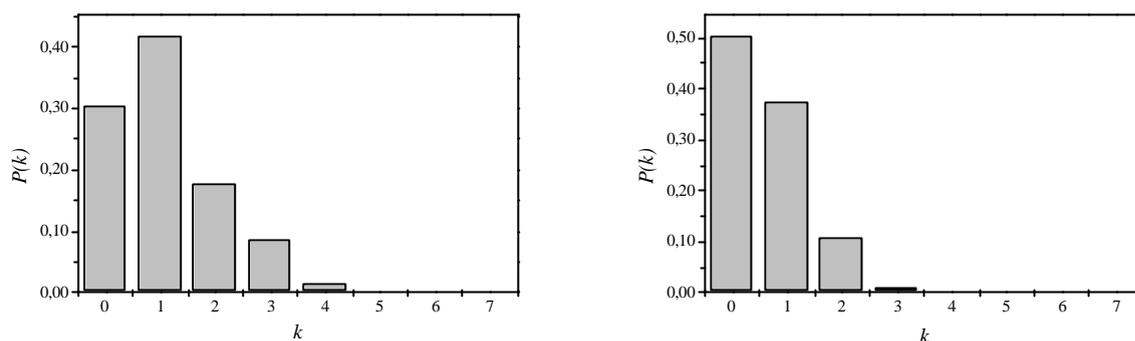


Figura 3.77: Grau de distribuição das ligações de hidrogênio das moléculas de água (esquerda) e metanol (direita) na mistura água-metanol 1:1 à 500 bar e 400°C.

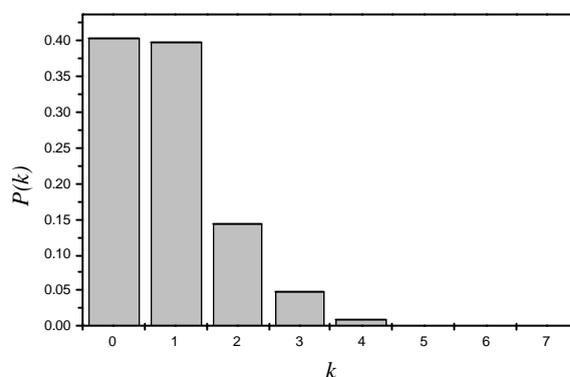


Figura 3.78: Grau de distribuição das ligações de hidrogênio de todas as moléculas presentes na mistura água-metanol 1:1 à 500 bar e 400°C.

Assim como para os líquidos puros, a análise de agregados da mistura usando estatística de ilhas apresenta um comportamento claramente distinto para as duas condições analisadas. Analogamente aos resultados apresentados nos tópicos 3.2.5 e 3.3.3 a visualização gráfica considera que para cada *bin* tem-se ilhas com número de componentes entre: $2^{bin-1} - (2^{bin} - 1)$. Na Figura 3.80 têm-se a estatística de ilhas para a mistura nas duas condições analisadas e para avaliar as diferenças entre os líquidos puros e a mistura os resultados são apresentados simultaneamente na Figura 3.80.

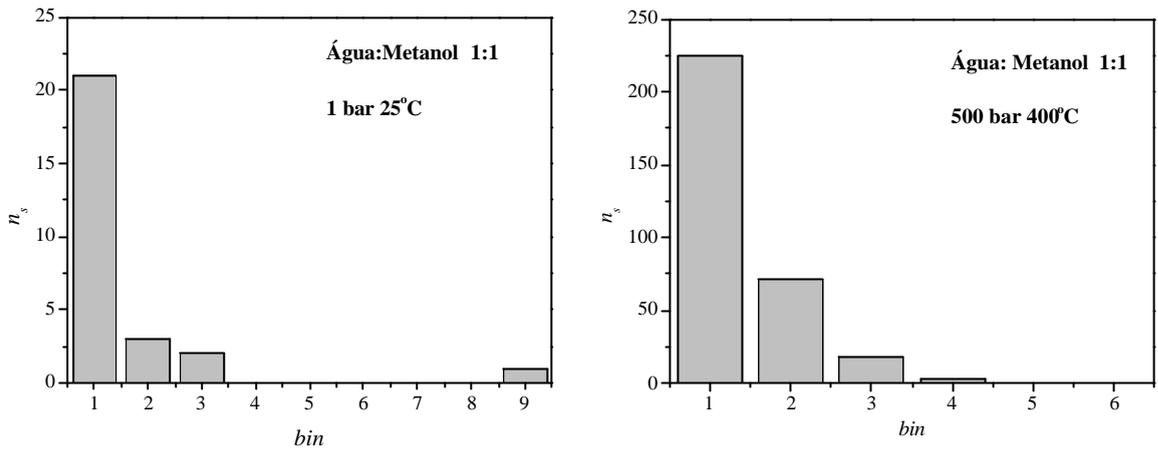


Figura 3.79: Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para uma das caixas de simulação da mistura binária água-metanol nas condições ambiente (esquerda) e na condição supercrítica (direita).

Na mistura, a estatística mostra um comportamento intermediário àquele observado para os líquidos puros, como pode ser visualizado melhor na Figura 3.80.

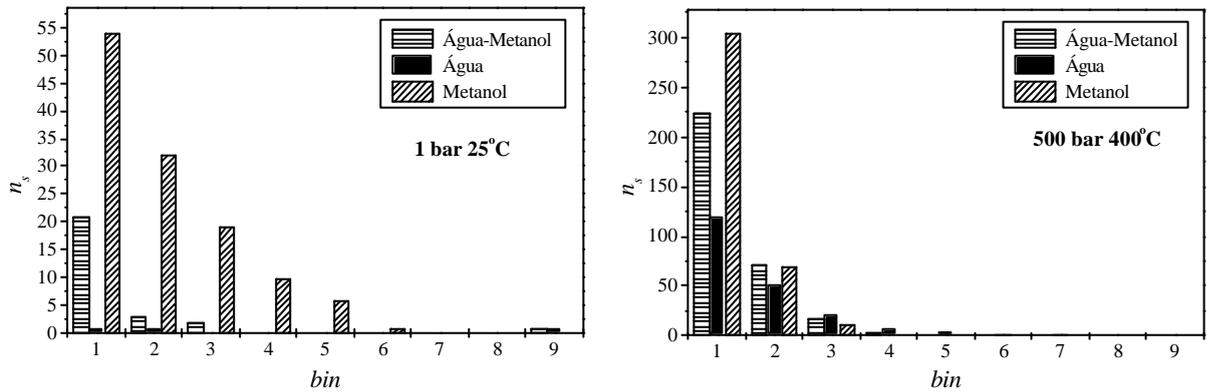


Figura 3.80: Número de ilhas com uma quantidade de componentes entre: $2^{bin-1} - (2^{bin} - 1)$ para caixas de simulação contendo líquidos puros e a mistura à temperatura ambiente (esquerda) e à 500 bar e 400°C (direita).

A Tabela 3.33 apresenta as ilhas de tamanho máximo médio a partir da análise das 500 caixas de simulação geradas. O maior agregado nas condições extremas contém um comprimento de caminho médio significativamente pequeno em comparação com aquele do agregado formado em condições ambiente, evidenciando a existência de padrão de *small-world* em condições extremas (observar Figuras 3.73 e 3.74).

Tabela 3.33: Tamanho máximo médio das ilhas para as 500 caixas obtidas nas duas condições analisadas.

p (bar)	T (°C)	Ilha de tamanho máximo médio
1	25	475
500	400	13

As densidades espectrais reescaladas de 500 redes de ligações de hidrogênio para a água líquida nas condições ambiente e supercrítica são apresentadas nas Figura 3.81 e 3.82.

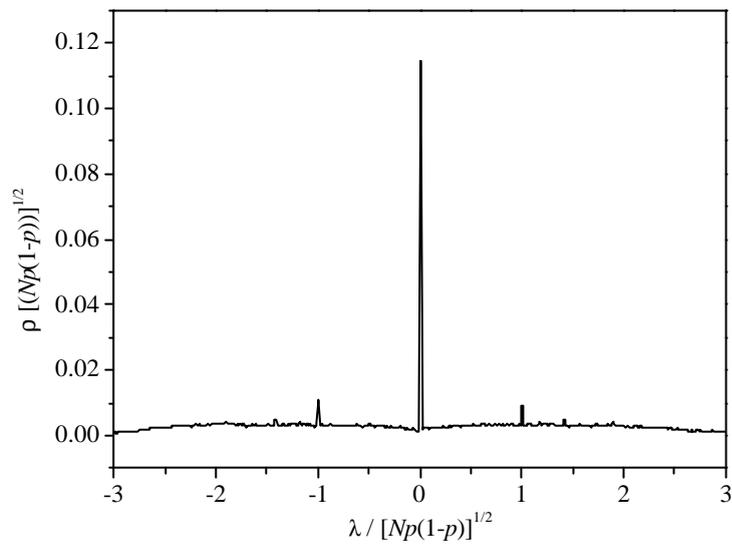


Figura 3.81: Densidade espectral média reescalada para as redes de ligações de hidrogênio da mistura água-metanol 1:1 nas condições ambiente.

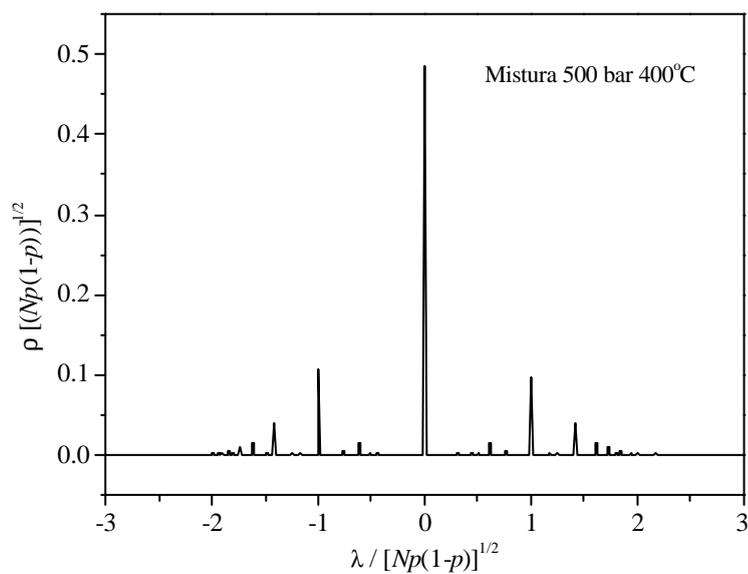


Figura 3.82: Densidade espectral média reescalada para as redes de ligações de hidrogênio da mistura água-metanol 1:1 na condição extrema (500 bar 400°C).

A análise da densidade espectral das ligações de hidrogênio presentes na mistura em condições ambiente, comparando com o resultado obtido para a água pura, indica que o número de sítios ou moléculas da rede ($N = 512$) da mistura equivale a escala obtida para a água usando 300 moléculas. Um aumento de N para a mistura levaria a uma melhor representação do comportamento randômico na condição ambiente. Ao passo que a densidade espectral da mistura na condição supercrítica aproxima-se tanto do comportamento observado para a água como para o metanol puro. Portanto tanto os líquidos puros como a mistura apresentam formação de estruturas, ou seja, não apresentam comportamento randômico.

3.6. Líquidos Iônicos

Como discutido na introdução, os líquidos iônicos estão se tornando cada vez mais importantes como meios alternativos para reações químicas que não prejudicam o meio ambiente. Trabalhos vêm sendo publicados no desenvolvimento de campos de força e validações para simulação destes líquidos¹²⁴, mas ainda são poucas as simulações computacionais destes compostos, bem como o grande número de átomos envolvidos permite um teste para aplicação de modelos baseados em RNAs, daí a motivação para o estudo destes sistemas.

3.6.1. Análise conformacional do cátion BMI (1,3-*n*-butil-metilimidazólio)

O cátion BMI (1,3-*n*-butil-metilimidazólio), Figura 3.83, foi escolhido para o trabalho inicial, tanto por sua importância, como pela possibilidade de interação com outros grupos de pesquisa. Quanto aos ânions, escolheu-se o BF_4^- , Figura 3.84, e Cl^- . Outros exemplos de ânions tiveram suas estruturas otimizadas (Figura 3.85) para estudos posteriores.

Os cálculos *ab initio* para otimização das estruturas moleculares dos cátions e ânions foram realizadas utilizando o *Gaussian 98*¹²⁵, com método MP2/6-31G(d,p), incluindo funções difusas para os ânions.

¹²⁴ J. N. C. Lopes, J. Deschamps, A. A. H. Pádua, *J. Phys. Chem. B*, 108: 2038-2047 (2004). J. Andrade, E.S. Bões, H. Stassem, *J. Phys. Chem. B*, 106: 13344-13351 (2002).

¹²⁵ M. J. Frisch, G. W. Trucks, *et al.*, *Manual Gaussian 98*, Pittsburgh, PA (1998).

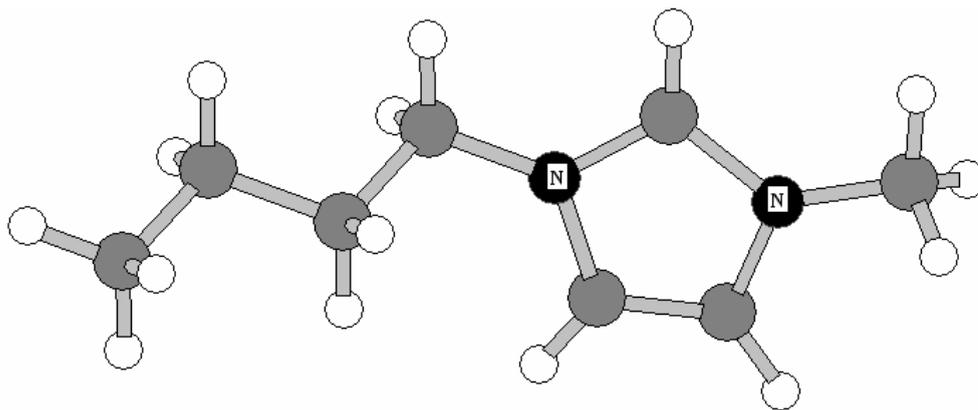


Figura 3.83: Estrutura do cátion 1,3-*n*-butil-metilimidazólio (BMI) - MP2/6-31G(d,p).

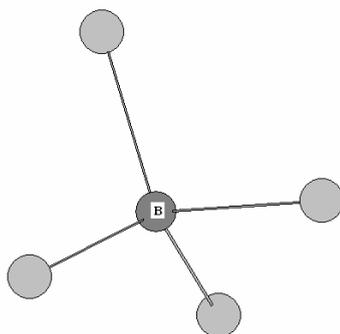


Figura 3.84: Estrutura do ânion BF_4^- - MP2/6-31+G(d,p).

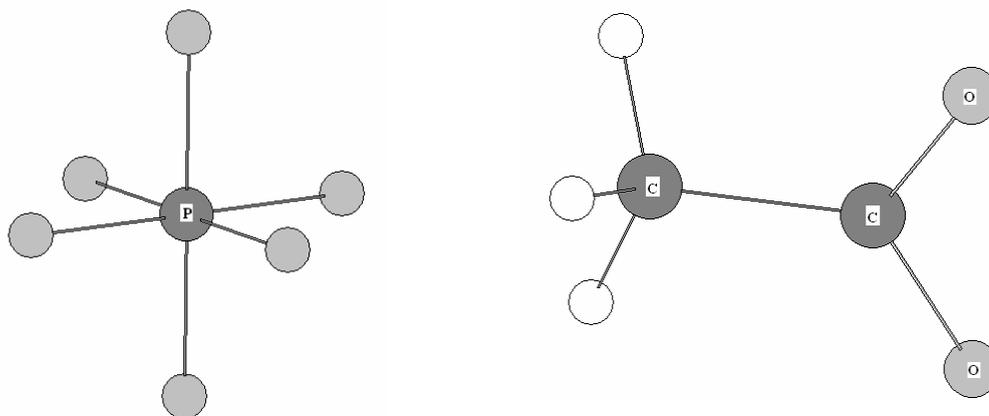


Figura 3.85: Estrutura dos ânions PF_6^- (esquerda) e CH_3COO^- (direita).

O cadeia *n*-butil ligada ao anel do cátion imidazólio gera diversas conformações estáveis e, portanto, estruturas do sal menos organizadas à temperatura ambiente, daí a obtenção dos líquidos iônicos. Como as conformações influenciam nas propriedades e comportamento dos fluidos, realizou-se uma análise conformacional usando o método de mecânica molecular MM^+ , variando-se os ângulos f_1 , f_2 e f_3 (Figura 3.86).

Os resultados da análise conformacional permitiram obter 27 conformações, com energias entre 22,3 e 23,8 kcal mol⁻¹, das quais foram escolhidas três estruturas distintas,

Tabela 3.34, que foram posteriormente otimizadas com o método MP2/6-31G(d,p). Os dados de energia de interação servirão para treinar mais três RNAs que poderão ser incluídas nas simulações de Monte Carlo.

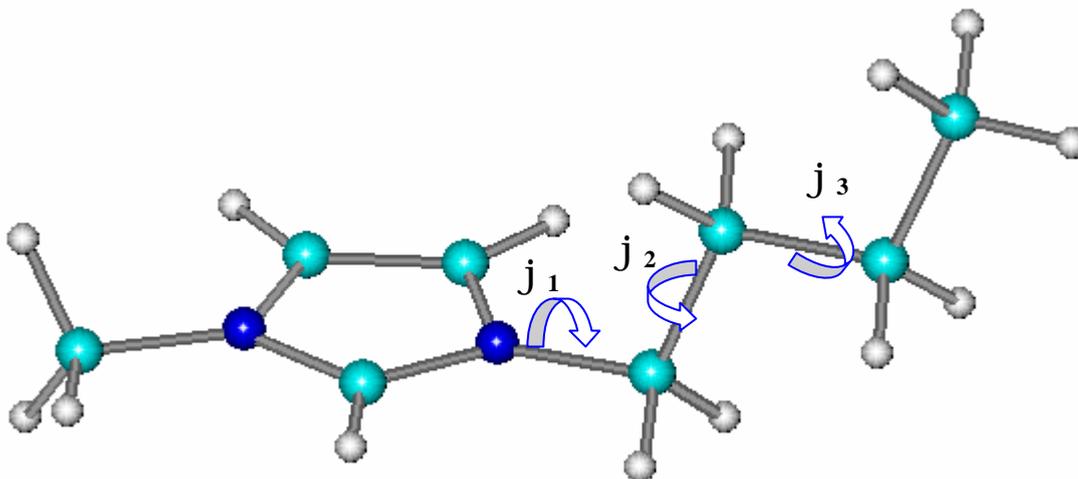


Figura 3.86: Ângulos diédricos f_1 , f_2 e f_3 para análise conformacional do cátion BMI.

Tabela 3.34: Resultados da análise conformacional do BMI usando MM⁺.

Cátion	f_1	f_2	f_3	Energia (kcal mol ⁻¹)
1	75,21	-63,51	177,03	22,44
2	102,37	63,24	177,12	22,45
3	73,21	59,87	66,22	22,97

3.6.2. Simulações de Monte Carlo Preliminares

As simulações de Monte Carlo foram executadas em um *ensemble NpT* usando um total de 250 cátions e 250 ânions em uma caixa de simulação cúbica aplicando condições de contorno periódicas. As cargas atômicas parciais selecionadas para o cátion foram obtidas pelo ajuste da superfície de potencial eletrostático (ESP – *electrostatic surface potential*) usando *CHelpG* através do módulo ESP do programa *NWChem 4.1*¹²⁶ (Figura 3.87) e os protocolos de simulação ajustados para as simulações em condições ambiente são apresentados na Tabela 3.35. O número de configurações utilizado para a equilibração e aquisição foram pelo menos 30×10^6 e 30×10^6 , respectivamente. O raio de corte adotado é igual a 17 Å.

¹²⁶ High Performance Computational Chemistry Group. *NWChem*, Richland, WA 99352 (2002).

Comparando o volume máximo para variação de volume verificou-se que mesmo para um valor muito pequeno (7% do valor utilizado para simulação da água – 186), a taxa de aceitação é de apenas 5%.

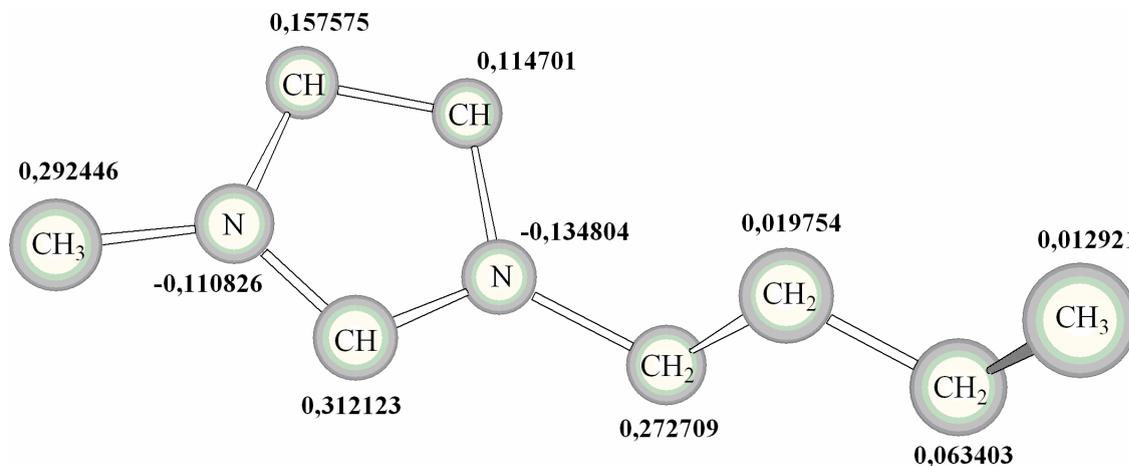


Figura 3.87: Cargas ajustadas usando o módulo ESP do programa *NWCHEM 4.1*.

Tabela 3.35: Protocolos de simulação incluindo razão de aceitação para mudanças translacionais ou rotacionais A (%), razão de aceitação das variações de volume A_v (%), máxima variação de volume $V_{m\acute{a}x}$ e dimensão final da caixa cúbica.

Líquido Iônico	p (bar)	T (°C)	A (%)		A_v (%)	$V_{m\acute{a}x}$	a (?)
			cátion	ânion			
[BMI 1][BF ₄ ⁻]	1	30	~ 0,02	~ 0,15	~ 8	6,0	43,12
[BMI 2][BF ₄ ⁻]	1	30	~ 1,3	~ 14	~ 10	8,0	43,12
[BMI 1][Cl ⁻]	1	25	~ 0,2	~ 0,3	~ 20	0,8	39,80
[BMI 3][Cl ⁻]	1	40	~ 0,9	~ 1,4	~ 16	0,8	39,80

A baixa aceitação para variação do volume acarreta uma pequena variação da densidade inicial.

A simulação considerando o modelo de molécula rígida para os cátions BMI 1 e BMI 2, Tabela 3.34 – confômero 1, gerou a caixa após configurações de equilíbrio com o líquido iônico [BMI 1][BF₄⁻] ilustrada na Figura 3.89.

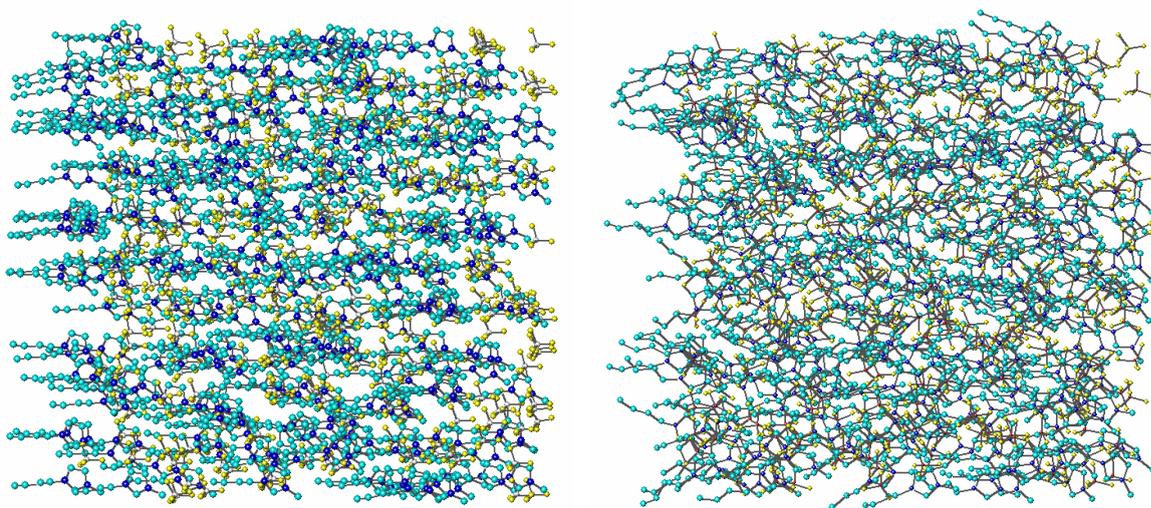


Figura 3.88: Caixa de simulação equilibrado contendo os líquidos iônicos [BMI 1][BF₄⁻] (esquerda) e [BMI 2][BF₄⁻] (direita) nas condições ambiente.

A energia e densidade do líquido [BMI 1][BF₄⁻] estão apresentadas na Tabela 3.36.

Tabela 3.36: Propriedades termodinâmicas obtidas nas simulações de Monte Carlo de líquidos iônicos [BMI 1][BF₄⁻]. Os valores entre parênteses representam o erro relativo porcentual com relação ao valor experimental (1,17 g cm⁻³)¹²⁷.

Líquido Iônico	p (bar)	T (°C)	ρ (g cm ⁻³)	-E/N (kcal mol ⁻¹)
[BMI 1][BF ₄ ⁻]	1	30	1,169 (0,08)	119,3
[BMI 2][BF ₄ ⁻]	1	30	1,169 (0,08)	80,61
[BMI 1][Cl]	1	25	1,149	121,7
[BMI 3][Cl]	1	40	1,149	79,18

Durante as simulações de Monte Carlo verificou-se que a densidade do líquido tem grande dependência com a densidade da caixa de simulação inicial, necessitando de um número grande de configurações para equilíbrio.

¹²⁷ J. de Andrade, E. S. Boes, H. Stassen, *J. Phys. Chem. B.* 106: 13344-13351 (2002), S. Takahashi, K. Suzuya, S. Kohara, N. Koura, L.A. Curtiss, M. L. Z. Saboungi, *Phys. Chem.* 209: 209 (1999).

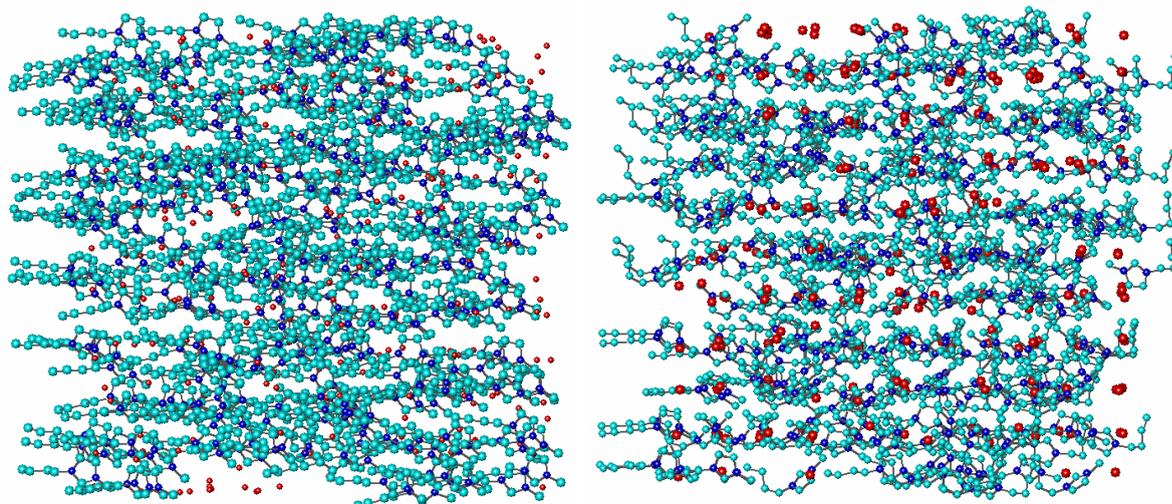


Figura 3.89: Caixa de simulação equilibrada contendo o líquido iônico [BMI 1][Cl] (esquerda) e [BMI 3][Cl] (direita).

A análise das Figuras 3.88 e 3.89 indica que a flexibilidade da cadeia *n-butyl* é uma das responsáveis pela desordem (líquido) dos sais.

3.6.3. Ajuste da energia de interação usando RNAs

As energias de interação *ab initio* MP2/6-31+G(d, p) foram obtidas para 337 configurações do par [BMI][BF₄⁻]. Para isto, utilizou-se o programa desenvolvido para geração de configurações, isto é, para distintas posições do ânion ao redor do cátion, conforme ilustradas na Figura 3.90. Estas energias serão utilizadas para o treinamento de RNAs que serão utilizadas para obtenção de superfícies de energia potencial destes líquidos.

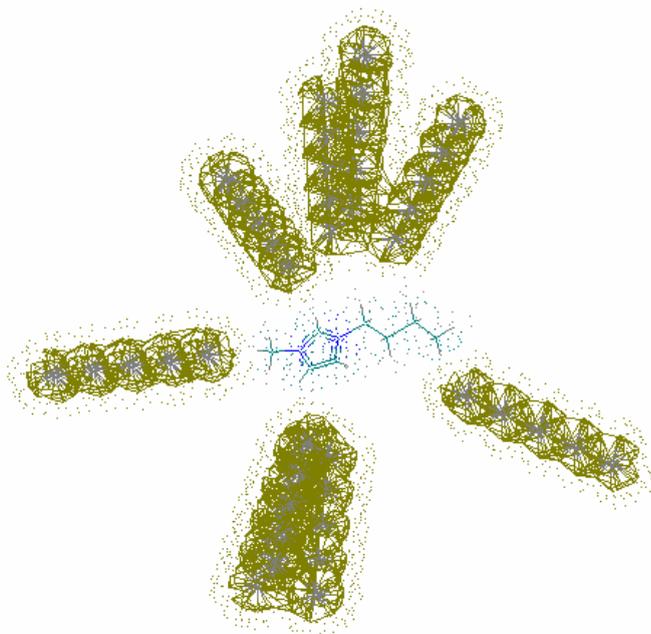


Figura 3.90: Diversas configurações superpostas para o par cátion BMI (centro – conformação 1), e para o ânion BF_4^- , ocupando várias posições ao redor do cátion.

Com base na análise de sensibilidade e análise de componentes principais escolheu-se as entradas (distâncias quadradas) para as RNAs: R1-14, R1-15, R4-11, R5-15 (Figura 3.91).

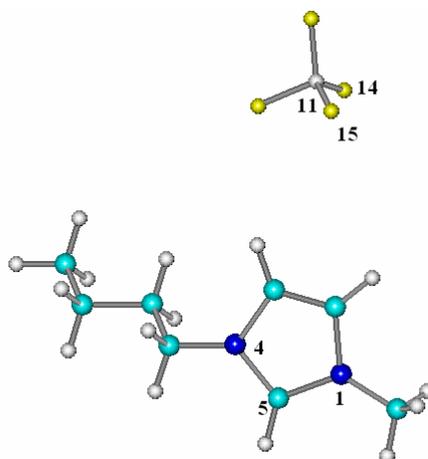


Figura 3.91: Par $[\text{BMI}][\text{BF}_4^-]$ com numeração dos átomos, cujas distâncias quadradas foram utilizadas como entrada para o modelo baseado em RNAs.

Como o intervalo de energia obtido foi pequeno, não se utilizou a função logarítmica e nem se adotou a constante C , necessária no item 3.5.4. Para o treinamento utilizou-se 308 dados e 29 para testar o desempenho das RNAs.

Os resultados a seguir foram obtidos para a primeira conformação (Tabela 3.34). As superfícies de energia potencial (SEP), para o conjunto de treinamento, a partir de energias *ab initio* e a partir de RNAs geradas são apresentadas nas Figuras 3.92 e 3.93.

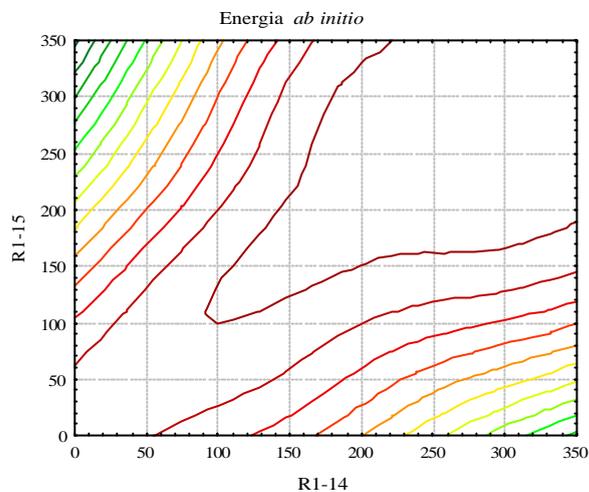


Figura 3.92: Superfície de energia potencial a partir de energias *ab initio*.

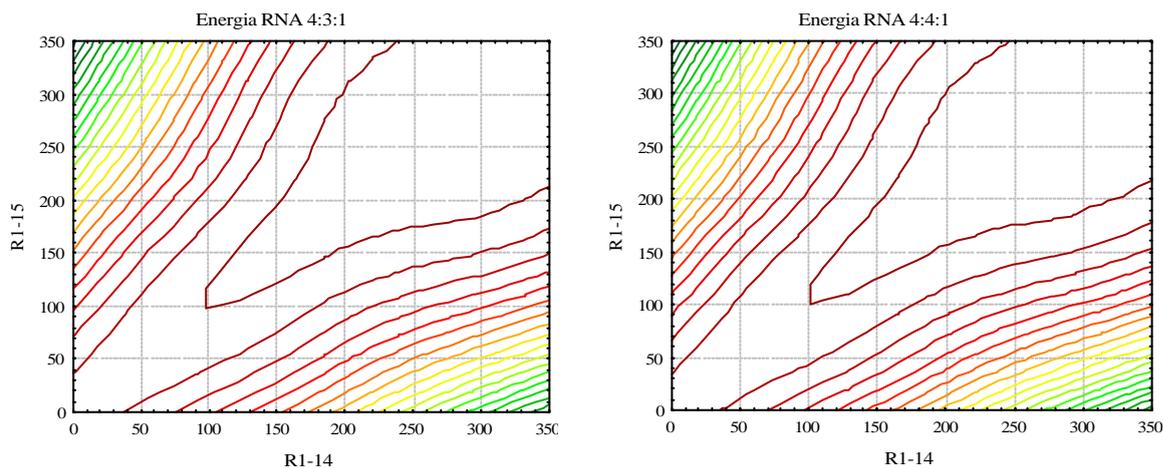


Figura 3.93: Superfície de energia potencial a partir de energias previstas pela RNA 4:3:1 (esquerda) e pela RNA 4:4:1 (direita).

O comportamento das SEPs foi capturado pelas RNAs. Entretanto, para verificar o desempenho das RNAs utilizou-se os 29 dados separados preliminarmente. A Figura 3.94 mostra as energias previstas.

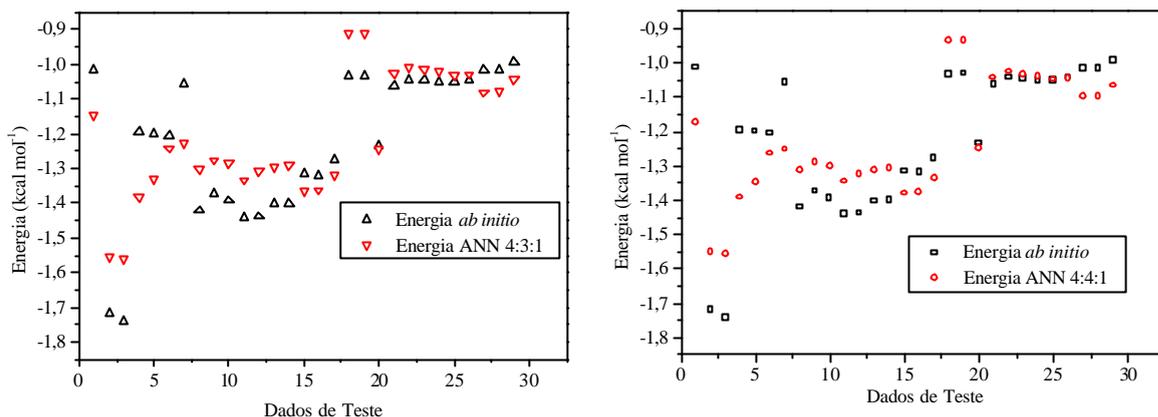


Figura 3.94: Energias *ab initio* e previstas pela RNA 4:3:1 (esquerda) e pela RNA 4:4:1 (direita).

As superfícies de energia potencial (Figura 3.96), entretanto, não representam fielmente a SEP obtida usando as energias *ab initio* (Figura 3.95). Pode ter ocorrido um treinamento no qual as RNAs perderam sua capacidade de generalização, tornando-se específicas para os dados que compõem o conjunto de treinamento. Nesta etapa seria necessário treinar novamente as redes aumentando o conjunto de treinamento ou parando o processo antes que ocorra o sobre-treinamento.

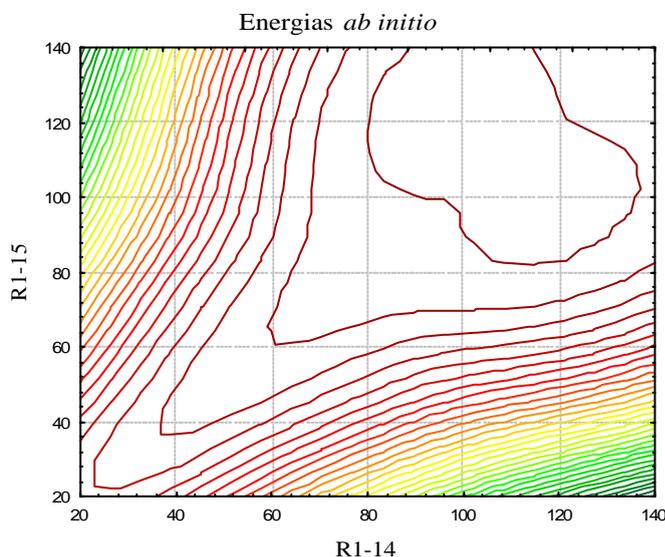


Figura 3.95: Superfície de energia potencial para o par $[\text{BMI}][\text{BF}_4^-]$ usando energias *ab initio* (conjunto de teste).

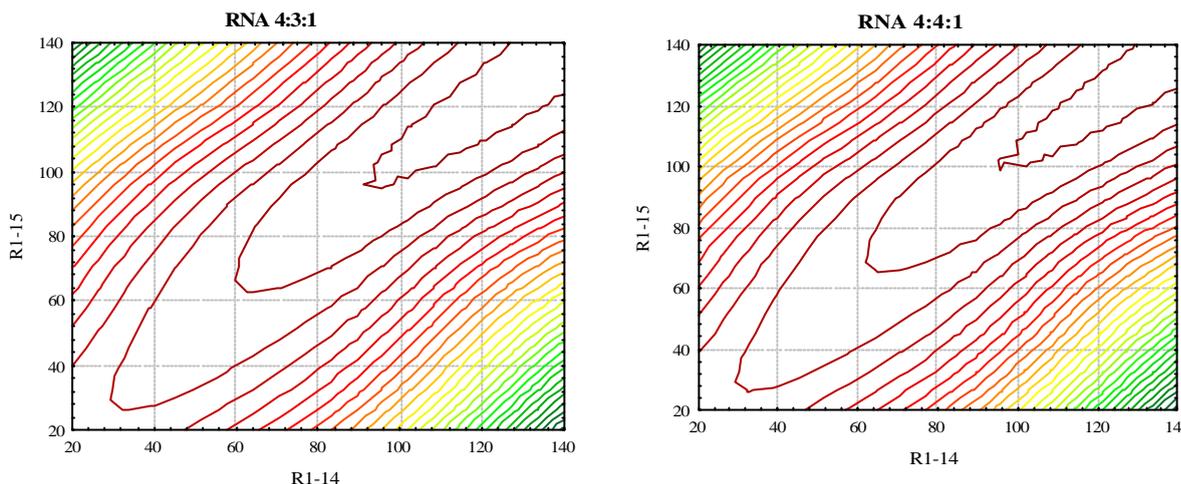


Figura 3.96: Superfícies de energia potencial para o par [BMI][BF₄⁻], obtidas usando os modelos de redes neurais.

Os tempos para execução da subrotina responsável pelo cálculo do potencial de interação através do modelo analítico e através do modelo de RNA com estrutura 4:4:1 estão apresentados na Tabela 3.37. Não é necessário calcular novamente o tempo de simulação para a RNA treinada, já que apresenta a mesma complexidade da RNA usada para simulação da água líquida, apenas mudando os valores dos parâmetros da rede neural.

Tabela 3.37: Demanda computacional para os modelos analítico e aquele baseado em RNAs.

	Par [BMI][BF ₄ ⁻]	Estimativas de uma simulação de Monte Carlo ^{a)}	
		Tempo total (h)	Razão
RNA (4:4:1)	11 μs	~17,6	1
Modelo analítico	143 μs	~228	12,4

a) 10 milhões de configurações com mudança de volume a cada 2000 configurações.

CAPÍTULO 4

CONCLUSÕES E PERSPECTIVAS

4.1. Simulação Computacional da água em diferentes condições de temperatura e pressão

Trabalhos como os de Jedlovsky & Vallauri (1996)¹²⁸ contribuíram significativamente para a correção dos dados experimentais publicados por Soper & Phillips em 1986¹²⁹ e novamente Soper *et. al.* em 1993¹³⁰ que após críticas reexaminaram seus dados e publicaram novos resultados em 1997¹³¹. Isto ressalta a utilidade da simulação computacional, em particular da água, e responde dúvidas e questionamentos, como o de Brodsky em 1996, sobre o valor preditivo da simulação computacional da água¹³². Também outros trabalhos concretizam o fato de que simulações computacionais, quando executadas e analisadas de forma criteriosa podem gerar informações úteis, daí a motivação para o desenvolvimento e aplicação deste tipo de ferramenta.

Os resultados da simulação computacional para os dois modelos (TIP4P e TIP5P) utilizados neste trabalho apresentaram-se razoáveis em comparação com os resultados experimentais. Particularmente o modelo TIP4P gerou melhores resultados para as propriedades termodinâmicas em condições extremas, como a densidade. Contudo, o modelo TIP5P forneceu resultados mais adequados para as funções de distribuição radial para todas as condições. Como as RDFs constituem um forte indicativo para escolha de um modelo este foi adotado para a geração de dados para o treinamento de RNAs.

Outro ponto a ser abordado é a diferença nos protocolos de simulação entre as condições analisadas, bem como o difícil ajuste dos percentuais de aceitação de novas configurações no método de Monte Carlo para as condições extremas de temperatura e pressão.

A análise da presença de interações de ligações de hidrogênio foi determinada a partir dos cálculos utilizando o modelo TIP5P. Verificou-se que em condições extremas a média de ligações de hidrogênio diminui significativamente de quatro, nas condições ambiente, para no máximo três, em uma condição subcrítica (100 bar e 150°C), duas, na maior parte

¹²⁸ P. Jedlovsky, R. Vallauri, *J. Chem. Phys.* 105, 2391-2398 (1996).

¹²⁹ K. Soper, M. G. Phillips, *Chem. Phys.* 107, 47-60 (1986).

¹³⁰ A. K. Soper, C. Andreani, M. Nardone, *Phys. Rev. E* 47, 2598-2605 (1993).

¹³¹ A. K. Soper, F. Bruni, M. A. Ricci, *J. Chem. Phys.* 106, 247-254 (1997).

¹³² A. Brodsky, *Chem. Phys. Lett.* 261, 563-568 (1996).

das condições analisadas, e até mesmo apenas uma ligação de hidrogênio, em média, à 500 bar e 400°C. Verificou-se ainda, que em altas temperaturas e pressões, além de se formar um número menor de ligações de hidrogênio, a maior parte destas ligações são mais fracas (entre 3 e 4 kcal/mol) que aquelas formadas à temperatura ambiente (entre 4 e 5 kcal/mol).

A diminuição das interações por ligações de hidrogênio está em acordo com a formação de agregados verificada em investigações experimentais e também quantitativamente pela análise topológica das redes de ligações de hidrogênio formadas (Item 4.3).

4.2. Simulação Computacional do Metanol e da Mistura Água-Metanol 1:1 nas Condições Ambiente e Extrema

A simulação computacional do metanol mostrou que na condição crítica analisada a altura dos primeiros picos das funções de distribuição radial diminuem, e então ocorre, assim como observado para a água, uma diminuição das interações de ligações de hidrogênio.

Verificou-se que nas condições ambiente a água presente na mistura água-metanol 1:1 apresenta-se mais organizada que na água pura, pois as funções de distribuição radial entre os átomos presentes nas moléculas de água têm o primeiro pico aumentado com relação as mesmas funções para a água pura. Ao passo que para o metanol, verifica-se que o primeiro pico diminui, de modo que se obtém moléculas de metanol menos agregadas quando na mistura do que quando no metanol líquido puro.

Em condições extremas observou-se que tanto as moléculas de água como as de metanol encontram-se mais agregadas entre si. A partir da análise das funções de distribuição radial para a mistura água-metanol em condições extremas, verifica-se que as RDFs entre os átomos de oxigênio do metanol e os átomos de oxigênio da água apresentaram picos menores nestas condições. Disto conclui-se que em condições extremas ocorre uma possível tendência a separação de fases.

Da análise das interações de ligações de hidrogênio, verificou-se que em condições extremas a média de ligações de hidrogênio diminui significativamente tanto no metanol como na mistura, assim como observado no caso da água líquida. A análise topológica destes sistemas foi realizada e as conclusões estão resumidas no item 4.3 a seguir.

4.3. Análise topológica da mistura água:metanol 1:1 e dos líquidos puros

A análise das propriedades topológicas locais das redes de ligações de hidrogênio na água pura supercrítica indicam a possibilidade da presença de padrões de *small-world*, ao passo que estas redes nas condições ambiente não apresentam padrões de *small-world*, apesar dos altos coeficientes de *clustering* ou de agregação e percoção no sistema. Este comportamento diferente também foi verificado pela análise da densidade espectral que fornece a topologia global. Pode-se ainda concluir que para a água, as propriedades locais são independentes do número de moléculas ou sítios para caixas de simulação maiores que 300. A Tabela 4.1 apresenta as propriedades topológicas locais para os três sistemas estudados (água pura, metanol puro e mistura água:metanol 1:1), todos para caixas de simulação contendo 512 sítios ou moléculas.

Ressaltamos que esta foi a primeira vez que a análise da presença de padrões de *small-world* foi realizada para sistemas químicos.

Tabela 4.1: Comprimento do caminho médio (L and L_{rand}), coeficiente de agregação (C and C_{rand}) e número de conexões médio ($\langle k \rangle$) para as redes de ligações de hidrogênio presentes na mistura água:metanol e naquelas formadas no líquidos puros. AMB: $p = 1$ bar e $T = 25^\circ\text{C}$, CRIT: $p = 500$ bar e $T = 400^\circ\text{C}$.

	Água		Metanol		Água: Metanol 1:1	
	AMB	CRIT	AMB	CRIT	AMB	CRIT
L	20,36	3,33	5,19	2,33	27,59	2,74
L_{rand}	5,23	23,78	11,76	9,99	6,81	59,2
L/L_{rand}	3,9	0,14	0,44	0,23	2,5330	0,0463
C	0,5456	0,2114	1,6595	0,5355	1,1425	0,3995
$C_{rand} = p$	0,0064	0,0025	0,0032	0,0010	0,0049	0,0017
C/C_{rand}	85,2	84,5	518,6	535,6	233,2	235,0
$\langle k \rangle$	3,3	1,3	1,65	0,536	2,5	0,9

Os altos coeficientes de agregação e pequenos comprimentos de caminhos médios (Tabela 4.1) obtidos para o metanol puro indicam uma estrutura de pequenos agregados neste líquido puro, tanto em condições ambiente quanto em condições críticas, mas para concluir sobre a presença ou não de padrões de *small-world* (SW), necessita-se que o número de sítios para este sistema seja aumentado, para que agregados maiores sejam

usados numa estatística mais confiável do cálculo do comprimento do caminho médio destas redes. A densidade espectral para o metanol puro também indica a presença de estruturas em ambas as condições, apesar deste sistema corresponder a um líquido, distanciando-se claramente de um comportamento randômico.

Para a mistura podemos observar resultados análogos aqueles obtidos para a água líquida pura. Observando a Tabela 4.1 confirmamos esta conclusão, como também a análise das densidades espectrais. Portanto, tanto a água pura quanto a mistura binária apresentam percolação no sistema em condições ambiente e apresentam padrão de *small-world* na condição supercrítica 500 bar 400°C.

Perspectivas

Realizar simulações de Monte Carlo com 1024 moléculas de metanol para simular o líquido puro em condições ambiente e crítica, seguida da análise topológica para confirmar as conclusões e padrões obtidos com as simulações com 512 moléculas.

Reparametrizar o modelo TIP5P utilizando dados experimentais relativos a condições extremas de temperatura e pressão.

Simular fluidos supercríticos incluindo solutos e verificar os efeitos através da análise topológica destes sistemas.

Relacionar quantitativamente a estrutura com separação de fases.

Explicar a presença de padrões de *SW* em sistemas tão distintos, como redes de computadores, de alimentos, de circuitos eletrônicos e redes de ligações de hidrogênio.

4.4. Simulação Computacional de Líquidos Iônicos

Resultados da análise conformacional do cátion 1-butil-3-metilimidazólio (BMI) permitiu a escolha de três confôrmeros. Como a molécula pode assumir estas diferentes conformações no líquido, a inclusão deste efeito nas simulações pode explicar a razão pela qual estes sais são líquidos nas condições ambiente. As caixas obtidas em condições ambientes apresentam-se mais estruturadas quando construídas considerando-se apenas o confôrmero 1, o qual apresenta a cadeia lateral linear, ao passo que aquelas obtidas com os confôrmeros com cadeia lateral não linear apresentam maior grau de desordem.

Perspectivas

Modificar o programa DIADORIM para realização de simulações de Monte Carlo considerando os três confôrmeros como solvente 1, solvente 2 e solvente 3 e a espécie X (ânion) como solvente 4.

Determinar os efeitos das proporções destes confôrmeros nas propriedades termodinâmicas e estruturais dos líquidos iônicos.

Modificar o programa DIADORIM para que permita a interconversão destes confôrmeros como um passo de Monte Carlo.

4.5. Ajuste de Superfícies de Energia Potencial usando Redes Neurais

Para o dímero da água, No *et al.*¹³³ treinaram redes neurais perceptron multicamadas a partir de dados *ab initio*, entretanto o tempo para execução das RNAs não viabilizam a sua aplicação em simulação computacional. Neste trabalho, conseguiu-se obter uma RNA com desempenho favorável do ponto vista de demanda computacional e previsão da energia para obtenção de superfícies de energia potencial.

Os resultados indicam que devem ser utilizadas apenas RNAs do tipo perceptron multicamadas (MLP) e a função de ativação sigmoidal.

No caso dos líquidos iônicos ainda não existiam modelos baseados em RNAs desenvolvidos. Para este caso, a redução do tempo computacional foi mais significativo dado o maior número de sítios de interação para uma descrição da interação via potencial analítico.

Cálculos *ab initio* da energia de interação para o par BMI^+ e BF_4^- , em configurações geradas sistematicamente, forneceram os dados para o treinamento de RNAs, os quais foram divididos de modo a permitirem o teste dos modelos na obtenção de superfícies de energia potencial intermolecular.

Perspectivas

Gerar conjuntos de treinamento mais representativos com relação a energia de interação.

¹³³ K. T. No, B. H. Chang, S. Y. Kim, M. S. Jhon, H. A. Scheraga, *Chem. Phys. Lett.* 271, 152-156 (1997).

4.6. Modelos baseados em Redes Neurais Artificiais em Simulações de Monte Carlo

A utilização de RNAs em simulações de Monte Carlo para o cálculo do potencial de interação de pares constituiu um dos aspectos originais deste trabalho.

O tempo necessário a realização de simulações de Monte Carlo da água com RNAs foi em média cerca de três vezes menor do que a simulação usando o modelo analítico TIP5P, apesar do logaritmo acrescentado, já que a estrutura da rede permaneceu de baixa complexidade, entretanto apesar das várias RNAs treinadas, os erros são ainda grandes para fornecer resultados adequados quando são utilizadas em simulações de Monte Carlo. O tempo previsto para uma simulação de Monte Carlo do líquido iônico $[BMI^+][BF_4^-]$ análoga aquela usada para o cálculo do tempo no caso da água, indica um redução de aproximadamente 12 vezes no tempo que levaria a simulação usando o modelo analítico.

Perspectivas

Embora as simulações de Monte Carlo para a água não tenham levado a bons resultados quanto as propriedades termodinâmicas e energia média (item 4.6), as RNAs obtidas a partir das energias de interação entre os íons presentes em líquidos iônicos, devem ser retreinadas a partir de dados mais completos, com melhor amostragem, e em seguida, testados em simulações de Monte Carlo.

Realizar uma amostragem mais aprimorada dos dados de treinamento.

Ampliar os valores de C a serem testados.

Testar a divisão dos dados em duas regiões, uma compreendendo energias baixas e a outra contendo as energias positivas maiores, de modo que RNAs de complexidades semelhantes seriam usadas durante a simulação, entretanto com valores dos pesos e *biases* distintos.

Modificar a função objetivo de treinamento, do quadrado da diferença entre o valor predito e o valor desejado, $E_p = 1/2 \sum_i (S_i - t_i)^2$, para uma função correspondente ao quadrado do erro relativo, $E_p = 1/2 \sum_i [(S_i - t_i)/t_i]^2$.

APÊNDICE 1

TRABALHOS APRESENTADOS EM CONGRESSOS

1) XI Simpósio Brasileiro de Química Teórica – XI SBQT. Caxambú 2001.

MOLECULAR SIMULATION OF LIQUID WATER USING THE TIP4P AND TIP5P MODELS AT ROOM AND NEAR CRITICAL CONDITIONS.

Vivianni Marques Leite dos Santos, Marcelo Zaldini Hernandes, Ricardo Luiz Longo.

2) XII Simpósio Brasileiro de Química Teórica – XII SBQT. Caxambú 2003.

OBTENÇÃO DE SUPERFÍCIES DE ENERGIA POTENCIAL USANDO REDES NEURAIAS ARTIFICIAIS

Vivianni Marques Leite dos Santos, Ricardo Luiz Longo.

3) Workshop sobre Simulação e Modelagem de Nanoestruturas. Brasília. 2003.

MONTE CARLO SIMULATION OF SUPERCRITICAL WATER WITH ANALYSIS OF HYDROGEN BONDS AND STUDY OF THE SMALL-WORLD PATTERNS.

Vivianni Marques Leite dos Santos, Ricardo Luiz Longo.

4) 55^a Reunião Anual da SBPC. Recife. 2003.

SIMULAÇÃO DE MONTE CARLO DO SISTEMA ÁGUA-METANOL EM CONDIÇÕES AMBIENTE E SUPERCRÍTICA.

Vivianni Marques Leite dos Santos, Ricardo Luiz Longo.

5) Participação Workshop: "Química Quântica: A solução ideal para os Problemas Reais. Instituto de Química da Universidade de São Paulo (IQ-USP).

APÊNDICE 2

Topology of the Hydrogen Bond Network in Liquid Water
at Room and Supercritical Conditions: A Small-World
Structure. *Chem. Phys. Lett.* 390, 157–161 (2004).

Topology of the hydrogen bond networks in liquid water at room and supercritical conditions: a small-world structure

Vivianni Marques Leite dos Santos ^a, F.G. Brady Moreira ^b, Ricardo L. Longo ^{a,*}

^a Departamento de Química Fundamental, Universidade Federal de Pernambuco, Av. Prof. Luiz Freire, sn, Cidade Universitária, 50740-540 Recife, PE, Brazil

^b Departamento de Física, Universidade Federal de Pernambuco, 50670-901, Cidade Universitária, Recife, PE, Brazil

Received 23 January 2004; in final form 1 April 2004

Available online 27 April 2004

Abstract

Monte Carlo NPT simulations were performed for water at room and supercritical thermodynamic conditions using the TIP5P model potential. The topology of the hydrogen bond patterns was characterized by local (clustering coefficient, path length and degree distribution) and global (spectral analysis) properties. The analysis was performed on 500 uncorrelated configurations at room and supercritical (500 bar and 673 K) conditions. Small-world behavior (highly clustered and small path lengths) was characterized for supercritical water, whereas for room condition, a giant cluster appears and percolates the whole system. These results might have important consequences upon using (near-) supercritical water as new reaction media.

© 2004 Elsevier B.V. All rights reserved.

1. Introduction

Water is the Green Chemistry solvent since it is readily available at low cost, non-toxic, non-inflammable, non-explosive, non-volatile, recyclable and easy to handle. In addition, the physical–chemical properties of liquid water, such as, density, viscosity, dielectric constant, auto-ionization (K_w), ionic and heat conductivity, can be drastically changed or finely tuned by taking it to supercritical or near the supercritical conditions [1]. As a result, water at (near-)supercritical condition has become an important reaction medium where several chemical reactions and synthesis can be performed [2], even without the need for acid–basis catalysts. It has long been established that the physical–chemical properties of water are primarily due to the hydrogen bond (H-bond) network connecting all the water molecules. The behavior of this H-bond network as a function of high temperature and pressure has only recently being probed (revised) by experiments, particularly by

neutron diffraction with isotope substitution [3], which yielded structural information concerning the atom-atom (O–O, O–H and H–H) correlations in supercritical water (673 K and 500 bar vs. $T_c = 647$ K, $p_c = 221$ bar and $\rho_c = 0.32$ g/cm³). However, these are averaged results and a detailed picture of the H-bond network differences for water at room and supercritical conditions is still lacking.

In this Letter, we have used computer simulation (Monte Carlo) to generate liquid structures at room and supercritical conditions from which the topology of the H-bond networks were analyzed in order to establish or not the presence of small-world patterns [4]. This class of disordered networks is widespread in a variety of complex networks [5], such as, the internet topology, neural nets, electronic circuits, food webs, ecosystems, social interactions, scientific collaboration, etc. Due to its apparent universality [5,6], small-world patterns might yield information about the properties of complex networks, such as, evolvability, fitness and robustness with respect to external fluctuations. In the present case it could be, for instance, the stability of the H-bond networks with respect to pressure and temperature

* Corresponding author. Fax: +55-81-2126-8442.

E-mail address: longo@ufpe.br (R.L. Longo).

changes, which might yield new information and perspectives about the relationship between the water properties and the H-bond network.

2. Methodology

The liquid structures were obtained by standard Metropolis Monte Carlo computer simulations [7] using the TIP5P model potential [8] for water-water interaction. The simulations were performed with the DIAD-ORIM program [9] using a constant number of molecules ($N = 300$ and 512), at temperature ($T = 298$ and 673 K) and pressure ($p = 1$ and 500 bar). A cubic simulation box was employed with an initial volume consistent with the experimental density. Periodical boundary conditions, cutoff distance of 900 pm, and a volume variation at every 2000 configurations were used during the simulations. In addition, the range of the molecular translations (maximum translation motion of 15.5 and 57 pm at room and supercritical conditions, respectively), and rotations (maximum rotation motion 15.5° and 57° at room and supercritical conditions, respectively), were adjusted to yield an acceptance ratio of approximately 40% . For equilibration, at least 20×10^6 configurations were generated from simulation boxes already equilibrated in conditions near to the target state. The equilibrium condition has been thoroughly tested using energy, density, radial distribution functions, amongst other properties. Then, the liquid structures used for the H-bond analysis were collected every 3000 equilibrium configurations until it performed 500 simulation boxes. The interval of 3000 configurations (≈ 6 MC cycles, for the large system simulated) has been chosen to guarantee that the liquid structures analyzed were statistically uncorrelated.

The following geometrical restrictions were used in order to establish the presence of an H-bond: $R_{O-O} \leq 330$ pm, for the distance between two oxygen atoms of the water molecules, and $R_{H-X} \leq 274$ pm, for the distance between the hydrogen atom of a donor water molecule and one of the dummy or pseudo-atoms ($X = LP$) that represents the lone-pair (LP) of the acceptor water molecule, in the TIP5P model. These criteria are consistent with those previously proposed [10] using distance and angles. Once the geometrical condition has been satisfied, an additional criterion for the interaction energy of the pair of water molecules is employed, namely, the interaction energy has to be negative (attractive). Finally, these H-bonds are translated into a connectivity matrix that is used in the topological analysis. The topological properties [5] of the H-bond networks were determined by calculating the clustering coefficient C , the path length L , the degree distribution $P(k)$ and the graph spectra $\rho(\lambda)$.

The clustering coefficient, C_i , at node or vertex (molecule) i in the network is defined as the number of existing edges or links (H-bonds) among the k_i neighbors of node i , E_i , divided by the number of possible connections between them, $\binom{k_i}{2}$, that is,

$$C_i = E_i / \binom{k_i}{2} = \frac{2E_i}{k_i(k_i - 1)}. \quad (1)$$

The clustering coefficient of the whole network is the average of all individual C_i 's, namely, $C = \langle C_i \rangle$. It can be formalized in terms of graph theory [11] by considering a set of links $\xi_{ij}(i, j = 1, 2, \dots, N)$, where $\xi_{ij} = 1$ if the nodes i and j are connected, and $\xi_{ij} = 0$ otherwise. Let $\Gamma_i = \{j | \xi_{ij} = 1\}$ be the set of nearest neighbors of node i , with k_i members, then the number of connections between them is

$$E_i = \sum_{j=1}^N \xi_{ij} \left[\sum_{k>j, k \in \Gamma_i}^{k_i} \xi_{jk} \right], \quad (2)$$

and the clustering coefficient at node i can be calculated by Eq. (1). For a random graph, that is, a network with N nodes randomly connected with probability p , the clustering coefficient is [5]

$$C_{\text{rand}} = p = \frac{\langle k \rangle}{N}, \quad (3)$$

where $\langle k \rangle$ is the average degree of the network. It should be emphasized that the nearest neighbors are related to the surroundings of a given node (molecule) in a network (of H-bonds), which does not imply into spatial closeness nor that spatial closeness implies into edges or links (H-bonds). As a result, the average of nearest spatial neighbors of a given node (molecule) should be slightly larger than the average number of connections.

Let $l_{\min}(i, j)$ be the minimum path length (number of edges or links) connecting nodes i and j in the same graph. The average path length of a given node is [11]

$$l(i) = \frac{1}{N} \sum_{j=1}^N l_{\min}(i, j) \quad (4)$$

and the path length is the average over the nodes, namely, $L = \langle l(i) \rangle$. For a random graph the path length scales as [5]

$$L_{\text{rand}} \approx \frac{\ln N}{\ln \langle k \rangle}. \quad (5)$$

A small-world pattern is characterized by $C \gg C_{\text{rand}}$ and $L < L_{\text{rand}}$ [4,5].

Since not all nodes in a network have the same number of edges, a degree distribution $P(k)$ can be defined as the (normalized) frequency of nodes having k edges or node degree. In a random graph the edges are

placed randomly, so the majority of nodes have approximately the same degree, close to the average degree $\langle k \rangle$ of the network. More specifically [5], for a random graph, $P(k)$ is a Poisson distribution peaked at $\langle k \rangle$. Also, for a small-world network described by the Watts–Strogatz model [4], the shape of the degree distribution is similar to that of a random graph [5]. However, for some complex large networks with small-world behavior, the degree distribution significantly deviates from a Poisson distribution, and are usually described as [6]: (i) scale-free networks, with power-law distribution, $P(k) \approx k^{-\gamma}$; (ii) broad-scale networks, $P(k) \approx k^{-\gamma} f(k/k^*)$ where k^* is the cutoff for the sharp cutoff function of the power-law distribution; (iii) single-scale distribution (exponential or Gaussian).

An adjacency matrix $A(G)$ with $N \times N$ elements A_{ij} can represent a graph G with N nodes, when the elements satisfy the following conditions: $A_{ij} = A_{ji} = 1$ if nodes i and j are connected, and 0 otherwise [12]. The diagonalization of the matrix $A(G)$ yields a set of N eigenvalues, λ_j , that is the spectrum of graph G and originates the following spectral density [12]:

$$\rho(\lambda) = \frac{1}{N} \sum_{j=1}^N \delta(\lambda - \lambda_j), \quad (6)$$

which approaches a continuous function as $N \rightarrow \infty$. The k -th moment of the spectral density can be written as the number of paths returning to the same node in the graph, thus yielding some topological features of the network. For a random graph with N nodes and edges generated randomly with probability p , the spectral density converges to a semicircular distribution expressed as [12],

$$\rho(\lambda) = \begin{cases} \frac{\sqrt{4Np(1-p) - \lambda^2}}{2\pi Np(1-p)} & \text{if } |\lambda| < 2\sqrt{Np(1-p)}, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The largest (principal) eigenvalue is isolated from the bulk of the spectrum and it decreases as pN .

In this work a FORTRAN program was coded, which reads the coordinates of the water molecules from the simulation boxes, determines the H-bond connections and then perform the topological analysis on the H-bond networks.

3. Results and discussions

Regarding the Monte Carlo simulations the calculated thermodynamic properties agree well with previous simulation results [8], including their dependence with the temperature and pressure. In addition, the calculated radial distribution functions are in good agreement with the experimental data even for supercritical conditions [3]. Table 1 presents the results for the

Table 1
Path length (L and L_{rand}), clustering coefficient (C and C_{rand}) and average number of connections ($\langle k \rangle$) for H-bond networks of water

	298 K and 1 bar		673 K and 500 bar	
	$N = 512$	$N = 300$	$N = 512$	$N = 300$
L	20.36	19.65	3.33	3.30
L_{rand}	5.23	4.90	23.78	31.28
L/L_{rand}	3.9	4.0	0.14	0.11
C	0.5456	0.5302	0.2114	0.2057
$C_{\text{rand}} = p$	0.0064	0.0106	0.0025	0.0041
C/C_{rand}	85.2	50.0	84.5	50.2
$\langle k \rangle$	3.3	3.2	1.3	1.2

clustering coefficient, path length and average number of connections, for 500 H-bond network structures at room and supercritical conditions, obtained from simulations with 300 and 512 water molecules.

For a given thermodynamic condition, it can be observed that the parameters (L and C) determining the local behavior of the H-bond networks, as well as their average degree, are not dependent upon the number of water molecules in the simulation box. This indicates that the simulated samples of 300 and 512 molecules, with periodic boundary conditions in all directions, are large enough to ensure the thermodynamic limit. A comparison between the parameters calculated for different conditions, shows that, at room temperature, the liquid water H-bond networks have $L \gg L_{\text{rand}}$ and $C \gg C_{\text{rand}}$, which suggests that the path length of the network increases with its clustering, characteristic of a giant cluster that percolates the whole system [5]. On the contrary, at the supercritical condition, these networks have much smaller path lengths ($L_{\text{rand}} \gg L$), however the clustering is still large ($C \gg C_{\text{rand}}$), thus yielding the small-world behavior [4] of the H-bond networks at this thermodynamic condition. The average degrees presented in Table 1 agree with the average number of hydrogen bonds, $\langle n_{\text{HB}} \rangle$, of earlier simulations [13]. In particular, it corroborates that at the supercritical condition the average degree ($\langle k \rangle \approx 1.2$ – 1.3) is below the percolation threshold, ~ 1.56 [14], so that a continuous network of H-bonds does not exist, whereas for the room temperature, the percolation persists as can be observed in Fig. 1. For comparison, Fig. 2 illustrates the same H-bond network (upper panel) and its largest cluster (lower panel) at the supercritical condition.

The distribution degree for the H-bond networks of water at room and supercritical conditions are presented in Figs. 3 and 4, respectively. It can be noted that at room temperature it is very similar to a Poisson distribution compatible with a random graph behavior [5]. However, for the supercritical water the distribution is skewed and quite distinct from a Poisson-like behavior. Indeed, it appears similar to broad-scale networks, where the power-law degree has as sharp cutoff [6]. The

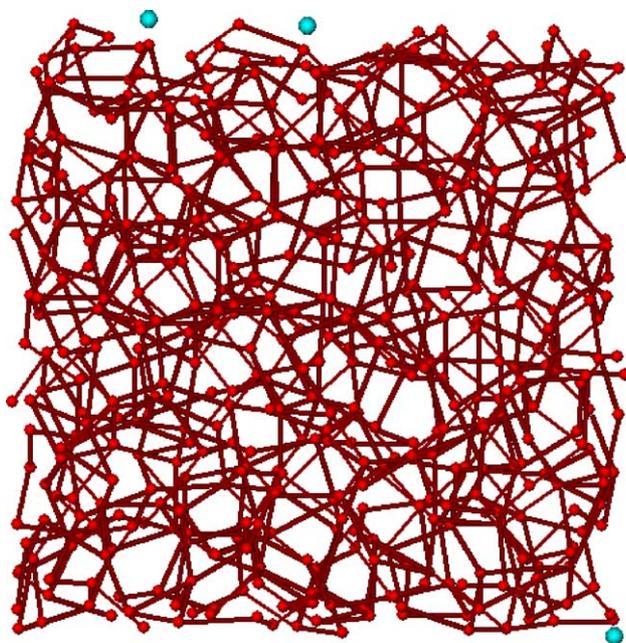


Fig. 1. H-bond network displayed by liquid water at room temperature ($N = 512$). Each node represents a water molecule and the edges the hydrogen bonds.

shape of the distribution has been related, for instance, to constraints limiting the number of links when the connections are costly. This conjecture [6] is consistent to the results observed for H-bond networks of supercritical water, since the rewiring is expensive due to the very low density and long average intermolecular distances, thus relating the shape of the distribution degree to the thermodynamic properties such as density and vaporization enthalpy. It should also be noted that distribution degree is also independent of the number of nodes (300 or 512 molecules).

An analysis using cluster number technique has been performed, but will not be presented and discussed since a similar study is already available [15] for the supercritical water. However, it should be noted that this cluster analysis is quantitatively represented in the averaged local topological properties, such as, clustering coefficients, path lengths and degree distributions.

A global property of complex network is the spectral analysis. The rescaled spectral density of all 500 networks of H-bond for liquid and supercritical water is presented in Fig. 5.

Clearly, the spectral density at room temperature is dominated by the first moment, as expected from the observed percolation process, that is, the formation of giant clusters. On the other hand, for supercritical water, the spectral density has significant contributions from the second and third moments, which is consistent with the small-world behavior of these H-bond networks. Also, it should be noted that when the number of nodes (molecules) increases from 300 to 512 the principal

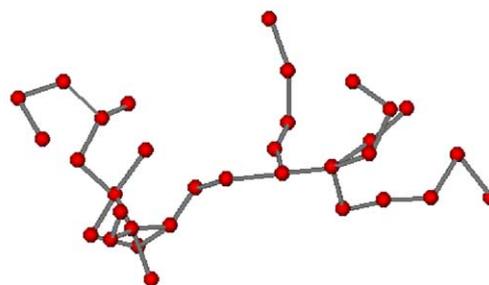
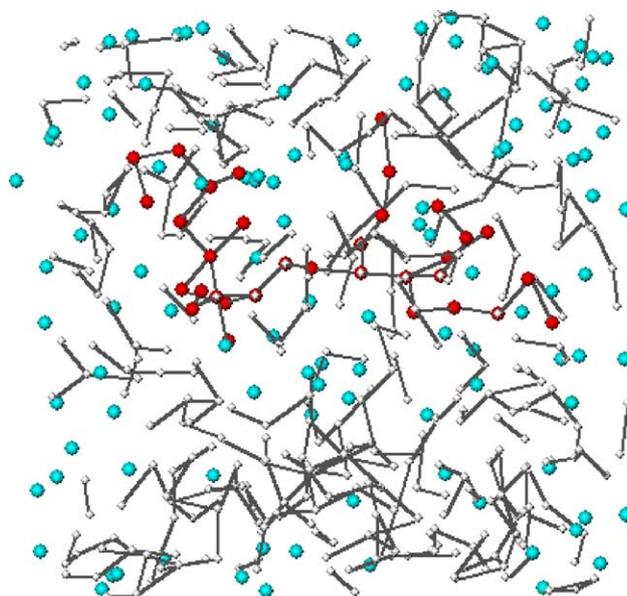


Fig. 2. H-bond network (top panel) and its largest cluster (lower panel) displayed by supercritical water ($N = 512$). Each node represents a water molecule and the edges the hydrogen bonds.

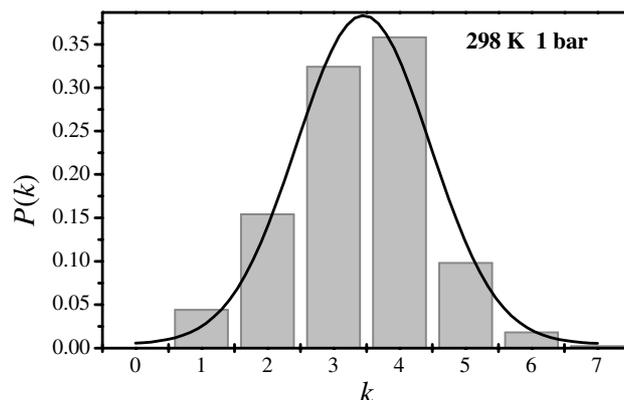


Fig. 3. Degree distribution for the H-bond network of water at room condition.

moment decreases significantly and the spectral density becomes more similar to that for a random graph (semicircular law).

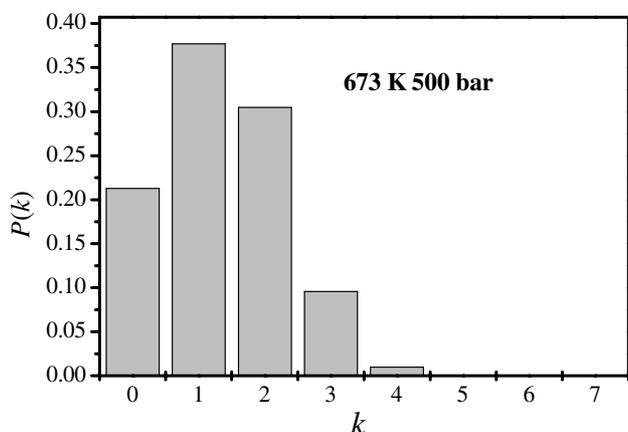


Fig. 4. Degree distribution for the H-bond network of supercritical water.

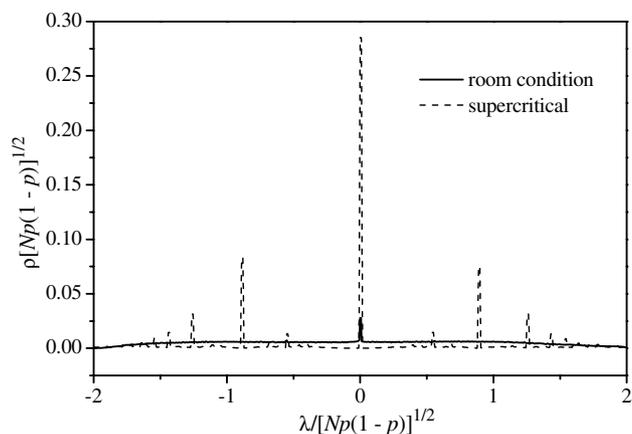


Fig. 5. Rescaled average spectral density for the H-bond networks of water ($N = 512$).

4. Conclusions

The analysis of local topological properties suggests that the H-bond networks in supercritical water present small-world pattern, whereas these networks at room

condition, despite of having large clustering coefficients, do not present small-world behavior, and percolate through the whole system. This distinct behavior has been also clearly observed in the spectral density global property of the connectivity graph. In addition, the local properties (clustering coefficients, path lengths and degree distributions) are independent upon the number of nodes (molecules) for simulation boxes larger than 300 molecules.

Acknowledgements

The authors acknowledge CAPES, CNPq, FINEP, IMMC, PADCT, PRONEX and RENAMI (Brazilian agencies) for financial support. One of the authors (V.M.L.S.) wishes to thank CNPq for providing a graduate fellowship.

References

- [1] P. Krammer, H. Vogel, *J. Supercrit. Fluids* 16 (2000) 189.
- [2] P.E. Savage, *Chem. Rev.* 99 (1999) 603; A.R. Katritzky, D.A. Nichols, M. Siskin, R. Murugan, M. Balasubramanian, *Chem. Rev.* 101 (2001) 837.
- [3] A.K. Soper, F. Bruni, M.A. Ricci, *J. Chem. Phys.* 106 (1997) 247.
- [4] D.J. Watts, S.H. Strogatz, *Nature* 393 (1998) 440.
- [5] R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* 74 (2002) 47.
- [6] L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, *Proc. Natl. Acad. Sci. USA* 97 (2000) 11149.
- [7] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1989.
- [8] M.W. Mahoney, W.L. Jorgensen, *J. Chem. Phys.* 112 (2000) 8910.
- [9] L.C.G. Freitas, *DIADORIM* Program, UFSCar, São Carlos, SP, Brazil, 1999.
- [10] A.A. Chialvo, P.T. Cummings, *J. Phys. Chem.* 100 (1996) 1309.
- [11] R.F. Cancho, C. Janssen, R.V. Solé, *Phys. Rev. E* 64 (2001) 046119.
- [12] I.J. Farkas, I. Derényi, A.-L. Barabási, T. Vicsek, *Phys. Rev. E* 64 (2001) 026704.
- [13] A.G. Kalinichev, J.D. Bass, *J. Phys. Chem. A* 101 (1997) 9720.
- [14] R.L. Blumberg, H.E. Stanley, A. Geiger, P. Mousbach, *J. Chem. Phys.* 80 (1984) 5230.
- [15] A.G. Kalinichev, S.V. Churakov, *Chem. Phys. Lett.* 302 (1999) 411.