André Alberto de Sousa Melo

Cálculos Teóricos em Sistemas com Interesse Biológico

Departamento de Química
Faculdade de Ciências da Universidade do Porto
Janeiro / 1999

André Alberto de Sousa Melo

Cálculos Teóricos em Sistemas Com Interesse Biológico

Tese submetida à Faculdade de Ciências da Universidade do Porto para a obtenção do grau de Doutor em Química, especialidade de Química Teórica

Departamento de Química
Faculdade de Ciências da Universidade do Porto
Janeiro / 1999

Agradecimentos

Desejo agradecer em primeiro lugar à minha supervisora, Professora Doutora Maria João Ramos, por toda a orientação prestada que muito contribuiu para a realização deste trabalho.

Aos meus colegas do colegas do grupo de Química Teórica agradeço o apoio que sempre me dispensaram.

Ao Dr. Pedro Rodrigues agradeço a recuperação "milagrosa" da informação contida num disco danificado.

Quero exprimir a minha gratidão ao Departamento de Química da Faculdade de Ciências do Porto por me ter aceite como estudante e por me ter concedido ajuda financeira para várias deslocações a encontros científicos.

À Fundação de Ciência e Tecnologia agradeço também o apoio financeiro na edição deste trabalho.

Finalmente, desejo expressar um agradecimento muito especial à minha família pelo apoio e estímulo que nunca deixou de me prestar.

Para a Luiza, David, Miguel e Ricardo

Da minha aldeia vejo quanto da terra se pode ver do Universo ...
Por isso a minha aldeia é tão grande como outra terra qualquer,
Porque eu sou do tamanho do que vejo.
E não do tamanho da minha altura ...

Fernando Pessoa, extracto de *O Guardador de Rebanhos*, *Poemas de Alberto Caeiro*, Ática, Lisboa (1983).

Resumo

No presente trabalho efectuou-se um estudo teórico de processos de associação entre proteases serinas e os respectivos inibidores. Este estudo desenvolveu-se segundo duas vertentes principais:

Numa primeira vertente, analisou-se detalhadamente a importância relativa das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI, utilizando mutações apropriadas na respectiva região de contacto e métodos de simulação molecular.

Numa segunda vertente, efectuou-se um estudo quântico das interacções arginina-carboxilato que desempenham um papel crucial no mecanismo de inibição do factor β-XIIa.

Foi ainda proposto um novo método, baseado na partição espacial das matrizes densidade e densidade de par reduzidas e no formalismo dos integrais indistinguíveis, para decompor as energias molecular e de estabilização em componentes associadas aos fragmentos que constituem o sistema. Este método é especialmente apropriado para estudar sistemas onde ocorre a transferência dum grupo ligado covalentemente, como é o caso das interacções arginina-carboxilato anteriormente mencionadas.

No âmbito deste trabalho, foi também elaborado um novo algoritmo de compressão de ficheiros de trajectórias obtidos através de dinâmica molecular, designado por Byte Structure Variable Length Coding (BS-VLC), que apresenta eficiências significativamente superiores aos algoritmos conservativos clássicos.

No primeiro capítulo, faz-se uma pequena descrição de aspectos gerais relacionados com os mecanismos de catálise e inibição das proteases serinas. A tripsina, o seu inibidor natural PTI e o factor β-XIIa são objecto de uma atenção especial por serem especificamente estudados neste trabalho.

No segundo capítulo, descrevem-se os métodos de simulação molecular utilizados no estudo das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI, designadamente os métodos da dinâmica molecular, da perturbação termodinâmica e da integração termodinâmica. O algoritmo de compressão BS-VLC, proposto no âmbito deste trabalho, é também apresentado neste capítulo.

No terceiro capítulo, apresentam-se os métodos quânticos utilizados no estudo das interacções arginina-carboxilato. O novo formalismo de decomposição das energias molecular e de estabilização, proposto neste trabalho, é aqui descrito de uma forma detalhada.

No quarto capítulo, são apresentados os resultados que se obtiveram nas simulações moleculares e nos cálculos quânticos efectuados.

No último capítulo, são apresentadas as conclusões gerais do presente estudo e algumas sugestões relativas ao seu desenvolvimento futuro.

Abstract

In the present work, a theoretical study of some association reactions between serine proteases and their inhibitors has been performed. This work has developed according to main lines:

The detailed study of the binding interactions responsible for the stability of the trypsin:PTI complex. This has been done performing relevant mutations in selected amino acids in the respective contact region and using molecular simulation methods.

The quantum study of arginine-carboxylate interactions which undertake a very important role in β-factor XIIa:inhibitor interactions.

A new method, based on the spatial decomposition of the reduced density and pair density matrices and on the indistinguishable integrals formalism, has been proposed also to partition the molecular and stabilization energies into meaningful fragments. This new partition scheme is specially appropriate to study systems in which a directly-bonded group transfer reaction occurs, such as the case of the arginine-carboxylate interactions mentioned earlier.

In this work, a new specific algorithm in to compress the trajectory files obtained by molecular dynamics simulations, named Byte Structure Variable Length Coding (BS-VLC), has been introduced giving significantly better efficiencies than the classical lossless algorithms.

The first chapter includes a small introduction of catalysis and inhibition mechanisms of serine proteases. Here, trypsin, PTI and β-factor XIIa are treated with some detail because these proteins are under study in this work.

In the second chapter, the molecular simulation methods, used in the study of the binding interactions responsible for the stability of the trypsin:PTI complex, are described. Special emphasis is given to molecular dynamics, thermodynamic perturbation and thermodynamic integration formalisms. Here, the BS-VLC compression algorithm is also introduced.

The quantum methods, namely *ab initio* and *semi-empirical* methods, used in the study of the arginine-carboxylate interactions are the subject of chapter 3. The partition scheme, introduced in this work, is also discussed in detail.

In the fourth chapter, the results obtained in the molecular simulations and quantum calculations are presented.

Finally, the last chapter gives the general conclusions of this work, as well as some suggestions for future studies in this field.

Résumé

Dans cet ouvrage un étude théorique des mécanismes d'association entre proteases serines et leurs inhibiteurs a étè accomplit. Le travail a étè developé en deux versants principaux et complémentaires: Dans une des versants on a etudié, en detail, l'ímportance relative des interactions qui ménent à la stabilisation du complexe formé par les deux proteines - tripsine:PTI. Ce but a étè accomplit en faisant des mutations appropiées dans la region de contact des proteines, par le moyen de méthodes de simulation moléculaire. Dans l'autre versant, on a fait un étude quantique des interactions arginine-carboxylate, qui jouent un rôle trés important dans le méchanisme d'inhibition du facteur β-XIIa. En plus, une nouvelle méthode, basée dans la partition spatial des matrices de la densité et de la densité de paire réduites et aussi dans le formalisme des intégraux indistincts, a étè proposée pour la decomposition des énergies moléculaire et de stabilisation d'un système dans des components associés aux fragments. Cette méthode est particulièrement approprié pour l'étude du transfer d'un group avec liaison covalent, comme c'est le cas de l'interaction arginine-carboxylate.

Dans ce travail, un noveau logiciel de compression de fichiers de trajectoires, obtenus dans les programmes de dynamique moléculaire, a été developé - Byte Structure Variable Lenght Coding (BS-VLC), dont l'efficacité est beaucoup plus grande que celle des algorithmes conservatives classiques.

Dans le primier chapitre, on fait une description de certains aspects génerales concernant les mécanismes de catalyse et d'inhibition des proteases sérines. La tripsine, son inhibiteur naturel PTI et le facteur β-XIIa, sont présentés avec plus de detail car ils sont étudiés dans ce travail.

Dans le deuxième chapitre, les méthodes de simulation moléculaire son décrites. Cettes méthodes sont utilisées dans ce travail pour étudier les interactions responsables de la stabilité du complexe tripsine:PTI, notamment les méthodes de dynamique moléculaire, de perturbation thermodynamique et de l'integration thermodynamique. L'algorithme de compression BS-VLC, ici proposé, est aussi presenté dans ce chapitre.

Dans le troisième chapitre sont presentés les méthodes quantiques utilisés dans l'étude des interactions arginine-carboxylate. Le noveau formalism de décomposition des énergies moléculaire et de stabilisation est ici presenté avec detail.

Le quatrième chapitre présente les résultats obtenus avec ces méthodes.

Le dernier chapitre présente une conclusion génerale du travail ainsi comme quelques sugestions pour son dévelopement future.

ÍNDICE

Prefácio 17

Referências 22

Capítulo 1 - PROTEASES SERINAS: CATÁLISE E INIBIÇÃO 25

1.1 - Mecanismo de acção catalítica 27

1.2 - Inibição 31

1.3 - Tripsina e PTI 35

1.4 - Factor β-XIIa 42

Referências 44

Capítulo 2 - SIMULAÇÕES MOLECULARES 47

2.1 - Introdução 49

2.2 - Modelos de energia potencial 56

2.3 - Cálculo de propriedades em simulações moleculares 60

2.3.1 - Cálculo de observáveis 60

3.3.2 - Cálculo de variações de energia de Helmholtz 62

2.4 - Método da Dinâmica Molecular 71

2.5 - Compressão de ficheiros de trajectórias obtidas por Dinâmica Molecular 85

Referências 99

Capítulo 3 - MÉTODOS QUÂNTICOS 105

3.1 - Introdução 107

3.2 - Formalismo das matrizes densidade 113

3.3 - Métodos aproximados 117

3.4 - Energias de estabilização 141

3.5 - Decomposição da energia molecular e de estabilização 152

3.5.1 - Considerações gerais 152

3.5.2 - Método de Morokuma 154

3.5.3 - Método de Ruedenberg 161

3.5.4 - Método de Clementi 175

3.5.5 - Método de Kollmar 177

3.5.6 - Método de Bofill 179

3.5.7 - Método de Melo e Ramos 184

Referências 210

Capítulo 4 - RESULTADOS 217

4.1 - Introdução 219

4.2 - Estudo das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI 219

4.3 - Compressão de ficheiros de trajectórias obtidas por Dinâmica Molecular 241

4.4 - Estudo quântico das interacções arginina-carboxilato 246

Referências 282

Capítulo 5 - CONCLUSÕES 287

Referências 293

Apêndice I - Variações de energia de Helmholtz obtidas nas simulações computacionais das mutações estudadas na região de contacto tripsina:PTI 295

Apêndice II - Desvios quadráticos médios totais obtidos nas simulações computacionais das mutações estudadas na região de contacto tripsina:PTI 307

Prefácio

A espantosa realidade das coisas.
É a minha descoberta de todos os dias.

Fernando Pessoa, extracto de *O Guardador de Rebanhos*, *Poemas de Alberto Caeiro*, Ática, Lisboa (1983).

As enzimas controlam a maioria dos processos químicos que ocorrem nos seres vivos. A melhor maneira de investigar o mecanismo de catálise destes biopolímeros seria cristalizar os complexos enzima:substrato e analisar as suas estruturas através de técnicas de difracção de raios X. Infelizmente, um complexo deste tipo é extremamente instável e uma vez formado a reacção catalítica inicia-se originando muito rapidamente o correspondente complexo enzima:produto.

Os inibidores naturais das enzimas são habitualmente pequenas proteínas que desempenham um papel biológico extremamente importante nos mecanismos de controle e defesa dos seres vivos. Os inibidores ligam-se às enzimas de uma forma muito semelhante à dos substratos. No entanto, um complexo enzima:inibidor é estabilizado por um número grande de interacções específicas que inactivam a enzima bloqueando o seu centro activo.

Neste contexto, o estudo teórico das reacções de associação enzima:inibidor é muito importante como modelo para os correspondentes processos enzima:substrato. Estes estudos podem também contribuir para uma melhor compreensão das interacções responsáveis pela estabilidade dos complexos enzima:inibidor e este conhecimento pode ser muito importante no desenho de novos inibidores específicos.

As proteases serinas constituem uma classe de enzimas muito importante cujos membros, embora com funções biológicas muito diversas, catalisam a ruptura das ligações peptídicas com base num mecanismo catalítico comum e no mesmo centro activo (constituído pela tríade Ser^{195}-His^{57}-Asp^{102}, cuja numeração é baseada na quimotripsina). A tripsina é uma enzima importante desta classe que catalisa a ruptura hidrolítica de ligações peptídicas em posições adjacentes a aminoácidos carregados positivamente [1,2]. O PTI (Pancreatic Trypsin Inhibitor) é um inibidor natural da tripsina que actua no pâncreas.

Um número significativo de estudos teóricos e experimentais foram efectuados com o objectivo de permitir uma melhor compreensão dos mecanismos biológicos fundamentais associados às proteases serinas. Estes estudos incluem cálculos quânticos e clássicos [3-8], determinações espectroscópicas [9-12], análise estatística de estruturas cristalográficas [13-15] e simulações moleculares [16-28]. No entanto, apesar das interacções responsáveis pela estabilidade complexo tripsina:PTI terem sido identificadas através da análise de estruturas cristalográficas [29-30], nenhum estudo sistemático da sua importância relativa tinha sido efectuado até ao momento. Este facto motivou uma das linhas de investigação desenvolvida neste trabalho em que se estudaram as interacções que ocorrem na superfície de

contacto do complexo tripsina:PTI. A metodologia utilizada consistiu em efectuar mutações de aminoácidos específicos por outros não específicos recorrendo a ciclos termodinâmicos apropriados e aos métodos da dinâmica molecular, da integração termodinâmica e da perturbação termodinâmica.

Nestas simulações produziram-se ficheiros de trajectórias que, no conjunto, totalizaram cerca de 30 *Gbytes* de informação. Neste contexto, desenvolveu-se um novo algoritmo de compressão, o Byte Structure Variable Length Coding (BS-VLC), que apresenta eficiências significativamente melhores que os algoritmos conservativos clássicos. A aplicação do algoritmo BS-VLC aos ficheiros de trajectórias produzidos permitiu reduzir significativamente o volume de informação armazenado.

O factor β-XII a é uma enzima, pertencente à classe das proteases serinas, que está envolvida no sistema de coagulação do sangue. Apesar da sua função biológica ser muito diferente da da tripsina, as duas enzimas apresentam um grau de homologia sequencial bastante elevado. No entanto, o PTI não inibe o factor β-XII a.

Neste contexto, o estudo quântico das interacções arginina-carboxilato, que parecem desempenhar um papel muito importante no mecanismo de inibição do factor β-XII a [31], apresenta bastante interesse. Este facto motivou uma segunda linha de investigação neste trabalho.

As interacções arginina-carboxilato podem ocorrer na forma de dois complexos ligantes (duplo ião e neutro) cuja estabilidade depende do meio. Nos dímeros factor β-XII a:inibidor esta interacção ocorre num meio hidrofóbico.

Neste trabalho, utilizaram-se os modelos moleculares metilguanidínio-acetato e metilguanidina-ácido acético para simular os complexos duplo ião e neutro, respectivamente. O estudo das interacções arginina-carboxilato foi efectuado no vazio, o que permite simular de uma forma razoável o meio hidrofóbico em que estas ocorrem nos dímeros factor β-XII a:inibidor. Este estudo pretendeu encontrar algum esclarecimento em relação às seguintes questões fundamentais:

Quais as conformações mais estáveis para cada um dos complexos ligantes?

Qual o nível mais baixo de cálculo quântico que permite obter uma descrição adequada destas interacções?

Qual dos dois complexos ligantes é mais estável em meios hidrofóbicos?

Qual é o mecanismo associado à reacção de transferência protónica que permite transformar um complexo ligante no outro?

Qual é a energia de activação associada a esta reacção?

Quais os efeitos responsáveis pela estabilização preferencial de cada um dos complexos ligantes?

Neste trabalho, foi ainda proposto um método de partição, dentro de um formalismo *ab initio* multiconfiguracional ou monoconfiguracional, que se baseia no facto da energia de um sistema molecular poder ser completamente determinada através das suas matrizes densidade e densidade de par. De acordo com este método, as energias molecular e de estabilização são decompostas em componentes associadas aos fragmentos que constituem o sistema. Esta metodologia permite uma interpretação útil e flexível dos efeitos responsáveis pela estabilidade de um dado sistema molecular e foi aplicada no estudo das interacções arginina-carboxilato.

Referências

[1] R. M. Stroud, *Sci. Amer.*, **231**, 74 (1974).

[2] J. Kaut, *Ann. Rev. Biochem.*, **46**, 331 (1977).

[3] V. Deggett, S. Shroder e P. A. Kollman, *J. Am. Chem. Soc.*, **113**, 8926 (1991).

[4] I. M. Kovach, D. Huhuta e S. Baptist, *J. Mol. Struct. (Theochem)*, **72**, 99 (1991).

[5] I. M. Kovach e D. Huhuta, *J. Mol. Struct. (Theochem)*, **79**, 335 (1991).

[6] A. Bencsura, I. Enyedy e M. Kovach, *Biochemistry*, **34**, 8989 (1995).

[7] G. Dive, D. Dehareng e D. Peeters, *Int. J. Quantum Chem.*, **58**, 85 (1996).

[8] A. Beveridge, *Protein Sci.*, **5**,1355 (1996).

[9] A. K. Whiting e W. L. Peticolus, *Biochemistry*, **33**, 552 (1994).

[10] N. Wellner e G. Zundell, *J. Mol. Struct.*, **317**, 249 (1994).

[11] P. A. Frey, S. A. Whitt e J. B. Tobin, *Science*, **264**, 1927 (1994).

[13] F. J. Kull, *Biochem. Arch.*, **7**, 245 (1991).

[14] S. J. Hubbard, J. M. Thornton e S. F. Campbell, *Faraday Discuss.*, **93**, 13 (1992).

[15] J. J. Perona e C. S. Craig, *Protein Sci.*, **4**, 337 (1995).

[16] C. F. Wong e J. A. McCammon, *J. Am. Chem. Soc.*, **108**, 3830 (1986).

[17] C. X. Xang, Y. Y. Shi, F. Zhou e L. Wang, *Proteins Struct. Funct. Genetics*, **15**, 5 (1993).

[18] A. P. Heiner, H. J. C. Berendsen e W. F. van Gunsteren, *Protein Eng.*, **6**, 397 (1993).

[19] S. Miyamoto e P. A. Kollman, *Proc. Natl. Acad. Sci.*, USA **90**, 8402 (1993).

[20] S. Nakagawa, H. A. Yu, M. Karplus e H. Unseyama, *Proteins Struct. Funct. Genetics*, **16**, 172 (1993).

[21] I. V. Kurinov e R. W. Harrison, *Nature Struct. Biol.*, **1**, 735 (1994).

[22] J. Åqvist, *J. Comp. Chem.*, **17**,1587 (1996).

[23} W. L. Jorgensen, E. M. Duffy, J. W. Essex, D. L. Severance, J. F. Blake, D. K. Joneshertzog, M. L. Lamb e J. Tiradorives, *Biomol. Struct. Dyn.*, **342**, 21 (1997).

[24] D. K. Joneshertzog e W. L. Jorgensen, *J. Med. Chem.*, **40**, 1539 (1997).

[25] J. W. Essex, D. L. Severance, J. Tiradorives e W. L. Jorgensen, *J. Phys. Chem. B*, **101**, 9663 (1997).

[26] H. Resat, T. J. Marrone e J.A. McCammon, *Biophys. J.*, **72**, 522 (1997).

[27] Z. Y. Guo e C. L. Brooks, *J. Am. Chem. Soc.*, **120**, 1920 (1998).

[28] Y. Sugita e A. Kitao, *Proteins Struct. Funct. Genetics*, **30**, 388 (1998).

[29] J. Janin and C. Chothia, J. Mol. Biol. **100**, 197 (1976).

[30] W. Bode, P. Schwager, and R. Huber, Miami Winter Symp. **11**, 43 (1976).

[31] R. Wynn e M. Laskowski, Jr., *Biochem. Biophys. Res. Commun.* , **166**, 1406 (1990).

Capítulo 1: Proteases Serinas: Catálise e Inibição

Se o Todo-Poderoso me tivesse consultado antes de dar início à criação, teria recomendado algo de mais simples.

Afonso X de Castela.

1.1 - Mecanismo de acção catalítica

As enzimas controlam praticamente todos os processos químicos que ocorrem em organismos biológicos. Estas proteínas apresentam habitualmente um elevado grau de especificidade relativamente aos processos químicos catalisados e às espécies (substratos) neles envolvidos. Na ausência de enzimas, quase todos as reacções biomoleculares seriam demasiado lentas tornando inviáveis os mecanismos fundamentais dos organismos vivos.

As proteases, que catalisam a decomposição de outras proteínas, constituem um dos mais importantes grupos de enzimas. Dentro deste grupo, as proteases serinas formam uma classe importante, cujos membros são essenciais numa série de mecanismos biológicos fundamentais. De facto, as proteases serinas participam na digestão, na coagulação do sangue, na fecundação do óvulo e em processos imunitários [1-2].

As proteases serinas podem ser classificadas em famílias tendo em conta as respectivas homologias sequenciais e estruturas tridimensionais. Neste contexto, incluem-se as famílias da tripsina e da subtilisina que serão especificamente abordadas neste capítulo. Os membros da família da tripsina ocorrem não só nos vertebrados, e noutros tipos de organismos superiores, mas também nos organismos procarióticos [2]. Por outro lado, as subtilisinas apenas surgem em bacilos [2].

Apesar das suas proveniências, funções fisiológicas, sequências de aminoácidos e estruturas tridimensionais serem muito diversas, todas as proteases serinas catalisam a ruptura hidrolítica das ligações peptídicas com base no mesmo mecanismo catalítico e no mesmo centro activo [1-2].

A actividade catalítica destas enzimas está associada à presença de uma serina que é comum aos respectivos centros activos. Consequentemente, estas enzimas serão representadas pela notação (E-OH), onde OH é o grupo hidróxido terminal da cadeia lateral da serina reactiva.

O mecanismo da ruptura hidrolítica de uma ligação peptídica catalisada por uma protease serina é apresentado na figura (1.1):

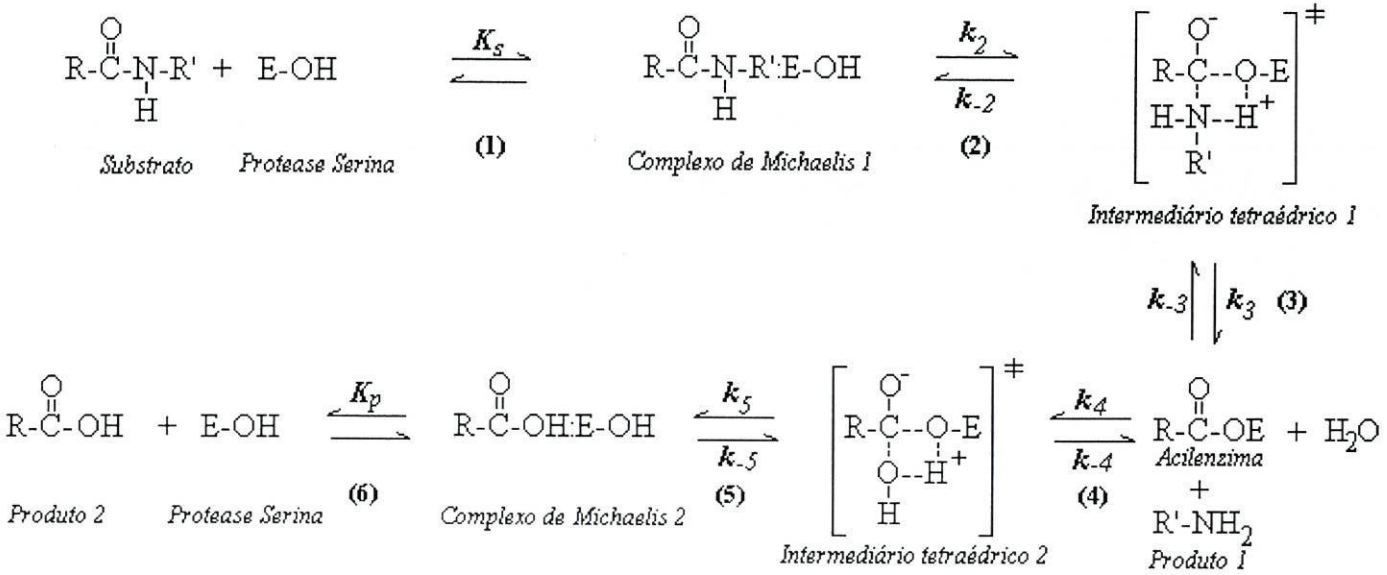


Figura 1.1. Mecanismo da ruptura hidrolítica de uma ligação peptídica catalisada por uma protease serina. Nesta figura, k, K_S e K_p representam uma constante de velocidade, a constante de equilíbrio para a reacção de formação do complexo de Michaelis 1 através da ligação da protease serina ao respectivo substrato e a constante de equilíbrio para a reacção de decomposição do complexo de Michaelis 2 originando a protease serina e o produto 2, respectivamente.

O mecanismo apresentado na figura (1.1) inclui os seguintes passos:

(1) A protease serina liga-se ao substrato para formar o complexo de Michaelis 1.

(2) Este complexo rearranja-se formando o intermediário tetraédrico 1.

(3) Este intermediário decompõe-se formando o produto 1 e a acilenzima.

(4) A acilenzima é hidrolisada dando origem a um novo intermediário tetraédrico.

(5) Este intermediário decompõe-se formando o complexo de Michaelis 2.

(6) Este complexo separa-se regenerando a protease serina e formando o produto 2.

No entanto, só após a determinação da estrutura cristalográfica de algumas proteases serinas, tais como a quimotripsina [3], a tripsina [4] e a elastase, e dos seus complexos com inibidores, como por exemplo o complexo tripsina:PTI, foi possível elaborar um modelo molecular detalhado, embora incompleto, do mecanismo de acção catalítica destas enzimas.

Estas estruturas são do tipo globular, e a sua comparação permitiu estabelecer que o centro activo destas enzimas é constituído por uma tríade de aminoácidos. Esta tríade constitui um sistema de deslocalização de carga que favorece os processos de transferência de protão associados ao mecanismo catalítico apresentado na figura (1.1). O centro activo das proteases serinas inclui, não só a serina reactiva anteriormente referida, mas também uma histidina e um aspartato. Num sistema de numeração convencional, baseado no alinhamento das proteases serinas com a quimotripsina, esta tríade é identificada pela sequência (Ser^{195}-His^{57}-Asp^{102}). Foi também adoptado um sistema de numeração dos resíduos do substrato a partir da ligação peptídica hidrolisada, designando-os por P_1, P_2, P_3, ... , P_n, na direcção do terminal amónio e P_1', P_2', P_3', ... , P_n', na direcção do terminal carboxilato. Os correspondentes resíduos na região de contacto da protease serina foram designados por S_1, S_2, S_3, ... , S_n e S_1', S_2', S_3'... , S_n'.

Neste contexto, o mecanismo apresentado na figura (1.1) pode ser adequadamente explicitado com base no modelo molecular descrito na figura (1.2).

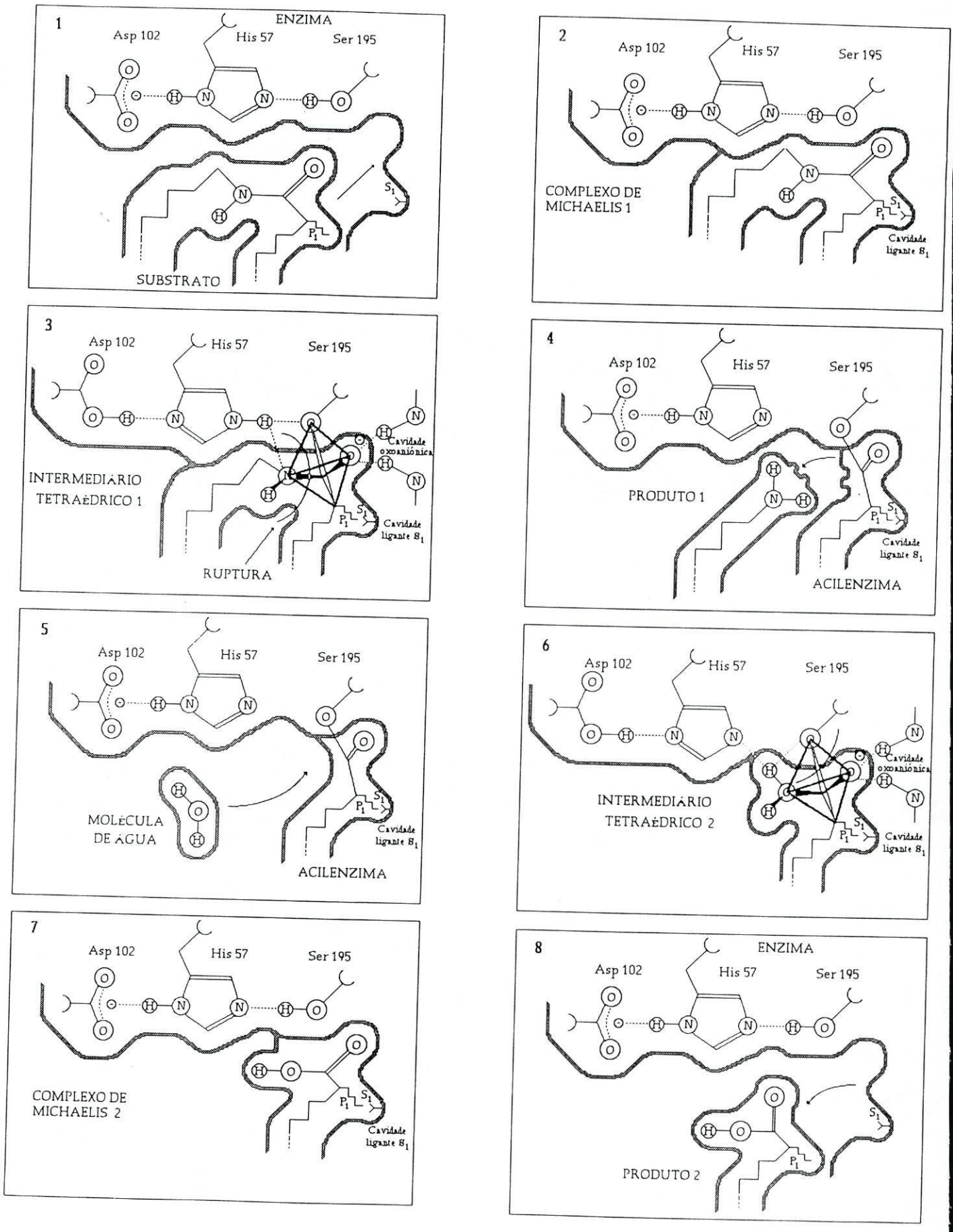


Figura 1.2. Modelo molecular do mecanismo de acção catalítica das proteases serinas, baseado em Stroud [1].

Na figura (1.2), podem ser observados alguns dos requisitos estruturais associados à actividade catalítica destas enzimas, designadamente:

i) Uma cavidade ligante (designada por cavidade ligante S_1) adjacente ao centro activo onde a cadeia lateral do resíduo P_1 do substrato interactua com a cadeia lateral do resíduo S_1 da protease serina (ver quadros 2 a 7). Este é um requisito estrutural de importância fundamental na especificidade catalítica apresentada por estas enzimas.

ii) Uma cavidade oxoaniónica onde os intermediários tetraédricos 1 e 2 são acomodados e estabilizados por interacções do tipo ligação de hidrogénio com grupos NH (ver quadros 3 e 6).

Existem ainda mais alguns requisitos estruturais adicionais não directamente observáveis na figura (1.2), tais como:

iii) A região ligante protease serina:substrato que apresenta uma simetria do tipo β-antiparalela.

iv) Uma cavidade específica onde o produto 1 ($R'NH_2$) é, em algumas proteases serinas, acomodado.

1.2 - Inibição

Os inibidores são moléculas que se ligam às enzimas de uma forma semelhante aos substratos. De facto, a estrutura tridimensional dos complexos enzima:inibidor é muito semelhante à dos correspondentes complexos enzima:substrato. No entanto, os inibidores bloqueiam os centros activos das enzimas estabelecendo com estas um complicado sistema de interacções. Desta forma, os inibidores desactivam as enzimas impedindo que estas exerçam a sua actividade catalítica.

Os inibidores podem classificar-se em duas categorias: inibidores irreversíveis e inibidores reversíveis. Os inibidores irreversíveis estabelecem interacções extremamente fortes (normalmente do tipo covalente) com as enzimas. Consequentemente, uma vez formado o complexo enzima:inibidor ele dificilmente se decompõe, o que implica em desactivação definitiva da enzima. Pelo contrário, os inibidores reversíveis têm interacções fracas (normalmente do tipo electrostático, ligação de hidrogénio ou de van der Waals) com as enzimas. Desta forma, estabelece-se um equilíbrio entre as formas complexada e livre da enzima que é facilmente reversível de acordo com as condições do meio.

Os inibidores naturais são habitualmente do tipo reversível e desempenham um importante papel biológico de controle e protecção. Nos seres vivos, as enzimas são muitas vezes sintetizadas em órgãos onde não interessa que iniciem a sua actividade catalítica. Por isso, são então desactivadas temporariamente por inibidores reversíveis específicos. Estes cessam a sua acção quando as enzimas penetram nos órgãos onde devem desempenhar o seu papel catalítico. Seguidamente referem-se as famílias mais importantes de inibidores naturais das proteases serinas.

A família da abóbora contém os inibidores mais pequenos das proteases serinas [6]. De facto, os inibidores desta família têm de 27 a 30 aminoácidos e são originários das sementes da abóbora e do pepino. A estrutura básica desta família de inibidores é apresentada na figura (1.3); a estrutura comum destes inibidores contém 6 cisteínas envolvidas em 3 pontes dissulfureto, uma localização específica para o resíduo P_1 e uma região razoavelmente conservada dentro dos diferentes inibidores da família incluindo a seguinte sequência de resíduos (C-P-x(5)-C-x(2)-D-x-D-C-x(3)-C-x-C).

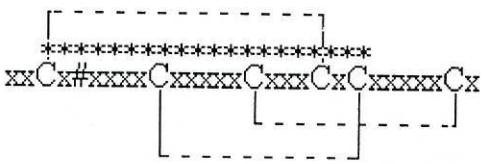


Figura 1.3. Representação esquemática da estrutura comum dos inibidores das proteases serinas da família da abóbora incluindo (C) cisteínas envolvidas nas pontes dissulfureto; (#) localização do resíduo P_1; (*) região melhor conservada.

Os inibidores pancreáticos da tripsina de Kunitz têm cerca de 50 aminoácidos e são de origem animal [7]. A sua estrutura básica está representada na figura (1.4), apresentando 6 cisteínas envolvidas em 3 pontes dissulfureto, uma localização específica para o resíduo P_1 e uma região razoavelmente dentro dos diferentes inibidores da família incluindo a seguinte sequência de resíduos (F-x(3)-G-C-x(6)-[FY]-x(5)-C).

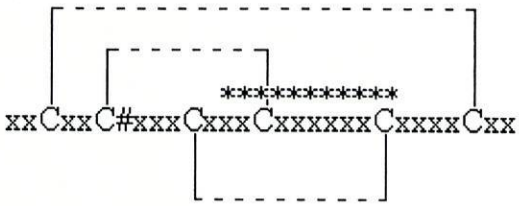


Figura 1.4. Representação esquemática da estrutura comum da família dos inibidores pancreáticos da tripsina de Kunitz incluindo (C) cisteínas envolvidas nas pontes dissulfureto; (#) localização do resíduo P_1; (*) região melhor conservada.

Os inibidores das proteases serinas da família de Bowman-Birk têm cerca de 70 aminoácidos e são originários das sementes de plantas leguminosas e dos grãos de cereais [7-8]. A sua estrutura básica é apresentada na figura (1.5) - podem observar-se 14 cisteínas envolvidas em 7 pontes dissulfureto, duas localizações possíveis para o resíduo P_1 e uma região razoavelmente dentro dos diferentes inibidores da família incluindo a seguinte sequência de resíduos (C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDKS]-[DEKRHSTA]-C).

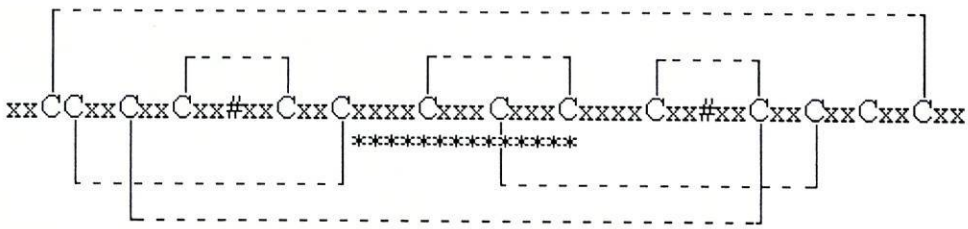


Figura 1.5. Representação esquemática da estrutura comum dos inibidores das proteases serinas da família de Bowman-Birk incluindo (C) cisteínas envolvidas nas pontes dissulfureto; (#) localizações possíveis do resíduo P_1; (*) região melhor conservada.

Os inibidores das proteases serinas da família da batata têm de 60 a 90 aminoácidos e são os únicos que ocorrem tanto em plantas como em animais [9]. De facto estes inibidores podem ser encontrados nas sementes da cevada e do feijão, na batata e no tomate, mas também nas sanguessugas. Estruturalmente são caracterizados por terem uma única localização para o resíduo P_1 e por serem os únicos que não apresentam qualquer ponte dissulfureto.

Os inibidores da tripsina e da α-amilase provenientes dos cereais têm cerca de 120 aminoácidos [10]. A sua estrutura básica é apresentada na figura (1.6); a estrutura comum destes inibidores contém 10 cisteínas envolvidas em 5 pontes dissulfureto, a localização do resíduo P_1 não é específica e existe uma região razoavelmente dentro dos diferentes inibidores da família incluindo a seguinte sequência de resíduos (C-x(4)-[SAGD]-x(4)-[SPAL]-[LF]-x(2)-C-[RH]-x-[LIVMFY](2)-x(3,4)-C).

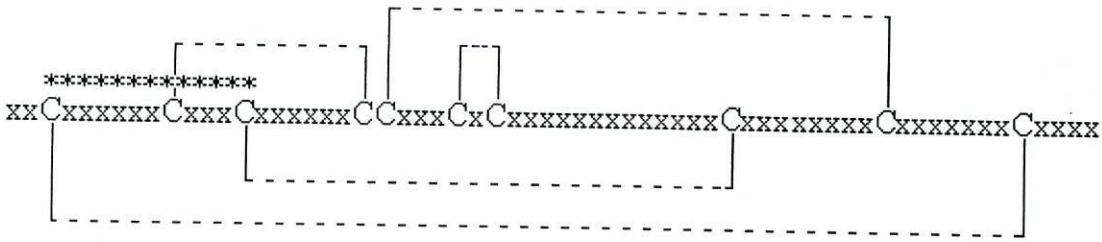


Figura 1.6. Representação esquemática da estrutura comum dos inibidores da tripsina e da α-amilase provenientes dos cereais incluindo (C) cisteínas envolvidas nas pontes dissulfureto; (*) região melhor conservada.

Os inibidores das proteases serinas de Kunitz da família da soja apresentam 170 a 200 aminoácidos e uma ou duas pontes dissulfureto [7]. Nesta família, a localização do resíduo P_1 não é específica e a sequência de aminoácidos mais bem conservada encontra-se na região terminal amónio incluindo a seguinte sequência de resíduos ([LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM])

Os inibidores das proteases serinas da família de Kazal são de origem animal [7]. A sua estrutura básica é apresentada na figura (1.7). Como se pode observar da análise desta figura , a estrutura comum destes inibidores contém 6 cisteínas envolvidas em 3 pontes dissulfureto, uma localização específica para o resíduo P_1 e uma região razoavelmente dentro dos diferentes inibidores da família incluindo a seguinte sequência de resíduos (C-x(7)-C-x(6)-Y-x(3)-C-x(2,3)-C).

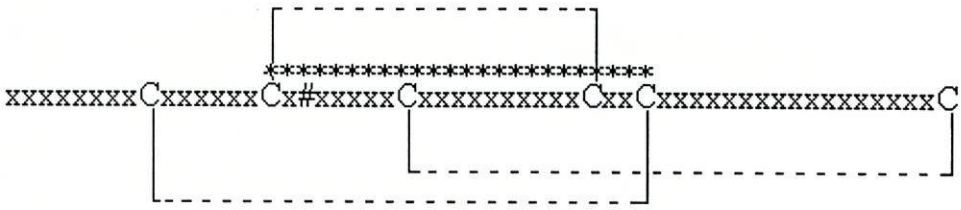


Figura 1.7. Representação esquemática da estrutura comum dos inibidores das proteases serinas da família de Kazal incluindo (C) cisteínas envolvidas nas pontes dissulfureto; (#) localização do resíduo P_1; (*) região melhor conservada.

Por fim, os inibidores da subtilisina da bactéria *Streptomices* [11] têm cerca de 110 aminoácidos e apresentam uma estrutura básica do tipo da figura (1.8) com determinadas características comuns: 4 cisteínas envolvidas em 2 pontes dissulfureto, uma localização específica para o resíduo P_1 e uma região razoavelmente dentro dos diferentes inibidores da família incluindo a seguinte sequência de resíduos (C-x-P-x(2,3)-G-x-H-P-x(4)-A-C-[ATD]-x-L).

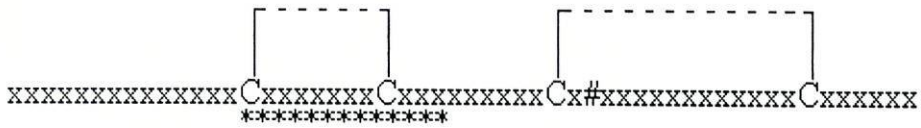


Figura 1.8. Representação esquemática da estrutura comum dos inibidores da subtilisina da bactéria *Streptomices* incluindo (C) cisteínas envolvidas nas pontes dissulfureto; (#) localização do resíduo P_1; (*) região melhor conservada.

1.3- Tripsina e PTI

A estrutura tridimensional da cavidade ligante S_1 (ver figura 1.2) e a natureza química dos resíduos nela contidos (principalmente do resíduo S_1) determinam fundamentalmente as especificidades apresentadas pelas proteases serinas relativamente a substratos e inibidores [1-2].

Na família da tripsina, os principais tipos de especificidade podem ser associados à elastase, à tripsina e à quimotripsina:

A elastase apresenta uma cavidade ligante S_1 parcialmente obstruída pelas cadeias laterais dos resíduos Val^{216} e Thr^{226}. Consequentemente, esta enzima apenas permite acomodar substratos com resíduos P_1 pequenos (por exemplo glicina ou alanina). Pelo contrário, a quimotripsina e a tripsina apresentam cavidades ligantes S_1 amplas o que permite acomodar substratos com resíduos P_1 de maior dimensão. No entanto, enquanto na tripsina o resíduo S_1 (Asp^{189}) apresenta uma cadeia lateral carregada negativamente, na quimotripsina este resíduo é substituído pela Ser^{189} de cadeia lateral neutra e a cavidade ligante S_1 é muito mais hidrofóbica. Consequentemente, a tripsina apresenta uma especificidade acentuada relativamente a substratos com resíduo P_1 de cadeia lateral carregada positivamente (lisina ou arginina), enquanto a quimotripsina prefere substratos com resíduo P_1 de cadeia lateral aromática. As especificidades catalíticas da quimotripsina, tripsina e elastase são apresentadas na figura (1.9).

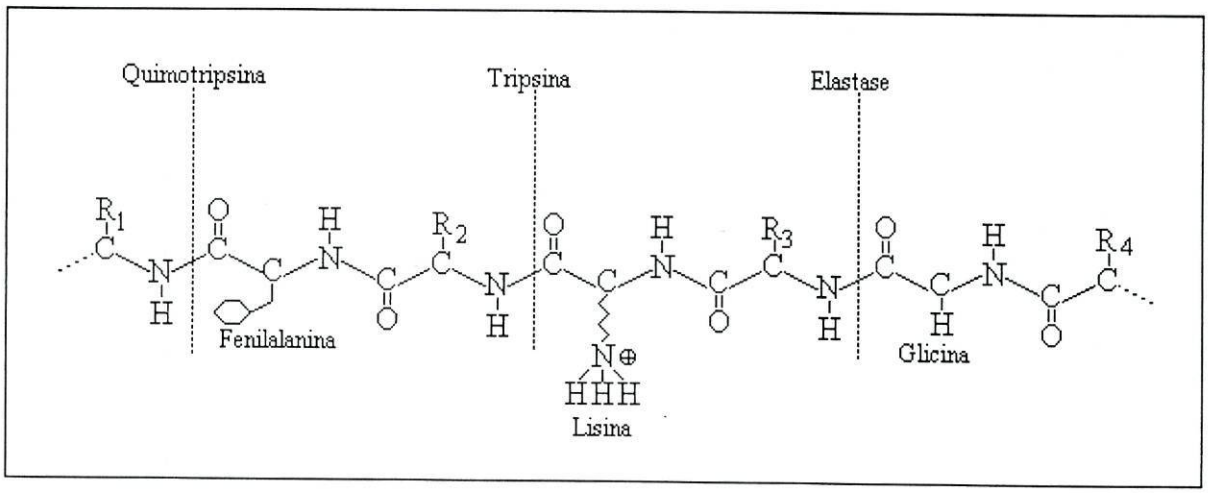


Figura 1.9. Especificidades catalíticas da quimotripsina, tripsina e elastase.

Na família da subtilisina, por exemplo o resíduo P_1 é pouco específico, embora muitas vezes apresente uma cadeia lateral aromática [2]. Pelo contrário, os resíduos P_3 e P_4 são fortemente específicos e têm de ser hidrofóbicos [2].

A tripsina é uma enzima intestinal, constituída por 223 aminoácidos e com uma massa molecular 24000 Da, que participa na decomposição das proteínas ingeridas nos respectivos aminoácidos. As estruturas tridimensionais desta enzima [4] e dos seus complexos com

diversos inibidores [4,12-15] foram determinadas por difracção de raios X com um bom grau de precisão. Por isso, a tripsina é muitas vezes utilizada como modelo para estudos sistemáticos das proteases serinas [16].

Algumas proteases serinas digestivas importantes, tais como a tripsina, a quimotripsina e a elastase, são originárias do pâncreas na forma de proenzimas (percursores inactivos) que são posteriormente activados no duodeno pela própria tripsina. Nestas circunstâncias, a activação prematura da tripsina no pâncreas seria suficiente para iniciar uma reacção catalítica em cadeia que destruiria todas as proteínas aí existentes. Esta catástrofe é evitada devido à presença no pâncreas de inibidores específicos que bloqueiam a tripsina mas não o seu precursor. O PTI (Pancreatic Trypsin Inhibitor) é um dos mais importantes inibidores deste tipo e a sua estrutura tridimensional tem sido amplamente estudada através de difracção de raios X [17-18]. O PTI pertence à família dos inibidores pancreáticos da tripsina de Kunitz, tem 58 aminoácidos e uma massa molecular 6504 Da.

Vincent e Lazdunski [22] determinaram experimentalmente a energia livre de estabilização do complexo tripsina:PTI,

$$\text{Tripsina(aq)} + \text{PTI(aq)} \rightleftharpoons \text{Tripsina:PTI(aq)}\,, \qquad (1)$$

tendo obtido um valor de -18,03 kcal mol^{-1}.

A estabilidade e especificidade dos complexos enzima:inibidor tem sido discutida com base nas interacções fundamentais responsáveis pela sua coesão e nas respectivas estruturas cristalográficas [16,23-24]. Com base nestes modelos, a energia de Helmholtz de estabilização de um complexo enzima:inibidor (ΔA) é decomposta em componentes intra-segmento (ΔA_{intra}) e inter-segmento (ΔA_{inter}):

$$\Delta A = \Delta A_{intra} + \Delta A_{inter} \qquad (2)$$

A componente intra-segmento (ΔA_{intra}) está associada aos rearranjos internos da enzima e do inibidor devido à formação do complexo. Esta componente deve dar uma contribuição repulsiva para a estabilização do complexo, fortemente influenciada por uma previsível diminuição de entropia associada à perda de graus de liberdade dos dois segmentos(enzima e inibidor) no processo de associação [23].

A componente inter-segmento (ΔA_{inter}) pode ser decomposta nos seguinte termos:

$$\Delta A_{inter} = \Delta A_{hidrof.} + \Delta A_{lig.} + \Delta A_{rep.} \quad (3)$$

O termo hidrofóbico ($\Delta A_{hidrof.}$) está essencialmente associado à reorganização do solvente. Este termo deve ser fortemente atractivo fundamentalmente devido ao aumento de entropia associado ao afastamento de um número grande de moléculas de água da enzima e do inibidor durante o processo de associação [16,23-24].

O termo ligante ($\Delta A_{lig.}$) está associado às interacções atractivas (do tipo ligação de hidrogénio, electrostática ou de van der Waals) envolvendo resíduos na superfície de contacto enzima:inibidor. No entanto, o estabelecimento de uma interacção envolvendo um resíduo A da enzima e um resíduo B do inibidor implica normalmente a remoção das interacções que cada um dos resíduos estabelece com o solvente nos respectivos segmentos isolados. Nesta circunstâncias, a contribuição de cada par de interacções A···B para a energia de Helmholtz é o resultado do seguinte processo esquemático:

$$A\cdots H_2O + B\cdots H_2O \rightleftharpoons A\cdots B + H_2O\cdots H_2O\ . \quad (4)$$

Consequentemente, mesmo pontes de hidrogénio do tipo amida apresentam um balanço global pouco significativo em água [23].

O termo repulsivo ($\Delta A_{rep.}$) está associado aos maus contactos que ocorrem na superfície de contacto enzima:inibidor. Esta componente deve ser residual em complexos estáveis.

Normalmente, assume-se que a especificidade das interacções enzima:inibidor está associada fundamentalmente ao termo ligante ($\Delta A_{lig.}$), admitindo que as restantes componentes têm um efeito global próximo de zero [23].

Com base na respectiva estrutura cristalográfica, foram identificadas uma série de interacções ligantes, apresentadas na figura (1.10), responsáveis pela estabilização do complexo tripsina:PTI [16,23].

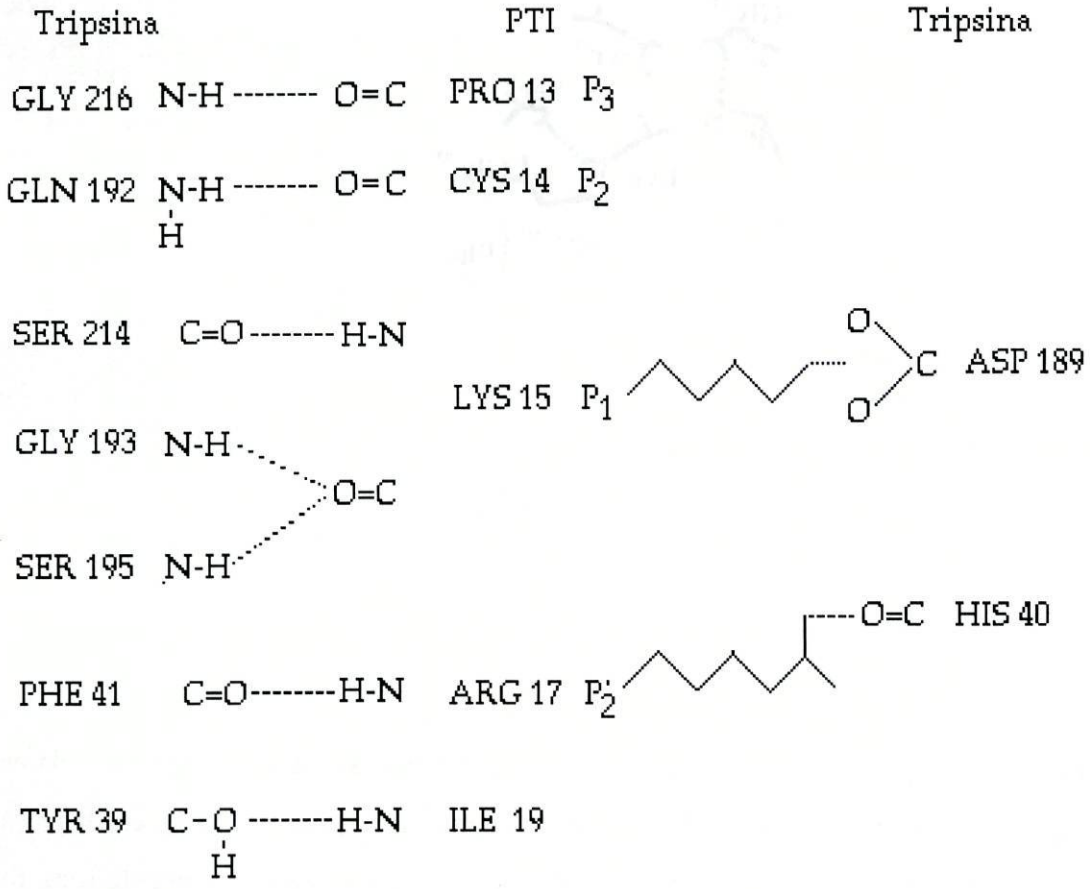


Figura 1.10. Interacções ligantes que estabilizam o complexo tripsina:PTI.

Estas interacções podem ser classificadas em duas classes: interacções envolvendo a cadeia principal (apresentadas na figura 1.11) e interacções envolvendo cadeias laterais (apresentadas nas na figura 1.12).

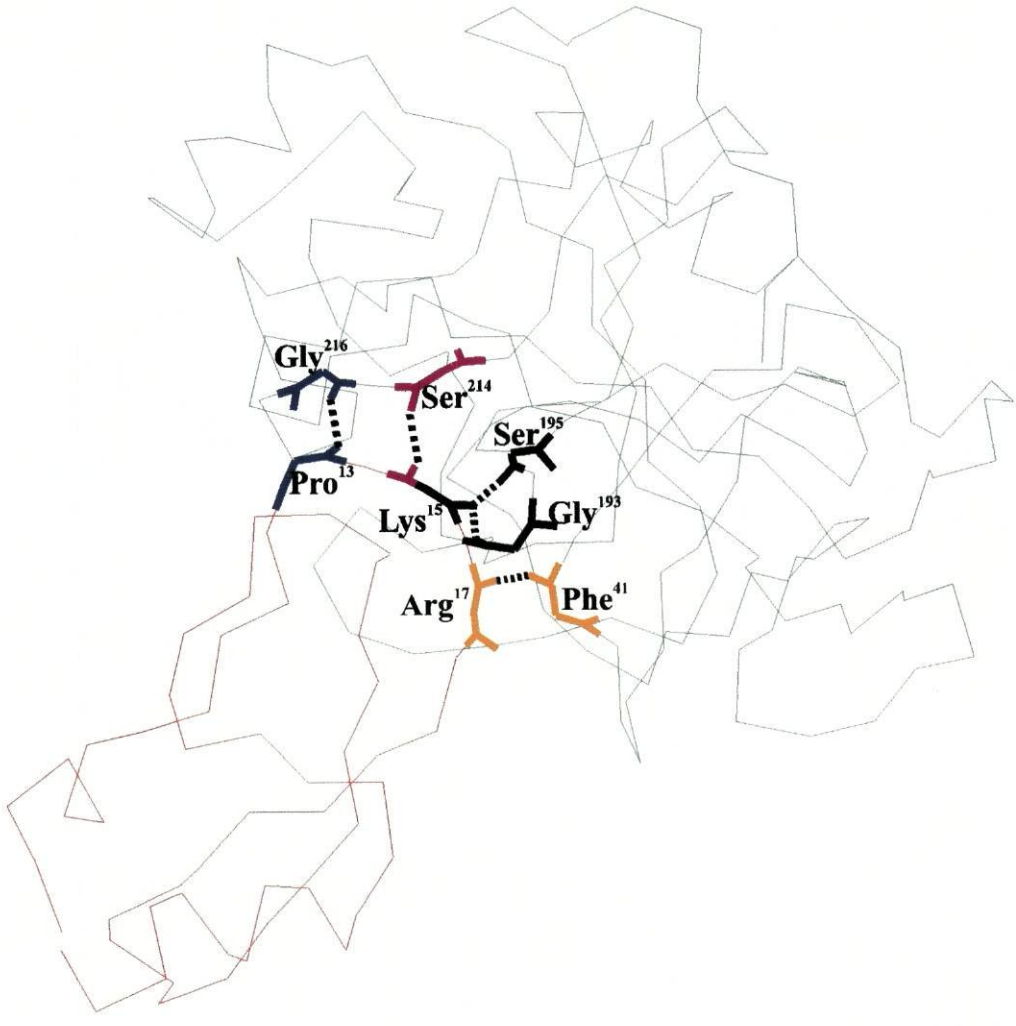


Figura 1.11. Complexo tripsina:PTI com a estrutura tridimensional dos carbonos α da enzima a verde e do inibidor a vermelho. As interacções ligantes, responsáveis pela estabilização do complexo, envolvendo a cadeia principal são também representadas: a cor-de-rosa forte a interacção Ser^{214}(tripsina)-Lys^{15}(PTI), a laranja a interacção Phe^{41}(tripsina)-Arg^{17}(PTI), a preto as interacções Ser^{195}(tripsina)-Lys^{15}(PTI) e Gly^{193}(tripsina)-Lys^{15}(PTI) e a azul a interacção Gly^{216}(tripsina)-Pro^{13}(PTI).

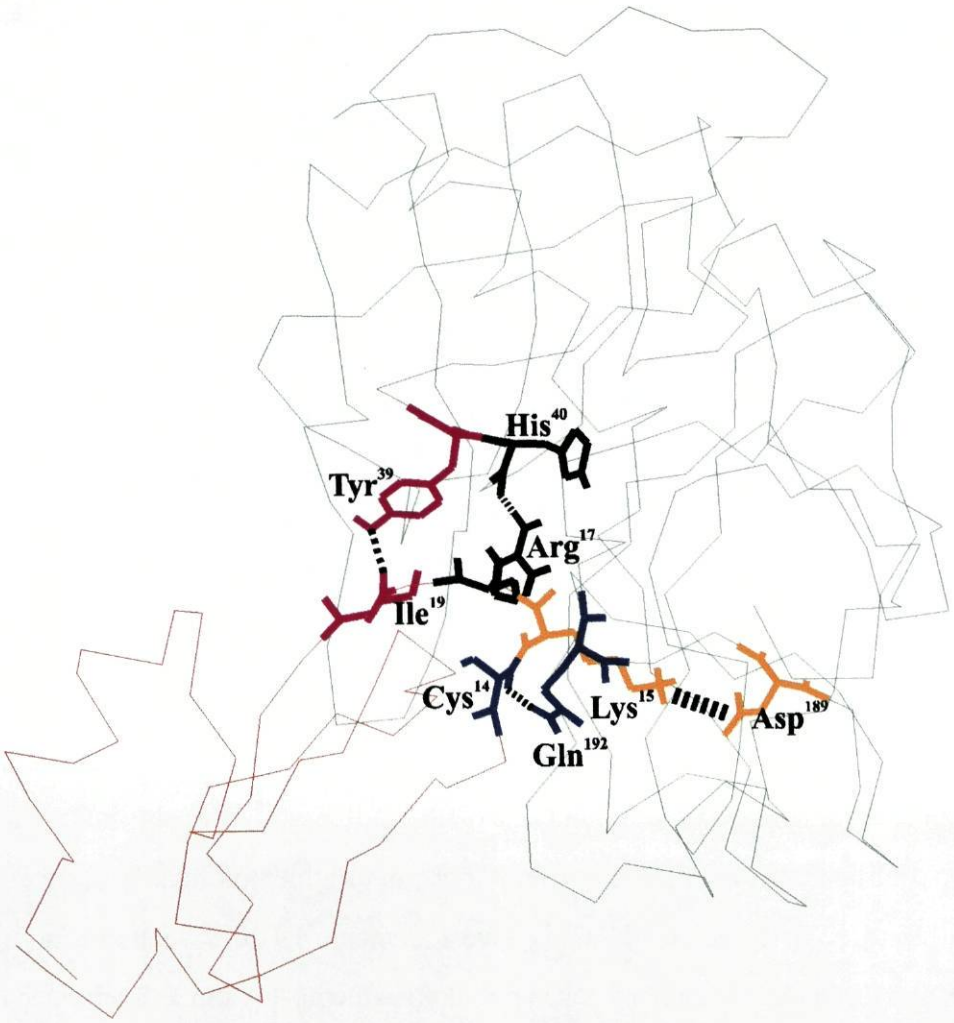


Figura 1.12. Complexo tripsina:PTI com a estrutura tridimensional dos carbonos α da enzima a verde e do inibidor a vermelho. As interacções ligantes, responsáveis pela estabilização do complexo, envolvendo cadeias laterais são também representadas: a laranja a interacção Asp^{189}(tripsina)-Lys^{15}(PTI), a preto a interacção His^{40}(tripsina)-Arg^{17}(PTI), a cor-de-rosa forte a interacção Tyr^{39}(tripsina)-Ile^{19}(PTI) e a azul a interacção Gln^{192}(tripsina)-Cys^{14}(PTI).

Não é possível, quer por métodos teóricos quer por métodos experimentais, determinar directamente a contribuição de cada uma das interacções referidas nas figuras (1.10), (1.11) e (1.12) para a estabilização do complexo tripsina:PTI. No entanto, é possível fazer uma avaliação destas contribuições através da determinação variações de energia de Helmholtz de ligação associadas à mutação de resíduos específicos, relativamente a interacções ligantes no complexo tripsina:PTI, por outros não específicos. A metodologia utilizada nessas determinações será discutida extensivamente nos capítulos 2 e 4.

1.4- Factor β-XIIa

Nos mamíferos, a coagulação do sangue constitui um sistema de defesa e controle que permite manter a integridade e a pressão do aparelho circulatório quando ocorre alguma ruptura num vaso sanguíneo [25-26]. Neste sistema, as glicoproteínas do plasma desempenham um papel de primordial importância. Estas proenzimas são precursores inactivos de proteases serinas, apresentam um alto grau de homologia com a tripsina, quimotripsina e elastase e são convertidas da sua forma inactiva para a sua forma activa através da protólise de apenas uma ou duas ligações peptídicas [25-26].

O sistema de coagulação do sangue é bastante complexo e envolve interacções entre cerca de 14 glicoproteínas [26-28]. O factor XII (factor de Hageman), constituído por 596 aminoácidos e com massa molecular 76000 Da, é um dos componentes deste sistema. Na presença de uma superfície aniónica, o factor XII é activado por calicreínas originando o factor β-XIIa. Esta enzima é constituída por dois segmentos ligados entre si por uma ponte dissulfureto [27]. O segmento mais leve (segmento-L) tem 9 aminoácidos e uma massa molecular 2000 Da. O segmento mais pesado (segmento-H) tem 243 aminoácidos, uma massa molecular 28000 Da e apresenta 6 pontes dissulfureto internas.

O factor β-XIIa (ou mais especificamente o seu segmento-H) apresenta um elevado grau de homologia sequencial com a tripsina, partilhando ainda com esta enzima o mesmo resíduo S_1 (Asp^{189}), o mesmo centro activo e o mesmo mecanismo de acção catalítica [29]. Contudo, o factor β-XIIa não é inibido pelo PTI.

A estrutura terciária do factor β-XIIa é desconhecida, porque estudos experimentais, incluindo difracção de raios X, não conduziram a quaisquer resultados positivos [30]. No entanto, foi elaborado o modelo tridimensional desta enzima [31] com base nas homologias sequenciais com as estruturas bem conhecidas da tripsina [4], quimotripsina [3] e elastase [5]. Posteriormente este modelo foi refeito [32], de acordo com conhecimentos actualizados conseguidos através do estudo experimental de estruturas tridimensionais de outras proteases serinas.

As diferentes afinidades da tripsina e do factor β-XIIa para o PTI devem reflectir-se em diferenças nas respectivas regiões de contacto. Consequentemente, o estudo detalhado das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI, efectuado neste

trabalho, poderá representar um contributo importante no desenho de novos inibidores para o factor β-XIIa, particularmente se o PTI for usado como ponto de partida.

O factor β-XIIa, por apresentar um resíduo S_1 com cadeia lateral carregada negativamente (Asp^{189}), apresenta uma afinidade acentuada relativamente a substratos e inibidores com resíduo P_1 de cadeia lateral carregada positivamente (lisina ou arginina). No entanto, os melhores inibidores desta enzima tem uma arginina como resíduo P_1 [33]. Por outro lado, a cavidade ligante S_1 apresenta um carácter predominantemente hidrofóbico.

As interacções arginina-carboxilato (Arg-Carb), que se estabelecem entre as cadeias laterais da arginina e de aminoácidos do tipo carboxilato (aspartato e glutamato), estão entre as mais importantes que ocorrem em proteínas. De facto, cerca de 40% dos pares iónicos encontrados em proteínas envolvem interacções deste tipo [34-35]; foi também sugerido que estas interacções desempenham um papel importante no reconhecimento de resíduos guanina por proteínas [36], na estabilização de folhas β [37], na estabilidade térmica de proteínas [38], na fotoquímica de proteínas retinais [39-43], no reconhecimento antigénio-anticorpo [44-45] e nas interacções enzima-inibidor [46-48]. Consequentemente, o estudo quântico das interacções Arg-Carb, efectuado neste trabalho, apresenta um grande interesse, não só no estudo do mecanismo de inibição do factor β-XIIa, mas também na compreensão de muitos outros mecanismos de interesse biológico relevante.

Referências

[1] R. M. Stroud, *Sci. Amer.*, **231**, 74 (1974).

[2] J. Kaut, *Ann. Rev. Biochem.*, **46**, 331 (1977).

[3] G. H. Cohen, E. W. Silverton e D. R. Davies, *J. Mol. Biol.*, **148**, 449 (1981).

[4] M. Marquart, J. Walter, J. Deisenhofer, W. Bode e R. Huber, *Acta Cryst.*, **B39**, 480 (1983).

[5] L. Sawyer, D. M. Shoton, J. W. Campbell, P. L. Wendell, H. Muirhead, H. C. Watson, R. Diamond e R. C. Ladner, *J. Mol. Biol.*, **118**, 137 (1978).

[6] J. Otlewski e D. Krowarsh, *Acta Biochem. Pol.*, **43**, 431 (1996)..

[7] M. Laskowski e I. Kato, *Annu. Rev. Biochem.*, **49**, 593 (1980).

[8] M. Tashiro, K. Hashino, M. Shiozaki, F. Ibuki e Z. Mazi, *J. Biochem.*, **102**, 297 (1987).

[9] H. Nozawa, H. Yamagata, Y. Aizono, M. Yoshikawa e T. Iwasaki, *J. Biochem.*, **106**, 1003 (1989).

[10] M.-F. Gaulier, R. Alary e P. Jondirer, *Plant. Mol. Biol.*, **14**, 313 (1990).

[11] S. Taguchi, S. Kojima, M. Terabe K-I. Miura e H. Momose, *Eur. J. Biochem.*, **220**, 911 (1994).

[12] W. Bode, H. J. Greyling, R. Huber, J. Otlevski e T. Wilusz, *Febs. Lett.*, **242**, 285 (1989).

[13] T. A. Holak, D. Gondol, J. Otlewski e T. Wilusz, *J. Mol.Biol.*, **210**, 635 (1989).

[14] M. E. McGrath, T. Erpel, C. Bystioff e R. J. Eletterick, *EMBO J.*, **94**, 8897 (1990).

[15] M. Nakamura, K. Tomoo, T. Ishida, H. Taguchi, Y. Tsuada, Y. Okada, A. Okunomiya, K. Wanaka e S. Okamoto, *Biochem. Biophys. Res. Commun.*, **213**, 583 (1995).

[16] W. Bode, P. Schwager e R. Huber, *Miami Winter Symp.*, **11**, 43 (1976).

[17] A. Wlodawer, J. Walter, R. Huber e L. J. Sjölin, *J. Mol. Biol.*, **180**, 301 (1984).

[18] S. Parkin, B. Rupp e H. Hope, *Acta Cryst.*, **D52**, 18 (1996).

[19] C. L. Brooks III, M. Karplus e B. Petit, *Adv. Chem. Phys.*, **71**, 1 (1988).

[20] W. F. van Gunsteren e M. Karplus, *Biochemistry*, **21**, 2259 (1992).

[21] D. J. Tobias, G. J. Martyna e M. L. Klein, *J. Phys. Chem.*, **97**, 12959 (1993).

[22] J. P. Vincent e M. Lazdunski, *Biochemistry*, **11**, 2967 (1972).

[23] J. Janin e C. Chothia, *J. Mol. Biol.*, **100**, 197 (1976).

[24] A. A. Rashin, *Prog. Biophys. Mol. Biol.*, **60**, 71 (1993).

[25] C. M. Jackson e Y. Nemerson, *Ann. Rev. Biochem.*, **49**, 765 (1980).

[26] B. Furie e B. C. Furie, *Cell*, **53**, 505 (1988).

[27] K. Fujikawa e B. A. McMullen, *J. Biol. Chem.*, **258**, 10924 (1983).

[28] G. J. Broze, *Annu. Rev. Med.*, **46**, 103 (1995).

[29] D. E. Cool e R. T. A. MacGillivray, *J. Biol. Chem.*, **262**, 13662 (1987).

[30] M. Baron, D. Norman, A. Willis e I. D. Campbell, *Nature*, **345**, 642 (1990).

[31] M. J. Ramos, *J. Mol. Graph.*, **9**, 91 (1991).

[32] E. S. Henriques, W. B. Floriano, A. Melo, N. Reuter, M. J. Ramos, B. Maigret, M. C. Nascimento e J. A. N. F. Gomes, *resultados a publicar.*

[33] R. Wynn e M. Laskowski, Jr., *Biochem. Biophys. Res. Commun.* , **166**, 1406 (1990).

[34] D. J. Barlow e J. M. Thornton, *J. Mol. Biol.*, **168**, 867 (1983).

[35] J. Singh, J. M. Thornton, M. Snarey e S. F. Campbell, *Febs. Lett.*, **224**, 161 (1987).

[36] G. Lancelot, R. Mayer e C. Helene, *J. Am. Chem. Soc.*, **101**, 1569 (1979).

[37] D. S. Eggleston e D. J. Hodgson, *Int. J. Peptide Protein Res.*, **25**, 242 (1985).

[38] F. S. Qaw e J. M. Brewer, *Mol. Cell Biochem.*, **71**, 121 (1986).

[39] C. Scharnagl, J. Hettenkofer e S. F. Fisher, *J. Phys. Chem.*, **99**, 7787 (1995).

[40] J. K. Lanyi, *Biophys. J.*, **56**, 143 (1995).

[41] S. P. Balashov, R. Govindjee, E. S. Imasheva, S. Misra, T. G. Ebrey, Y. Feng, R. K. Crouch e D. R. Menich, *Biochemistry*, **34**, 8820 (1995).

[42] C. Scharnagl e S. F. Fisher, *Chem. Phys.*, **212**, 231 (1996).

[43] S. Dickopf e M.P. Heyn, *Biophys. J.*, **73**, 3171 (1997).

[44] S. Sheriff, E. W. Silverton, E. A. Padlan, G. H. Cohen e S. J. Smith-Gill, *Proc. Natl Acad. Sci. USA*, **84**, 8075 (1987).

[45] A. L. Magalhães, B. Maigret, J. Hoplack, J. A. N. F. Gomes e H. A. Scheraga, *J. Protein Chem.*, **13**, 195 (1994).

[46] E. Gawlita, V. E. Anderson e P. Paneth, *Eur. Biophys. J.*, **23**, 353 (1994).

[47] S. K. Sreedharan, C. Veerna, L. S. D. Cases, S. M. Brocklehurst, S. E. Garbia, H. N. Shah e K. Brocklehurst, *Biochem. J.*, **316**, 777 (1996).

[48] D. H. Kim e J. I. Park, *Biorg. Med. Chem. Lett.*, **6**, 2967 (1996).

Capítulo 2: Simulações Moleculares

No nosso esforço para compreender a realidade, a nossa posição lembra a de um homem que procura adivinhar o maquinismo de um relógio fechado. Esse homem vê o mostrador e os ponteiros, ouve o tiquetaque, mas não tem meios para abrir a caixa que esconde maquinismo. Se é um homem engenhoso, pode fazer ideia de um maquinismo responsável por tudo o que observa exteriormente, mas não poderá nunca ter a certeza de que o maquinismo que imagina seja o único que possa explicar os movimentos exteriores.

Albert Einstein e Leopold Infeld, *A Evolução da Física*, Livros do Brasil, Lisboa

2.1 - Introdução

Em termos macroscópicos, o estado termodinâmico (macroestado) de um sistema pode ser completamente determinado através da especificação numérica de um certo número de propriedades designadas por funções de estado. Por exemplo, o macroestado de um sistema com um só componente pode ser determinado por três funções de estado { (NVT); (NVE); (NPT); (μVT) etc., onde N representa o número de partículas, V o volume, T a temperatura, P a pressão e μ o potencial químico}. As restantes propriedades podem ser calculadas através das relações fundamentais da termodinâmica.

A mecânica estatística permite estabelecer a ligação teórica entre as propriedades termodinâmicas macroscópicas e a descrição microscópica do sistema. Habitualmente recorre-se ao conceito de *ensemble* que permite evitar os problemas inerentes ao cálculo de médias temporais em sistemas reais. Um *ensemble* consiste num grande número de réplicas do sistema em estudo, cada uma das quais sujeita às mesmas restrições termodinâmicas do sistema original. Neste contexto, podem definir-se diversos tipos de *ensemble* tendo em conta as funções de estado especificadas { *microcanónico* (NVE); *canónico* (NVT); *isotérmico-isobárico* (NPT); *grande canónico* (μVT) etc. }. No entanto, estas réplicas diferem a nível microscópico, porque existe um grande número de estados quânticos (microestados) compatíveis com um dado macroestado. Neste contexto, as propriedades termodinâmicas podem ser classificadas em propriedades mecânicas ou observáveis e propriedades não mecânicas. As observáveis podem ser determinadas para cada microestado do sistema e o seu valor esperado é calculado como uma média de *ensemble*. Pelo contrário, as propriedades não mecânicas são propriedades globais cujo valor depende de todo o espaço de microestados acessível.

Considerando um sistema cujos macroestados são determinados pelas funções de estado (NVT), o valor esperado de uma observável O pode ser calculado através da equação:

$$\langle O \rangle = \sum_{I} \langle \psi_I | \hat{O} | \psi_I \rangle P_I \qquad (2.1)$$

Na equação (2.1), $\hat{O}$ é o operador associado à observável O, ψ_I é a função de estado do estado quântico I e P_I é a probabilidade de ocorrência deste estado dada por,

$$P_I = \frac{\exp(-\beta E_I)}{Q(NVT)}, \tag{2.2}$$

$$\text{onde } \beta = \frac{1}{k_B T} \tag{2.3}$$

Nas equações (2.2) e (2.3), E_I é a energia do estado quântico I, k_B é a constante de Boltzmann e Q(NVT) é a função de partição do *ensemble* definida como:

$$Q(NVT) = \sum_I \exp(-\beta E_I) \tag{2.4}$$

A energia de Helmholtz (A) constitui um exemplo de uma propriedade não mecânica que, nas condições especificadas, pode ser calculada através da equação:

$$A = -k_B T \ln Q(NVT) \tag{2.5}$$

O uso da equação (2.1) para a determinação do valor esperado da observável O implica a resolução da equação de Schrödinger para o sistema em estudo e a determinação dos valores esperados do operador $\hat{O}$ para todos os microestados acessíveis. No entanto, esta equação só tem soluções analíticas para sistemas monoelectrónicos, e mesmo a utilização de métodos aproximados, que serão discutidos no capítulo 3, se revela inviável pois o número de estados quânticos que contribui para a média de *ensemble* (2.1) é astronómico (da ordem de $10^{10^{25}}$ [1]).

A aproximação semiclássica constitui uma alternativa ao formalismo quântico anteriormente apresentado. Neste contexto, cada microestado é caracterizado por um conjunto de posições $\Omega = (X_1, X_2, \ldots X_N)$ e momentos $\Pi = (p_1, p_2, \ldots, p_N)$ para todas as N partículas do sistema e a sua energia é uma função contínua,

$$E(\Pi;\Omega) = \Theta(\Pi) + V(\Omega) \ , \tag{2.6}$$

onde $\Theta(\Pi)$ representa a energia cinética definida como,

$$\Theta(\Pi) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} \ , \tag{2.7}$$

e $V(\Omega)$ representa a energia potencial cuja modelação será discutida posteriormente na secção (2.2).

Dentro da aproximação semiclássica, o valor esperado de uma observável O é dado por,

$$\langle O \rangle = \int_{\Pi} \int_{\Omega} O(\Pi;\Omega)\, P(\Pi;\Omega)\, d\Pi\, d\Omega \ , \tag{2.8}$$

onde,

$$d\Pi = dp_1\, dp_2 \ldots dp_N \ \text{ e } \ d\Omega = dX_1\, dX_2 \ldots dX_N \tag{2.9}$$

Na equação (2.8), $O(\Pi;\Omega)$ representa a expressão clássica da observável O expressa em termos das coordenadas de posição e dos momentos das N partículas do sistema e $P(\Pi;\Omega)$ é a função probabilidade definida como,

$$P(\Pi;\Omega) = \frac{\exp(-\beta\, E(\Pi;\Omega))}{h^{3N}\, I\, Q(NVT)} \ , \tag{2.10}$$

onde h é a constante de Planck, I é um factor introduzido para impor a indistinguibilidade das partículas idênticas e Q(NVT) é a função de partição semiclássica dada por:

$$Q(NVT) = \frac{\int_{\Pi}\int_{\Omega} \exp(-\beta\, E(\Pi;\Omega))}{h^{3N}\, I} \tag{2.11}$$

O valor esperado de uma observável O pode habitualmente ser decomposto nas suas componentes cinética $\langle O \rangle_{\Pi}$ e conformacional $\langle O \rangle_{\Omega}$:

$$\langle O \rangle = \langle O \rangle_{\Pi} + \langle O \rangle_{\Omega} = \int_{\Pi} O(\Pi)\, P(\Pi)\, d\Pi + \int_{\Omega} O(\Omega)\, P(\Omega)\, d\Omega \tag{2.12}$$

Na equação (2.12), $O(\Pi)$ e $O(\Omega)$ representam as componentes cinética e conformacional da expressão clássica da observável O e as funções probabilidade $P(\Pi)$ e $P(\Omega)$ são definidas, respectivamente, como:

$$P(\Pi) = \frac{\exp(-\beta\, \Theta(\Pi))}{\int_{\Pi} \exp(-\beta\, \Theta(\Pi))\, d\Pi} = \frac{\exp(-\beta\, \Theta(\Pi))}{K(NVT)} \tag{2.13}$$

e

$$P(\Omega) = \frac{\exp(-\beta\, V(\Omega))}{\int_{\Pi} \exp(-\beta\, V(\Omega))\, d\Omega} = \frac{\exp(-\beta\, V(\Omega))}{Z(NVT)} \tag{2.14}$$

As amostragens utilizando uma função probabilidade do tipo $P(\Pi;\Omega)$ tendem a privilegiar as zonas do espaço de fases correspondentes às energias mais baixas. Quando se pretendem explorar zonas dificilmente acessíveis numa amostragem de Boltzmann, é necessário recorrer a metodologias alternativas. Valleau e Torrie [2] sugeriram um método que permite extrair médias de *ensemble* de Boltzmann a partir de amostragens efectuadas utilizando funções probabilidade de outros tipos. Neste método, a energia do sistema é alterada para,

$$E'(\Pi;\Omega) = \Theta(\Pi) + V'(\Omega) = \Theta(\Pi) + V(\Omega) + W(\Omega) = E(\Pi;\Omega) + W(\Omega) \ , \qquad (2.15)$$

onde $W(\Omega)$ é uma função peso cuja forma depende da aplicação particular. O formalismo de Valleau e Torrie pode ser derivado através dos seguintes passos:

Substituindo as equações (2.10) e (2.11) na equação (2.8) esta é modificada para:

$$\langle O \rangle = \frac{\int_{\Pi}\int_{\Omega} O(\Pi;\Omega)\exp[-\beta E(\Pi;\Omega)]d\Pi\, d\Omega}{\int_{\Pi}\int_{\Omega} \exp[-\beta E(\Pi;\Omega)]d\Pi\, d\Omega} \qquad (2.16)$$

Inserindo na equação 2.16 o factor unidade na forma de,

$$1 = \exp[-\beta W(\Omega)]\exp[\beta W(\Omega)] \ , \qquad (2.17)$$

obtém-se,

$$\langle O \rangle = \frac{\int_{\Pi}\int_{\Omega} O(\Pi;\Omega)\exp[-\beta E(\Pi;\Omega)]\exp[-\beta W(\Omega)]\exp[\beta W(\Omega)]d\Pi\, d\Omega}{\int_{\Pi}\int_{\Omega} \exp[-\beta E(\Pi;\Omega)]\exp[-\beta W(\Omega)]\exp[\beta W(\Omega)]d\Pi\, d\Omega} \ , \qquad (2.18)$$

ou:

$$\langle O \rangle = \frac{\int_{\Pi}\int_{\Omega} O(\Pi;\Omega)\exp[\beta W(\Omega)]\exp[-\beta E'(\Pi;\Omega)]\ d\Pi\, d\Omega}{\int_{\Pi}\int_{\Omega} \exp[\beta W(\Omega)]\exp[-\beta E'(\Pi;\Omega)]\ d\Pi\, d\Omega} \ , \qquad (2.19)$$

Finalmente, dividindo o numerador e o denominador da equação (2.19) por,

$$\int_{\Pi}\int_{\Omega} \exp[-\beta\, E'(\Pi;\Omega)]\, d\Pi\, d\Omega \ , \tag{2.20}$$

chega-se à equação,

$$\langle O \rangle = \frac{\int_{\Pi}\int_{\Omega} O(\Pi;\Omega)\exp[\beta\, W(\Omega)]\; P'(\Pi;\Omega)\; d\Pi\, d\Omega}{\int_{\Pi}\int_{\Omega} \exp[\beta\, W(\Omega)]\; P'(\Pi;\Omega)\; d\Pi\, d\Omega} \ , \tag{2.21}$$

onde,

$$P'(\Pi;\Omega) = \frac{\exp[-\beta\, E'(\Pi;\Omega)]}{\int_{\Pi}\int_{\Omega} \exp[-\beta\, E'(\Pi;\Omega)]\, d\Pi\, d\Omega} \tag{2.22}$$

Numa notação mais concisa, a equação (2.21) pode ser escrita como:

$$\langle O \rangle = \frac{\langle O(\Pi;\Omega)\exp[\beta\, W(\Omega)] \rangle_{E'}}{\langle \exp[\beta\, W(\Omega)] \rangle_{E'}} \tag{2.23}$$

Na equação (2.23), o subscrito E' indica que as correspondentes médias de *ensemble* são geradas com base na função probabilidade $P'(\Pi;\Omega)$ e na energia $E'(\Pi;\Omega)$.

Como $E(\Pi;\Omega)$ e $E'(\Pi;\Omega)$ apenas diferem na componente conformacional, a equação (2.23) pode ser reescrita como,

$$\langle O \rangle = \langle O \rangle_{\Pi} + \frac{\langle O(\Omega)\exp[\beta\, W(\Omega)] \rangle_{V'}}{\langle \exp[\beta\, W(\Omega)] \rangle_{V'}} \ , \tag{2.24}$$

onde o subscrito V' significa que as correspondentes médias de *ensemble* são geradas com base na função probabilidade P'(Ω) definida como:

$$P'(\Omega) = \frac{\exp[-\beta V'(\Omega)]}{\int_{\Pi}\int_{\Omega}\exp[-\beta V'(\Omega)]d\Pi\, d\Omega} \tag{2.25}$$

Habitualmente, a função peso apresenta a forma,

$$W(\Omega) = k_w (\Omega - \Omega_0)^2 , \tag{2.26}$$

onde k_w é uma constante de força efectiva. Neste caso, W(Ω) toma valores muito elevados para pontos do espaço afastados de Ω_0. Esta técnica é chamada de *umbrella sampling* por ser preferencialmente concentrada nas região de Ω_0, com uma extensão que depende de k_w.

As equações (2.13) e (2.14) sugerem que a função de partição pode ser expressa como um produto de uma componente cinética K(NVT) e de uma componente conformacional Z(NVT):

$$Q(NVT) = \frac{K(NVT)\, Z(NVT)}{h^{3N}\, I} \tag{2.27}$$

Em analogia, a energia de Helmholtz pode também ser decomposta em componentes cinética (C) e conformacional (F),

$$A = C + F , \tag{2.28}$$

onde

$$C = -k_B T \ln \frac{K(NVT)}{h^{3N}\, I} \tag{2.29}$$

e

$$F = -k_B T \ln Z(NVT) \tag{2.30}$$

Na metodologia semiclássica apresentada para o cálculo de observáveis (equação 2.12) e da energia de Helmholtz (equação 2.27), as respectivas componentes cinéticas podem, em geral, ser determinadas analiticamente. No entanto, o calculo das componentes conformacionais não pode ser efectuado deste modo pois implicaria a determinação prévia de Q(NVT), o que na esmagadora maioria dos casos é impossível. Neste contexto, os métodos de simulação molecular (designadamente o método da Dinâmica Molecular (MD) utilizado neste trabalho) desempenham um papel essencial para ultrapassar este tipo de dificuldades. Esta problemática será discutida posteriormente na secção 2.3.

2.2 - Modelos de energia potencial

Na formulação semiclássica da mecânica estatística a modelação da energia potencial desempenha um papel de enorme relevância. De facto, a exactidão no cálculo de propriedades termodinâmicas, espectroscópicas ou estruturais depende da possibilidade de encontrar uma função energia potencial de boa qualidade. Em termos ideais, esta função deve ter as seguinte características genéricas [1]:

Simplicidade - A função energia potencial deve ter a forma matemática mais simples possível para facilitar a implementação computacional dos métodos da mecânica estatística.

Generalidade - A função energia potencial deve poder ser usada para calcular diferentes propriedades.

Transferibilidade - A função energia potencial deve ser transferível entre diversas classes de moléculas.

Exactidão - A função energia potencial deve permitir reproduzir os valores experimentais das propriedades calculadas.

Habitualmente, não é possível conjugar de uma forma ideal estas quatro características. De facto, para aumentar a exactidão é necessário recorrer a funções energia potencial com formas matemáticas mais complexas [3-5], a transferibilidade entre as diversas classes de moléculas está longe de ter sido conseguida de uma forma inequívoca [5] e cada

parametrização reproduz preferencialmente certas propriedades em detrimento de outras. É pois necessário estabelecer compromissos entre os diversos critérios apresentados de forma a encontrar as melhores soluções para cada aplicação particular.

Ao contrário do formalismo quântico, os modelos habitualmente utilizados não incluem o tratamento explícito dos electrões. Este procedimento baseia-se na aproximação de Born-Oppenheimer que estabelece que os movimentos nucleares e electrónicos podem ser desacoplados e tratados separadamente. Recorre-se então ao conceito de superfície de energia potencial em que os núcleos se movem num campo de potencial médio correspondente à distribuição electrónica óptima. Neste contexto, o sistema molecular é encarado como um conjunto de cargas e massas pontuais mantido coeso por forças clássicas. Estas forças são descritas em termos de uma função energia potencial com a forma genérica,

$$V(\Omega) = V_{lig.}(\Omega) + V_{n/lig.}(\Omega) \ , \tag{2.31}$$

onde $V_{lig.}(\Omega)$ e $V_{n/lig.}(\Omega)$ representam as energias potenciais ligante e não ligante, respectivamente.

A energia potencial ligante inclui componentes associadas às coordenadas internas relevantes do sistema e a energia potencial não ligante inclui as interacções envolvendo átomos separados por três ou mais ligações [4,6]. As componentes da energia potencial são então parametrizadas com base numa combinação de cálculos quânticos e de dados espectroscópicos.

Com base neste formalismo, vários potenciais têm sido propostas para diferentes classes de moléculas. Para moléculas orgânicas, destacam-se os potenciais MM2 [7], MM3 [8-12], TRIPOS [13], DREIDING [14] e COSMIC [15]. Para macromoléculas, destacam-se os potenciais CHARMM [16-17], AMBER [18-19], ECEPP [20-24], OPLS [25], GROMOS [26] e CFF [27-28].

Neste trabalho, é utilizado o potencial *united atoms* do CHARMM22, em que apenas os hidrogénios polares e átomos pesados são tratados explicitamente [29]. Os hidrogénios não polares são apenas tidos em conta implicitamente para corrigir os parâmetros dos átomos pesados a que se encontram ligados. No CHARMM22, a energia potencial ligante inclui

componentes associadas ao elongamento de ligações ($V_s(\Omega)$), à deformação de ângulos de ligação ($V_b(\Omega)$) e de ângulos diedros impróprios ($V_{imp}(\Omega)$) e à torção de ângulos diedros ($V_{tor}(\Omega)$). A energia potencial não ligante inclui as componentes electrostática ($V_{es}(\Omega)$) e de van der Waals ($V_{vdW}(\Omega)$):

$$V_{lig}(\Omega) = V_s(\Omega) + V_b(\Omega) + V_{imp}(\Omega) + V_{tor}(\Omega) \tag{2.32}$$

$$V_{n/lig}(\Omega) = V_{es}(\Omega) + V_{vdW}(\Omega) \tag{2.33}$$

As diferentes componentes da energia potencial ligante são definidas como,

$$V_s(\Omega) = \frac{1}{2}\sum_{l} k_l (l - l_0)^2, \tag{2.34}$$

$$V_b(\Omega) = \frac{1}{2}\sum_{\theta} k_\theta (\theta - \theta_0)^2, \tag{2.35}$$

$$V_{imp}(\Omega) = \frac{1}{2}\sum_{\omega} k_\omega (\omega - \omega_0)^2, \tag{2.36}$$

e

$$V_{tors}(\Omega) = \frac{1}{2}\sum_{\phi} k_\phi \left[1 + \cos(n\phi - \gamma)\right], \tag{2.37}$$

onde k_l, k_θ, k_ω e k_ϕ são constantes de força e l_0, θ_0, ω_0 e γ são parâmetros de referência associados ao comprimento de ligação (l), ao ângulo de ligação (θ), ao ângulo diedro impróprio (ω) e ao ângulo diedro (ϕ), respectivamente. Na equação (2.37), n representa o número de mínimos conformacionais associados ao ângulo diedro (ϕ).

As componentes da energia potencial não ligante são dadas pela expressão de Coulomb

$$V_{es}(\Omega) = \sum_{X<Y}^{\{3\,lig.\}} \frac{q_i\, q_k}{\varepsilon\, r_{ik}}, \tag{2.38}$$

e pelo potencial de Lennard-Jones

$$V_{vdW}(\Omega) = \sum_{X<Y}^{\{3\,lig.\}} \left[\frac{A_{ik}}{r_{ik}^{12}} - \frac{B_{ik}}{r_{ik}^{6}} \right], \tag{2.39}$$

onde a notação {3 lig.} significa que os correspondentes somatórios são extensíveis a todos os pares de átomos separados por três ou mais ligações, q_i e q_k são respectivamente as cargas dos átomos i e k, r_{ik} é a respectiva distância internuclear, ε é a constante dieléctrica do meio e A_{ik} e B_{ik} são constantes características do par (i,k).

Combinando as equações (2.38) e (2.39), a energia potencial não ligante pode ser decomposta em componentes de interacção diatómicas,

$$V_{n/lig}(\Omega) = \sum_{i<k}^{\{3\,lig.\}} V_{ik}(r_{ik}) \ , \tag{2.40}$$

onde $V_{ik}(r_{ik})$ representa a energia de interacção entre os átomos i e k dada por:

$$V_{ik}(r_{ik}) = \frac{q_i\, q_k}{\varepsilon\, r_{ik}} + \frac{A_{ik}}{r_{ik}^{12}} - \frac{B_{ik}}{r_{ik}^{6}} \tag{2.41}$$

A determinação da conformação correspondente à energia potencial mínima pode ter algum interesse na descrição semiclássica de um sistema molecular. Isto pode ser efectuado recorrendo a métodos apropriados de minimização da energia [30]. Neste trabalho serão usados os métodos da *steepest descent* e do *gradiente conjugado*, que estão especialmente vocacionados para minimizar a energia em macromoléculas [30].

Em geral, o conhecimento do mínimo global da energia potencial não fornece grande informação relativamente a um sistema molecular. De facto, tanto as propriedades termodinâmicas como as estruturais dependem de todo o espaço conformacional acessível. No entanto, a determinação de valores médios de propriedades que dependam fundamentalmente das regiões vizinhas do mínimo global pode ser acelerada se esta conformação particular tiver sido previamente localizada.

2.3 - Cálculo de propriedades em simulações moleculares

2.3.1 - Cálculo de observáveis

O valor esperado de uma observável calculado pela equação (2.8) é obtido considerando um *ensemble* de microestados gerados de uma forma independente. Uma metodologia alternativa de abordar este problema consiste em utilizar um *ensemble* de microestados gerados de uma forma correlacionada. Neste contexto, o referido valor esperado pode ser calculado através da equação,

$$\langle O \rangle = \frac{1}{L} \sum_{j=1}^{L} O(\Pi_j;\Omega_j) \quad , \tag{2.42}$$

onde $O(\Pi_j;\Omega_j)$ representa o valor da observável O no microestado $(\Pi_j;\Omega_j)$ correspondente à réplica j do sistema molecular.

Embora a média de *ensemble* (2.42) só convirja rigorosamente para o valor esperado (2.8) no chamado limite termodinâmico $(L \to \infty)$, para a grande maioria das observáveis o equilíbrio estatístico é atingido para um número de réplicas desde que elas tenham sido geradas de acordo com a função probabilidade $P(\Pi;\Omega)$. Os métodos de simulação molecular, designadamente o método de Monte Carlo (MC) e o método da Dinâmica Molecular (MD), permitem gerar *ensembles* compatíveis com a referida função probabilidade, pelo que desempenham um papel da maior relevância no cálculo teórico de propriedades termodinâmicas.

Neste contexto, a determinação do valor esperado de uma observável, através de simulação molecular, é composta por duas fases: uma fase de equilibração e uma fase de

produção. Na primeira fase, tenta-se atingir o equilíbrio estatístico gerando um número de réplicas adequado. Na segunda fase, são calculadas as médias para a propriedade em estudo.

Considera-se que se atingiu o equilíbrio estatístico quando os valores de $O(\Pi_j;\Omega_j)$ passam a oscilar em torno de um valor estável que corresponde a $\langle O \rangle$. Nestas condições, a análise estatística da simulação pode ser efectuada pelo método dos blocos [31], que se baseia na divisão das L réplicas do *ensemble* em K blocos de dimensão n_K. O valor médio (O_s) da observável no bloco s é então calculado como:

$$O_s = \frac{1}{n_K} \sum_{j=s(n_K-1)+1}^{s \times n_K} O(\Pi_j;\Omega_j) \tag{2.43}$$

A equação (2.42) pode agora ser expressa em termos de médias de bloco:

$$\langle O \rangle = \frac{1}{K} \sum_{s=1}^{K} O_s \tag{2.44}$$

O erro estatístico da determinação do valor esperado $\langle O \rangle$ através da equação (2.44) é calculado como:

$$\sigma(O) = \left[\frac{1}{K\,(K-1)} \sum_{s=1}^{K} (O_s - \langle O \rangle)^2 \right]^{1/2} \tag{2.45}$$

Para que a equação (2.45) seja válida, deve escolher-se o comprimento de bloco n_K de forma a que os valores O_s não estejam correlacionados. Recorre-se então à chamada função de correlação que é definida como:

$$\rho(\nu) = \frac{\langle O(\Pi_\mu;\Omega_\mu)\, O(\Pi_{\mu+\nu};\Omega_{\mu+\nu}) \rangle}{\langle O^2 \rangle - \langle O \rangle^2}$$

$$= \frac{\frac{1}{L-\nu}\sum_{\mu=1}^{L-\nu}\left[O(\Pi_\mu;\Omega_\mu)-\langle O\rangle\right]\left[O(\Pi_{\mu+\nu};\Omega_{\mu+\nu})-\langle O\rangle\right]}{\langle O^2\rangle-\langle O\rangle^2} \qquad (2.46)$$

A função $\rho(\nu)$ permite estimar a correlação existente entre os sucessivos valores $O(\Pi_\mu;\Omega_\mu)$ e $O(\Pi_{\mu+\nu};\Omega_{\mu+\nu})$ espaçados de ν. Só é possível utilizar a equação (2.45) como boa estimativa de $\sigma(O)$ se a condição (2.47) for satisfeita:

$$\lim_{\nu \to \nu_0} \rho(\nu) = 0 \quad , \nu_0 << L \qquad (2.47)$$

Neste caso, o comprimento de correlação (n_K^{min}), que corresponde ao comprimento mínimo de bloco a partir do qual os valores O_s passam a estar descorrelacionados, é dado por:

$$n_K^{min} = \int_0^{\nu_0} \rho(\nu)\, d\nu \qquad (2.48)$$

O erro estatístico indica a precisão com que o cálculo do valor esperado da observável foi efectuado na simulação. A exactidão deste cálculo está mais directamente associada aos chamados erros sistemáticos que podem ser originados por uma grande variedade de factores, tais como o uso de um modelo de energia potencial inadequado ou deficientemente parametrizado, insuficiente equilibração e não inclusão de interacções coulômbicas de longo alcance.

2.3.1 - Cálculo de variações de energia de Helmholtz

Na grande maioria dos sistemas químicos não é possível determinar a energia de Helmholtz através de simulações moleculares. No entanto, foram propostos vários métodos para determinar, por esta via, diferenças de energia de Helmholtz expressas em termos de

médias de *ensemble* apropriadas [32-33]. Neste contexto, os métodos da perturbação termodinâmica [34] e da integração termodinâmica [35] são os mais usados.

De acordo com o método da perturbação termodinâmica (TP), a variação da energia de Helmholtz para uma reacção (R→P) é dada por:

$$\Delta A_{R\to P} = A_P - A_R = -k_B T \ln\left\langle \exp[-\beta\Delta E]\right\rangle_R \tag{2.49}$$

Na equação (2.49), ΔE é a diferença de energia entre os produtos (P) e os reagentes (R),

$$\Delta E = E_P(\Pi;\Omega) - E_R(\Pi;\Omega)\ , \tag{2.50}$$

e a notação $\left\langle ... \right\rangle_R$ indica uma média de *ensemble* gerada de acordo com a energia potencial dos reagentes ($V_R(\Omega)$) em condições NVT.

A variação da energia de Helmholtz conformacional para a mesma reacção pode ser calculada como,

$$\Delta F_{R\to P} = F_P - F_R = -k_B T \ln\left\langle \exp[-\beta\Delta V]\right\rangle_R\ , \tag{2.51}$$

onde ΔV é a diferença de energia potencial entre os produtos e os reagentes,

$$\Delta V = V_P(\Omega) - V_R(\Omega)\ , \tag{2.52}$$

A obtenção das quantidades $\Delta A_{R\to P}$ e $\Delta F_{R\to P}$ através de simulações moleculares depende dos valores das correspondentes perturbações associadas (ΔE e ΔV) . Na prática, só se conseguem obter boas convergências nestas simulações quando as espécies R e P são estruturalmente muito semelhantes. Quando as perturbações necessárias para transformar reagentes em produtos forem muito grandes é necessário recorrer à aproximação do parâmetro de acoplamento [32-33]. Esta aproximação consiste em definir uma série de estados híbridos,

entre os reagentes e os produtos, caracterizados pelo parâmetro de acoplamento (λ) e pelas energias total ($E_\lambda(\Pi;\Omega)$) e potencial ($V_\lambda(\Pi;\Omega)$):

$$E_\lambda(\Pi;\Omega) = (1-\lambda)\,E_R(\Pi;\Omega) + \lambda\,E_P(\Pi;\Omega) \tag{2.53}$$

$$V_\lambda(\Omega) = (1-\lambda)\,V_R(\Omega) + \lambda\,V_P(\Omega) \tag{2.54}$$

O parâmetro de acoplamento (λ) apresenta valores entre 0 (correspondente a uma situação em que existem apenas reagentes) e 1 (correspondente a uma situação em que existem apenas produtos).

Neste contexto, as variações de energia de Helmholtz total ($\Delta A_{\lambda_i\to\lambda_{i+1}}$) e conformacional ($\Delta F_{\lambda_i\to\lambda_{i+1}}$) entre os estados híbridos λ_i e λ_{i+1} são calculadas, respectivamente, como,

$$\Delta A_{\lambda_i\to\lambda_{i+1}} = A_{\lambda_{i+1}} - A_{\lambda_i} = -k_B T \ln\left\langle \exp\left[-\beta\Delta E_{\lambda_i\to\lambda_{i+1}}\right]\right\rangle_{\lambda_i}, \tag{2.55}$$

e

$$\Delta F_{\lambda_i\to\lambda_{i+1}} = F_{\lambda_{i+1}} - F_{\lambda_i} = -k_B T \ln\left\langle \exp\left[-\beta\Delta V_{\lambda_i\to\lambda_{i+1}}\right]\right\rangle_{\lambda_i}, \tag{2.56}$$

onde:

$$\Delta E_{\lambda_i\to\lambda_{i+1}} = E_{\lambda_{i+1}}(\Pi;\Omega) - E_{\lambda_i}(\Pi;\Omega) \tag{2.57}$$

e

$$\Delta V_{\lambda_i\to\lambda_{i+1}} = V_{\lambda_{i+1}}(\Omega) - V_{\lambda_i}(\Omega) \tag{2.58}$$

As médias de *ensemble* (2.55) e (2.56) são calculadas através de uma simulação utilizando o potencial híbrido ($V_{\lambda_i}(\Omega)$). É necessário um número total de M simulações (janelas) para fazer variar o parâmetro de acoplamento λ do seu valor inicial para o seu valor

final. As contribuições obtidas em cada janela são então adicionadas para obter as variações finais:

$$\Delta A_{R\to P} = A_P - A_R = \sum_{i=1}^{M} \Delta A_{\lambda_i \to \lambda_{i+1}} \tag{2.59}$$

$$\Delta F_{R\to P} = F_P - F_R = \sum_{i=1}^{M} \Delta F_{\lambda_i \to \lambda_{i+1}} \tag{2.60}$$

Os erros estatísticos associados a $\Delta A_{\lambda_i \to \lambda_{i+1}}$ e a $\Delta F_{\lambda_i \to \lambda_{i+1}}$ são calculados como,

$$\sigma(\Delta A_{\lambda_i \to \lambda_{i+1}}) = k_B T \frac{\sigma(W)}{W}\ , \tag{2.61}$$

e

$$\sigma(\Delta F_{\lambda_i \to \lambda_{i+1}}) = k_B T \frac{\sigma(S)}{S}\ , \tag{2.62}$$

onde W e S são, respectivamente, as médias de *ensemble* exponenciais nas equações (2.55) e (2.56) e $\sigma(W)$ e $\sigma(S)$ são os correspondentes erros estatísticos calculados pelo método dos blocos exposto na secção (2.3.1).

Os erros estatísticos associados às variações totais ($R \to P$) são então dados por:

$$\sigma(\Delta A_{R\to P}) = \left[\sum_{i=1}^{M} \sigma^2(\Delta A_{\lambda_i \to \lambda_{i+1}}) \right]^{1/2} \tag{2.63}$$

$$\sigma(\Delta F_{R\to P}) = \left[\sum_{i=1}^{M} \sigma^2(\Delta F_{\lambda_i \to \lambda_{i+1}}) \right]^{1/2} \tag{2.64}$$

O número de janelas pode ser significativamente reduzido se o método *double wide sampling* [36] for utilizado. De acordo com esta metodologia, a mesma simulação efectuada

para um dado estado híbrido λ_i é utilizada para calcular as variações de energia de Helmholtz em relações aos estados λ_{i+1} e λ_{i-1}:

$$\Delta A_{R\to P} = A_P - A_R = \sum_{i=1}^{M'} (\Delta A_{\lambda_i \to \lambda_{i+1}} - \Delta A_{\lambda_i \to \lambda_{i-1}}) \tag{2.65}$$

$$\Delta F_{R\to P} = F_P - F_R = \sum_{i=1}^{M'} (\Delta F_{\lambda_i \to \lambda_{i+1}} - \Delta F_{\lambda_i \to \lambda_{i-1}}) \tag{2.66}$$

Neste caso, os erros estatísticos totais são calculados como:

$$\sigma(\Delta A_{R\to P}) = \left[\sum_{i=1}^{M'} \left[\sigma^2(\Delta A_{\lambda_i \to \lambda_{i+1}}) + \sigma^2(\Delta A_{\lambda_i \to \lambda_{i-1}}) \right] \right]^{1/2} \tag{2.67}$$

$$\sigma(\Delta F_{R\to P}) = \left[\sum_{i=1}^{M'} \left[\sigma^2(\Delta F_{\lambda_i \to \lambda_{i+1}}) + \sigma^2(\Delta F_{\lambda_i \to \lambda_{i-1}}) \right] \right]^{1/2} \tag{2.68}$$

O método da integração termodinâmica (TI) é baseado nas seguintes equações:

$$\Delta A_{R\to P} = A_P - A_R = \int_0^1 \left\langle \frac{\partial E_\lambda(\Pi;\Omega)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{2.69}$$

e

$$\Delta F_{R\to P} = F_P - F_R = \int_0^1 \left\langle \frac{\partial V_\lambda(\Pi;\Omega)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{2.70}$$

Assumindo um acoplamento linear em λ para as energias total (equação 2.53) e potencial (equação 2.54), as equações anteriores podem ser simplificadas para:

$$\Delta A_{R\to P} = \int_0^1 \langle \Delta E \rangle_\lambda \, d\lambda \tag{2.71}$$

$$\Delta F_{R\to P} = \int_0^1 \langle \Delta V \rangle_\lambda \, d\lambda \tag{2.72}$$

Usando o método dos trapézios [37-38], os integrais (2.71) e (2.72) podem ser calculados, respectivamente, como:

$$\Delta A_{R\to P} \approx \sum_{i=1}^{G-1} \left\{ \left[\langle \Delta E \rangle_{\lambda_i} + \langle \Delta E \rangle_{\lambda_{i+1}} \right] \frac{[\lambda_{i+1} - \lambda_i]}{2} \right\} \tag{2.73}$$

e

$$\Delta F_{R\to P} \approx \sum_{i=1}^{G-1} \left\{ \left[\langle \Delta V \rangle_{\lambda_i} + \langle \Delta V \rangle_{\lambda_{i+1}} \right] \frac{[\lambda_{i+1} - \lambda_i]}{2} \right\} \tag{2.74}$$

É necessário um número total de G simulações (janelas) para fazer variar λ_i do seu valor inicial para o seu valor final. O erro estatístico (σ_i) associado a cada integração numérica i é dado por,

$$\sigma_i = \left[\sigma^2_{\lambda_{i+1}} + \sigma^2_{\lambda_i} \right]^{1/2} \frac{[\lambda_{i+1} - \lambda_i]}{2} , \tag{2.75}$$

onde $\sigma^2_{\lambda_{i+1}}$ e $\sigma^2_{\lambda_i}$ são, respectivamente, os erros estatísticos associados às médias de *ensemble* $\langle \Delta E \rangle_{\lambda_{i+1}}$ e $\langle \Delta E \rangle_{\lambda_i}$ (no caso da equação 2.73) e às de *ensemble* $\langle \Delta V \rangle_{\lambda_{i+1}}$ e $\langle \Delta V \rangle_{\lambda_i}$ (no caso da equação 2.74). Os erros totais associados às variações ($R \to P$) são então calculados como:

$$\sigma(\Delta X_{R\to P}) = \left[\sum_{i=1}^{M'} \sigma_i^2 \right]^{1/2} \quad ; X = A \text{ ou } F \tag{2.76}$$

A convergência das médias de *ensemble* (2.55), (2.56), (2.71) e (2.72) apresenta problemas específicos em duas regiões: sempre que estas quantidades mudam de sinal e nos *pontos extremos* ($\lambda \to 0$ ou $\lambda \to 1$).

Na primeira região, os problemas de convergência são originados pela instabilidade numérica associada aos valores muito pequenos (próximos de zero) das referidas médias de *ensemble*. Nestas circunstâncias, as fases de equilibração e de produção devem ser prolongadas para assegurar uma convergência adequada.

Na segunda região, podem ocorrer problemas de convergência se a implementação da perturbação implicar o aniquilamento de átomos. Neste caso, estes átomos têm energias potenciais, sentem forças que tendem para zero e teoricamente podem encontrar-se em qualquer ponto do espaço. Isto pode originar interacções repulsivas (do tipo electrostático ou de van der Waals) com outros átomos [32-33,39-41] e um grande aumento nas componentes ligantes harmónicas da energia potencial [42]. Nestas condições, as referidas médias de *ensemble* têm grandes variações e flutuações, o que obriga a prolongar as fases de equilibração e de produção e a diminuir o espaçamento entre as janelas para melhorar a sua convergência. Este problema é por vezes designado por *catástrofe quando* λ *tende para os pontos extremos* e várias metodologias alternativas têm sido propostas para o abordar [39-40,43-45]: os métodos Cross [43], Mezei e Beveridge [39] e Mezei [45] baseiam-se no escalamento não linear das componentes não ligantes da energia potencial. O método de Pearlman e Kollman [44] utiliza comprimentos de ligação de equilíbrio dependentes de λ. O método de Beuler *et al* [40] é implementado através da definição de uma esfera impenetrável para cada átomo que impede as colisões ou quase colisões com outros átomos. No entanto, estes métodos representam apenas soluções parciais para o problema e, em alguns casos, apresentam justificações físicas insuficientemente fundamentadas, pelo que nenhum deles foi adoptado universalmente.

Os dois formalismos (TP e TI) anteriormente apresentados são, em princípio, equivalentes devendo conduzir a resultados similares no cálculo de variações de energia de Helmholtz [32-33,46-47]. No entanto, cada um dos métodos apresenta problemas de convergência particulares que derivam da forma como é efectuada a respectiva implementação.

Dentro do formalismo TP, o calculo de variações de energia de Helmholtz envolve a determinação de médias de *ensemble* exponenciais de tipo (2.55). Estas quantidades

dependem simultaneamente dos estados híbridos λ_i e λ_{i+1}, mas são calculadas através de simulações utilizando a energia potencial $V_{\lambda_i}(\Omega)$. Se os mínimos de energia associados aos estados λ_i e λ_{i+1} estiverem localizados em regiões afastadas do espaço conformacional, o correspondente *ensemble* pode não conter conformações suficientes associadas aos valores mais baixos das variações de energia $\Delta E_{\lambda_i \rightarrow \lambda_{i+1}}$ [32-33,46]. Nestas circunstâncias, o número de janelas deve ser significativamente aumentado na implementação computacional deste método para obter a convergência adequada.

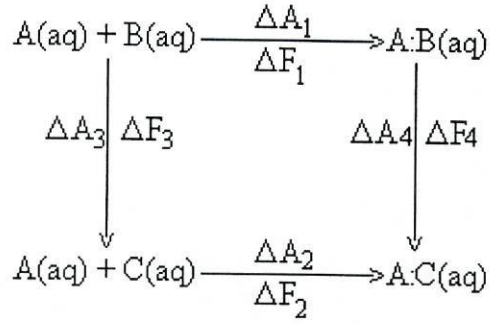
O método da integração termodinâmica, por não apresentar os problemas de ortogonalidade inerentes ao formalismo perturbacional, permite normalmente melhores eficiências estatísticas no cálculo de variações de energia de Helmholtz [37,46-47]. Isto traduz-se numa redução do número de janelas associadas a um cálculo deste tipo. No entanto, este método apresenta a desvantagem de envolver o cálculo numérico de integrais a partir de um número discreto de pontos. Além disso, o formalismo TI é bastante mais sensível aos problemas de convergência associados às transições para os *pontos extremos* do que o formalismo perturbacional [32-33].

Diversas metodologias alternativas, aos dois formalismos anteriormente apresentados, têm sido propostas nos últimos anos [48-63]. No entanto, estes métodos não serão aqui discutidos porque não forma utilizados neste trabalho.

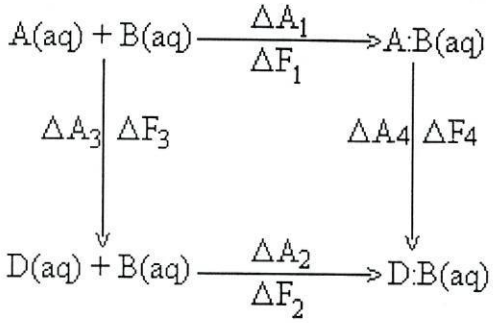
As reacções de associação estão entre as mais importantes que podem ser estudadas através de simulações moleculares [64-65]. Em solução aquosa, uma reacção deste tipo pode ser representada pela equação química,

$$A(aq) + B(aq) \xrightarrow[\Delta F]{\Delta A} A{:}B(aq) \quad , \tag{2.77}$$

mas a sua simulação é, em geral, impossível visto envolver a remoção de um grande número de moléculas de solvente das espécies A(aq) e B(aq) para formar o complexo A:B(aq) [66]. No entanto, é possível determinar diferenças de energia de Helmholtz de ligação recorrendo aos ciclos termodinâmicos apropriados:



(2.78)



(2.79)

$$\Delta\Delta A = \Delta A_2 - \Delta A_1 = \Delta A_4 - \Delta A_3 \tag{2.80}$$

$$\Delta\Delta F = \Delta F_2 - \Delta F_1 = \Delta F_4 - \Delta F_3 \tag{2.81}$$

A simulação dos processos 1 e 2 (processos físicos) é normalmente impossível pelos motivos anteriormente apresentados que se traduzem em tempos de computação extraordinariamente elevados. Pelo contrário, a simulação dos processos 3 e 4 (processos não físicos) é possível, utilizando a aproximação do parâmetro de acoplamento e os métodos da perturbação termodinâmica ou da integração termodinâmica, desde que as espécies envolvidas não sejam demasiado diferentes do ponto de vista estrutural.

Em ciclos termodinâmicos do tipo (2.78) e (2.79), é razoável considerar que a contribuição da componente cinética total seja próxima de zero [32], ou seja:

$$\Delta\Delta A \approx \Delta\Delta F \approx \Delta F_4 - \Delta F_3 \tag{2.82}$$

Neste contexto, pode ser utilizado o método de extrapolação proposto por Brooks III [67] para calcular diferenças de energia de Helmholtz, que se baseia numa aplicação particular do formalismo de Valleau e Torrie descrito na secção (2.1). Assumindo na equação (2.15),

$$V'(\Omega) = V_{\lambda_{i'}}(\Omega) \ , \tag{2.83}$$

$$V(\Omega) = V_{\lambda_i}(\Omega) \ , \tag{2.84}$$

e

$$W(\Omega) = \Delta V_{\lambda_i \to \lambda_{i'}} \ , \tag{2.85}$$

a média de *ensemble* $\langle \Delta V \rangle_{\lambda_i}$ pode ser calculada com base numa simulação efectuada com a energia potencial $V_{\lambda_{i'}}$:

$$\langle \Delta V \rangle_{\lambda_i} = \frac{\left\langle \Delta V \exp\left[\beta \, \Delta V_{\lambda_i \to \lambda_{i'}}\right] \right\rangle_{\lambda_{i'}}}{\left\langle \exp\left[\beta \, \Delta V_{\lambda_i \to \lambda_{i'}}\right] \right\rangle_{\lambda_{i'}}} \tag{2.86}$$

A expressão (2.86) constitui a base do método de extrapolação de Brooks III que pode ser usado para calcular diferenças de energia de Helmholtz dentro do formalismo geral da integração termodinâmica. De facto, considerando um número limitado de simulações efectuadas com as energias potenciais $V_{\lambda_{i'}}$, os valores esperados $\langle \Delta V \rangle_{\lambda_i}$ para os restantes estados híbridos λ_i podem ser obtidos pela equação (2.86).

2.4- Método da Dinâmica Molecular

O método da Dinâmica Molecular (MD) é uma técnica de simulação que permite determinar as propriedades dinâmicas e de equilíbrio para sistemas de N partículas cujos movimentos obedecem às leis da Mecânica Clássica.

As simulações MD são em muitos aspectos similares às experiências reais. Quando se efectua uma experiência no laboratório procede-se geralmente do seguinte modo:

(i) Prepara-se uma amostra do material que se pretende estudar.

(ii) Introduz-se na amostra um aparelho de medida adequado (termómetro, manómetro, etc.).

(iii) Mede-se a propriedade física em estudo durante o tempo suficiente para se obterem valores estáveis.

(iv) Determinam-se o valor médio da propriedade e correspondente erro estatístico.

(v) Se a propriedade estiver sujeita a ruído estatístico o tempo de medição tem de ser prolongado para se obter uma precisão aceitável.

Numa simulação MD procede-se exactamente da mesma forma:

(i) Selecciona-se um modelo (amostra) constituído por N partículas do sistema em estudo.

(ii) Resolvem-se as equações do movimento de Newton para o sistema e determinam-se os valores da propriedade em estudo para os microestados assim gerados.

(iii) A simulação é prolongada até que se obtenham valores estáveis da propriedade em estudo (fase de equilibração).

(iv) Quando o sistema está equilibrado, determinam-se o valor médio da propriedade e o erro estatístico associado (fase da produção).

(v) Os tempos de simulação necessários para obter uma boa precisão dependem da natureza da propriedade em estudo.

As principais fontes de erros na determinação de uma propriedade são também comuns às experiências laboratoriais e computacionais (designadamente amostras preparadas incorrectamente, tempos de medição demasiado curtos ou transformações irreversíveis durante a experiência).

No formalismo MD, a sequência de L microestados do sistema é descrita em termos das matrizes de trajectórias ($\mathbf{r}(i,t_j)$) e dos momentos ($\mathbf{p}(i,t_j)$),

$$(\Pi_j;\Omega_j) \Leftrightarrow (p_{1j}, p_{2j}, ..., p_{Nj}; X_{1j}, X_{2j}, ..., X_{Nj}) \Leftrightarrow (\mathbf{r}(i,t_j);\mathbf{p}(i,t_j)) \quad i=1,...,N\,;\, j=1,...,L\,, \quad (2.87)$$

onde,

$$\mathbf{p}(i,t_j) = p_{ij} \ , \tag{2.88}$$

e

$$\mathbf{r}(i,t_j) = X_{ij} \ , \tag{2.89}$$

podendo ser obtida por integração das equações do movimento de Newton,

$$\frac{\partial^2 \mathbf{r}(i,t_j)}{\partial t_j^2} = \frac{\mathbf{F}(i,t_j)}{m(i)} \ , \tag{2.90}$$

para uma série de L instantes de tempo separados por Δt (intervalo de tempo de integração).

Na equação (2.90), m(i) é a massa da partícula i e $\mathbf{F}(i,t_j)$ é a força que actua sobre esta partícula no instante de tempo t_j dada por:

$$\mathbf{F}(i,t_j) = -\frac{\partial V(\Omega_j)}{\partial \mathbf{r}(i,t_j)} = -\frac{\partial V(\mathbf{r}(i,t_j))}{\partial \mathbf{r}(i,t_j)} \tag{2.91}$$

Uma das dificuldades inerentes à implementação do método MD resulta das equações do movimento de Newton (2.90) não poderem ser integradas analiticamente para energias potenciais do tipo (2.31). Neste contexto, vários algoritmos aproximados têm sido propostos para efectuar estas integrações [68]. Os algoritmos mais utilizados para este efeito são os de *Runge-Kutta* [69], *Gear* [70], *Beeman* [71], *Verlet* [72] e *leap-frog* [68].

O algoritmo de *leap-frog*, utilizado neste trabalho, baseia-se na expansão das velocidades na forma de séries de Taylor em torno de cada ponto t_j,

$$\mathbf{v}(i,t_j + \Delta t/2) = \mathbf{v}(i,t_j) + \frac{\mathbf{F}(i,t_j)}{2m(i)}\Delta t + \frac{\partial \mathbf{F}(i,t_j)}{\partial t_j}\frac{(\Delta t/2)^2}{2} + O(\Delta t^3) \ , \tag{2.92}$$

$$\mathbf{v}(i,t_j - \Delta t/2) = \mathbf{v}(i,t_j) - \frac{\mathbf{F}(i,t_j)}{2m(i)}\Delta t + \frac{\partial \mathbf{F}(i,t_j)}{\partial t_j}\frac{(\Delta t/2)^2}{2} + O(\Delta t^3) \ , \tag{2.93}$$

onde $O(\Delta t^3)$ representa o erro resultante das expansões serem truncadas nos termos de 2ª ordem. Subtraindo a equação (2.93) à equação (2.92) obtém-se:

$$\mathbf{v}(i, t_j + \Delta t/2) = \mathbf{v}(i, t_j - \Delta t/2) + \frac{\mathbf{F}(i, t_j)}{m(i)} \Delta t + O(\Delta t^3) \ , \tag{2.94}$$

Usando o mesmo tipo de expansão para as posições obtém-se:

$$\mathbf{r}(i, t_j + \Delta t) = \mathbf{r}(i, t_j + \Delta t/2) + \mathbf{v}(i, t_j + \Delta t/2) \frac{\Delta t}{2} + \frac{\mathbf{F}(i, t_j)}{m(i)} \frac{(\Delta t/2)^2}{2} + O(\Delta t^3) \ , \tag{2.95}$$

e

$$\mathbf{r}(i, t_j) = \mathbf{r}(i, t_j + \Delta t/2) - \mathbf{v}(i, t_j + \Delta t/2) \frac{\Delta t}{2} + \frac{\mathbf{F}(i, t_j)}{m(i)} \frac{(\Delta t/2)^2}{2} + O(\Delta t^3) \tag{2.96}$$

Finalmente, subtraindo a equação (2.96) à equação (2.95) chega-se à equação (2.97),

$$\mathbf{r}(i, t_j + \Delta t) = \mathbf{r}(i, t_j) + \mathbf{v}(i, t_j + \Delta t/2) \, \Delta t \ + O(\Delta t^3) \ , \tag{2.97}$$

que permite determinar cada conformação a partir da imediatamente anterior.

O intervalo de tempo de integração (Δt) escolhido é limitado pelo valor da frequência mais alta (ν_{max}) associada aos movimentos que ocorrem no sistema:

$$\Delta t << \nu_{max}^{-1} \tag{2.98}$$

Habitualmente, este tipo de frequência está associado ao elongamento de ligações envolvendo pelo menos um átomo de hidrogénio. Neste caso, o valor típico de Δt é 0,5 fs. Consequentemente, os algoritmos MD com restrições, que permitem congelar os movimentos com frequências mais altas, apresentam grande interesse. Deste modo, é possível aumentar significativamente o valor de Δt mantendo uma boa descrição do sistema.

Neste contexto, o algoritmo SHAKE [73], utilizado neste trabalho, é especialmente apropriado para simulações em macromoléculas. De acordo com este algoritmo, as N_c restrições impostas aos movimentos do sistema apresentam a forma geral:

$$\sigma_n(\Omega_j) = 0 ; n = 1, \ldots, N_c \ , \tag{2.99}$$

com

$$\sigma_n(\Omega_j) = \mathbf{r}^2(n_1 \rightarrow n_2, t_j) - l_0^2(n) \tag{2.100}$$

Na equação (2.100), $l_0^2(n)$ representa o comprimento de ligação de equilíbrio associado à restrição n e $\mathbf{r}(n_1 \rightarrow n_2, t_j)$ representa o vector posição relativa dos átomos n_1 e n_2 envolvidos nesta ligação, na conformação Ω_j.

As equações de movimento são então alteradas introduzindo novas componentes $\mathbf{G}(i, t_j)$ que compensam as forças $\mathbf{F}(i,t_j)$ segundo as direcções em que os movimentos atómicos são restringidos:

$$\frac{\partial^2 \mathbf{r}(i, t_j)}{\partial t_j^2} = \frac{\mathbf{G}(i, t_j) + \mathbf{F}(i, t_j)}{m(i)} \ , \tag{2.101}$$

Estas equações têm de ser integradas de forma a satisfazer as N_c restrições impostas ao sistema. Isto pode ser efectuado utilizando o método dos multiplicadores de Lagrange [69] para construir as componentes $\mathbf{G}(i, t_j)$,

$$\mathbf{G}(i, t_j) = - \sum_{n=1}^{N_c} \lambda_n(t_j) \frac{\partial \sigma_n(\Omega_j)}{\partial \mathbf{r}(i, t_j)} \ , \tag{2.102}$$

onde os multiplicadores $\lambda_n(t_j)$ são determinados de forma a que as restrições (2.99) sejam satisfeitas para cada tempo t_j.

Dentro do algoritmo *leap-frog* as equações (2.94) e (2.97) são então alteradas, respectivamente, para,

$$\mathbf{v}(i,t_j+\Delta t/2)=\mathbf{v}(i,t_j-\Delta t/2)+\frac{\mathbf{F}(i,t_j)+\mathbf{G}(i,t_j)}{m(i)}\Delta t \quad , \tag{2.103}$$

e

$$\mathbf{r}(i,t_j+\Delta t)=\mathbf{r}(i,t_j)+\mathbf{v}(i,t_j+\Delta t/2)\,\Delta t \;=\mathbf{r}'(i,t_j+\Delta t)+\delta\mathbf{r}(i,t_j+\Delta t)\,, \tag{2.104}$$

onde $\mathbf{r}'(i,t_j+\Delta t)$ são as posições obtidas através de uma simulação MD convencional e $\delta\mathbf{r}(i,t_j+\Delta t)$ são as correcções efectuadas devido às restrições a que o sistema é sujeito, dadas por:

$$\delta\mathbf{r}(i,t_j+\Delta t)=\frac{\mathbf{G}(i,t_j)}{m(i)}(\Delta t)^2 \tag{2.105}$$

Substituindo a equação (2.102) na equação (2.105), obtém-se,

$$\delta\mathbf{r}(i,t_j+\Delta t)=\;-\,m^{-1}(i)\;(\Delta t)^2\sum_{n=1}^{N_c}\lambda_n(t_j)\frac{\partial\sigma_n(\Omega_j)}{\partial\mathbf{r}(i,t_j)}\quad , \tag{2.106}$$

ou, recorrendo à definição explícita de $\sigma_n(\Omega_j)$ (2.100),

$$\delta\mathbf{r}(i,t_j+\Delta t)=\;-\,2\,m^{-1}(i)\;(\Delta t)^2\sum_{n=1}^{\{(n_1,n_2)=(ik)\}}\lambda_n(t_j)\,r_{ik}(\Omega_j)\;, \tag{2.107}$$

onde o somatório é apenas extensível às restrições envolvendo o átomo i e $r_{ik}(\Omega_j)$ representa a distância entre os átomos i e k na conformação Ω_j. Além disso, as posições $\mathbf{r}(i,t_j+\Delta t)$ estão sujeitas às restrições (2.99), pelo que, recorrendo às equações (2.100) e (2.104), pode definir-se o seguinte sistema de equações adicionais:

$$\left|\mathbf{r}'(n_1, t_j + \Delta t) + \delta\mathbf{r}(n_1, t_j + \Delta t) - \mathbf{r}'(n_2, t_j + \Delta t) + \delta\mathbf{r}(n_2, t_j + \Delta t)\right|^2 = l_0^2(n) \qquad (2.108)$$

Desta forma, é possível determinar os multiplicadores de Lagrange $\lambda_n(t_j)$ para cada tempo t_j. No entanto, em sistemas moleculares os comprimentos de ligação restringidos estão muitas vezes correlacionados. Isto significa que o algoritmo SHAKE tem de ser resolvido iterativamente até que todas as restrições sejam simultaneamente satisfeitas dentro de um critério de precisão previamente estabelecido. A aplicação do algoritmo SHAKE a todas as ligações envolvendo pelo menos um átomo de hidrogénio permite diminuir o valor de Δt par 1,0 fs.

Nas simulações MD pretendem obter-se informações acerca das propriedades termodinâmicas de sistemas macroscópicos. No entanto, os recursos computacionais actualmente disponíveis apenas permitem simular amostras com um número limitado (na melhor das hipóteses da ordem das dezenas de milhar) de partículas. Nestas condições, não pode ser rigorosamente estabelecido que os *efeitos de superfície* sejam desprezáveis no cálculo de propriedades termodinâmicas do sistema. Estes efeitos resultam de, ao se simular um sistema molecular numa caixa fechada (normalmente cúbica), uma percentagem significativa das suas partículas se encontrarem na superfície [1]. Estas partículas estão sujeitas a uma energia potencial muito diferente das partículas que se encontram no interior da caixa.

Uma forma de ultrapassar este problema consiste em utilizar *condições periódicas fronteira* [1]. Dentro deste modelo, a caixa cúbica central, com raio r e contendo N partículas, é tratada como a célula original de um rede infinita de células idênticas (ver figura 2.1). Uma partícula na caixa central interactua não só com as outras partículas da mesma caixa mas também com as suas imagens. Durante a simulação, quando uma partícula muda de posição na caixa central o mesmo deve também ocorrer nas suas réplicas, para que a simetria da rede seja mantida. Em particular, quando uma partícula sair da caixa central por uma dos seus lados, uma das suas imagens deve entrar simultaneamente pelo lado oposto.

Num sistema do tipo descrito na figura (2.1), o número de termos a calcular é astronómico se não for utilizada nenhuma metodologia de truncagem da energia potencial não

ligante. As duas aproximações mais usadas para este efeito são as da *imagem mínima* e a do *corte esférico* [74].

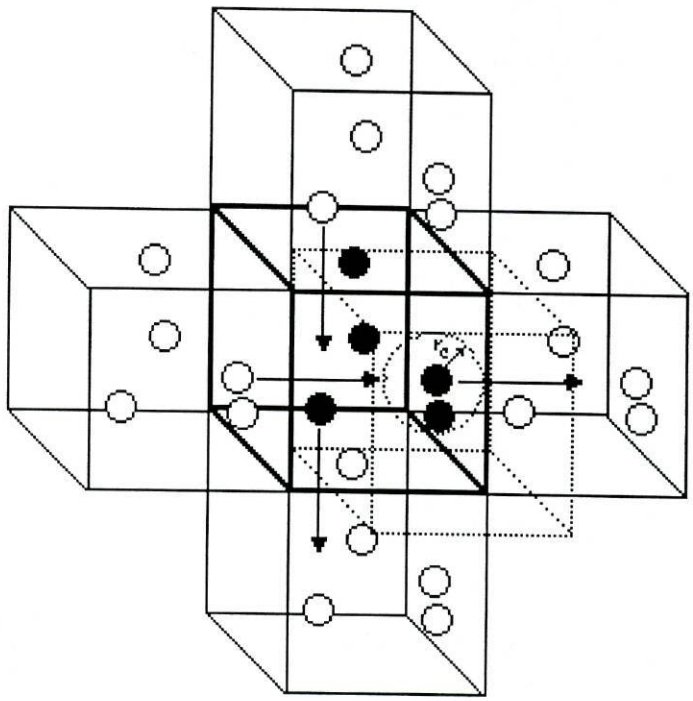


Figura 2.1. Representação esquemática das aproximações da *imagem mínima* e do *corte esférico* para uma rede periódica cúbica.

A aproximação da *imagem mínima* utiliza uma truncagem cúbica da energia potencial não ligante. Desta forma, uma partícula i só pode interactuar com as N-1 partículas mais próximas, quer elas se encontrem na caixa central ou nas respectivas réplicas. Estas partículas são seleccionadas centrando uma caixa cúbica, igual à caixa central, na partícula i (ver figura 2.1). Desta forma, não se permite que esta partícula interactue com as respectivas imagens, nem simultaneamente com outra partícula e uma das suas imagens.

A aproximação do *corte esférico* utiliza uma truncagem esférica da energia potencial não ligante. Isto é implementado centrando uma esfera, de raio r_c menor que r/2, na referida partícula i (ver figura 2.1).

O uso de funções energia potencial truncadas em simulações MD pode originar problemas sérios. Estes são originados por os algoritmos de integração das equações de movimento se basearem em expansões na forma de séries de Taylor das posições e velocidades das partículas. Isto implica que não podem existir descontinuidades na função

energia potencial. Uma forma de ultrapassar este problema consiste em introduzir correcções na forma da energia potencial de interacção entre duas partículas i e k (equação 2.41) de forma a restaurar a sua continuidade. No CHARMM22, isto é implementado através da seguinte equação,

$$V_{ik}^{corr}(r_{ik}) = V_{ik}(r_{ik})\, sw(r_{ik}, r_c, r_{on}) \ , \tag{2.109}$$

onde $V_{ik}^{corr}(r_{ik})$ é a energia potencial de interacção corrigida e $sw(r_{ik}, r_c, r_{on})$ é a função de correcção. Esta função é definida de forma a não ter qual quer efeito para distâncias curtas ($r_{ik} \leq r_{on}$), a fazer decair $V_{ik}(r_{ik})$ lentamente para zero para distâncias intermédias ($r_{on} < r_{ik} \leq r_c$) e a anular esta função para distâncias longas ($r_{ik} > r_c$):

$$sw(r_{ik}, r_c, r_{on}) = \begin{cases} 1 & ; \ r_{ik} \leq r_{on} \\ \dfrac{(r_c^2 - r_{ik}^2)^2 \left(r_c^2 + 2r_{ik}^2 - 3r_{on}^2\right)}{(r_c^2 - r_{on}^2)^3} & ; r_{on} < r_{ik} \leq r_c \\ 0 & ; \ r_{ik} > r_c \end{cases} \tag{2.110}$$

Em sistemas moleculares de grandes dimensões, a metodologia das condições periódicas fronteira, mesmo utilizando as aproximações da *imagem mínima* ou do *corte esférico*, apresenta grandes dificuldades de aplicação. Neste contexto, o método da *região activa* constitui uma alternativa muito utilizada [75-82]. Este método baseia-se na definição de uma região activa, de uma região fronteira e eventualmente de uma região inerte (ver figura 2.2).

A região activa tem habitualmente a forma de uma esfera cujo centro se situa na zona do sistema molecular que se pretende estudar. As partículas contidas nesta região são simuladas através de uma Dinâmica Molecular convencional.

A região fronteira tem habitualmente a forma de uma calote esférica exterior à região activa. Nesta região, os movimentos atómicos estão sujeitos a determinadas restrições.

A região inerte é constituída pelos restantes átomos e em muitos trabalhos [75,77-82] não é considerada na simulação do sistema. Pelo contrário, na formulação de Menziani *et al*

[76], utilizada neste trabalho, os átomos desta região são incluídos na determinação da energia potencial e das forças, mas são fixados nas suas posições iniciais.

O método da região activa é habitualmente aplicado em conjugação com a aproximação do *corte esférico*.

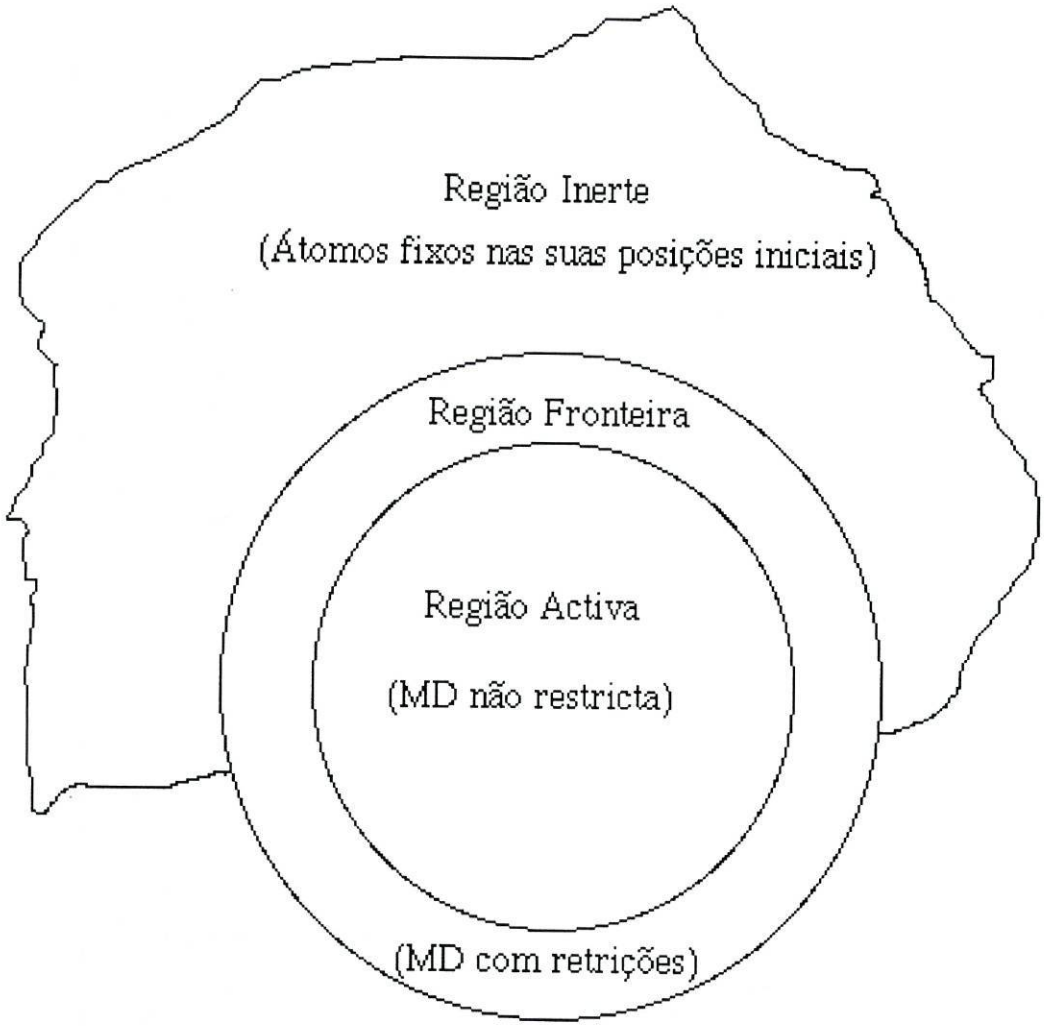


Figura 2.2. Representação esquemática do método da *região activa* segundo a formulação de Menziani *et al* [76].

O método da Dinâmica Molecular, tal como foi anteriormente apresentado, constitui uma técnica que permite simular a evolução com o tempo de um sistema clássico de N partículas contidas num volume V. Nestas condições, a energia total (E) constitui também uma constante de movimento. Consequentemente, numa simulação MD convencional são geradas médias de *ensemble* do tipo NVE. Este tipo de simulações apresenta o grande

inconveniente de não permitir efectuar comparações directas com os resultados de experiências laboratoriais, efectuadas em condições NVT ou NPT [83].

Diversa técnicas têm sido propostas para controlar a temperatura em simulações MD [83]. Todas elas se baseiam numa definição clássica da temperatura ($T(t_j)$),

$$\frac{1}{2} N_F \, k_B \, T(t_j) = \Theta(\Pi_j) = \frac{1}{2} \sum m(i) \, \mathbf{v}^2(i, t_j) \ , \tag{2.111}$$

onde N_F representa o número de graus de liberdade do sistema.

Nosé [84] introduziu os chamados métodos de *extended system*, que envolvem a introdução de um grau de liberdade adicional correspondente ao banho térmico. A temperatura é mantida constante através de trocas de calor entre o banho e o sistema molecular. Posteriormente, foram apresentadas diversas modificações e generalizações do método de Nosé [85-88].

Outro método frequentemente utilizado é o do *escalamento simples* [89], em que as velocidades são periodicamente multiplicadas por um factor que escala a energia cinética para o valor desejado. Por outro lado, no método das *colisões estocásticas* [90] as partículas do sistema colidem ocasionalmente com partículas virtuais de um banho térmico imaginário sendo as suas novas velocidades escolhidas aleatoriamente de uma distribuição de Maxwell-Boltzmann associada à temperatura pretendida.

Um modelo mais elaborado consiste em utilizar as equações de Langevin onde o acoplamento com o banho térmico é acompanhado pela introdução de termos estocásticos e de fricção nas equações de movimento [91]:

$$\frac{\partial^2 \mathbf{r}(i, t_j)}{\partial t_j^2} = \frac{\mathbf{F}(i, t_j)}{m(i)} - \gamma(i) \ \mathbf{v}(i, t_j) + \frac{\mathbf{R}(i, t_j)}{m(i)} \ , \tag{2.112}$$

Na equação (2.112), $\gamma(i)$ é o coeficiente de fricção da partícula i e $\mathbf{R}(i,t_j)$ é uma força estocástica gaussiana com média zero e variância

$$\left\langle \mathbf{R}(i,t_j)\,\mathbf{R}(k,t_j+\tau)\right\rangle = 6\,m(i)\,\gamma(i)\,k_B\,T_0\,\delta_{ik}\,\delta_{t_j t_{j+\tau}}\,, \tag{2.113}$$

onde T_0 representa a temperatura de referência que se pretende impor ao sistema.

O método de Berendsen *et al* [92], utilizado neste trabalho, constitui uma implementação particular do formalismo de Langevin. Neste método, considera-se que a constante fricção apresenta o mesmo valor (γ) para todas as partículas e, nestas circunstâncias, a derivada na energia cinética em ordem ao tempo é dada por:

$$\frac{\partial\Theta(\Pi_j)}{\partial t_j} = \sum_{i=1}^{N} \mathbf{F}(i,t_j)\,\mathbf{v}(i,t_j) + 2\gamma\left[\frac{3}{2}N\,k_B\,T_0 - \Theta(\Pi_j)\right] \tag{2.114}$$

O primeiro termo do lado direito da equação (2.114) está associado ao sistema molecular (é igual ao simétrico da derivada da energia potencial em ordem ao tempo) e o segundo termo representa o acoplamento global com o banho. Este termo adicional pode ser expresso em termos da temperatura:

$$\frac{\partial T(t_j)}{\partial t_j} = 2\gamma\left[T_0 - T(t_j)\right] \tag{2.115}$$

Nestas condições, as equações de movimento de Langevin são alteradas para:

$$\frac{\partial^2 \mathbf{r}(i,t_j)}{\partial t_j^2} = \frac{\mathbf{F}(i,t_j)}{m(i)} + \gamma\left[\frac{T_0}{T(t_j)} - 1\right]\mathbf{v}(i,t_j)\,, \tag{2.116}$$

O acoplamento global com o banho térmico implica pois um escalamento das velocidades em cada intervalo de tempo de integração,

$$\mathbf{v}(i,t_j) \rightarrow \alpha(t_j)\,\mathbf{v}(i,t_j)\,, \tag{2.117}$$

onde o coeficiente de escalamento $\alpha(t_j)$ é dado por:

$$\alpha(t_j)=\left[1+\frac{\Delta t}{\tau_T}\left(\frac{T_0}{T(t_j)}-1\right)\right]^{1/2} \quad ; \quad \tau_T=\frac{1}{2\gamma} \tag{2.117}$$

Na equação (2.117), τ_T é o tempo de relaxação do sistema e está associado à rapidez com que este reage aos estímulos exteriores. Quanto menor for τ_T (maior acoplamento ao banho térmico) maior será a rapidez com que a temperatura do sistema atinge o valor desejado T_0. No entanto, tempos de relaxação demasiado baixos (acoplamentos demasiado fortes) provocam perturbações no sistema e originam dificuldades na convergência das propriedades dinâmicas [92]. Um valor de compromisso aceitável corresponde a um tempo de relaxação de 0.1 ps [92].

Dentro do formalismo MD, diversas metodologias têm sido sugeridas nos últimos anos com o objectivo de simular melhor os sistema químicos e bioquímicos [93-102]. Neste contexto, o PTI, por ser relativamente pequeno e por apresentar uma estrutura cristalográfica de boa qualidade, tem sido utilizado como sistema modelo para comparação de métodos de simulação molecular [83,103-124].

Habitualmente, as simulações MD são efectuadas no espaço cartesiano. Nestas circunstâncias, a matriz $[\mathbf{r}(i,t_j)]$ pode ser substituída por três matrizes representando as trajectórias nas três direcções (x, y e z) do espaço,

$$[\mathbf{r}_x(i,j)],[\mathbf{r}_x(i,j)]\text{e}[\mathbf{r}_x(i,j)] \; ; \; i=1,\ldots,N \, , \, j=1,\ldots,L \; , \tag{2.119}$$

onde, por simplicidade, a notação t_j foi substituída pela notação j.

Em muitos casos, apresentar uma única estrutura que represente amostragem efectuada na simulação MD. Neste caso, a chamada estrutura média $(\overline{\mathbf{r}}_u(i), i=1,\ldots,N)$ é a escolha mais apropriada:

$$\bar{\mathbf{r}}_u(i) = \frac{\sum_{j=1}^{L} \mathbf{r}_u(i,j)}{L} \quad , \; i = 1, \ldots, N \tag{2.120}$$

Na equação (2.120) e no resto deste capítulo **u=x, y ou z.**

No entanto, a estrutura média não é suficiente para descrever adequadamente as trajectórias obtidas através duma simulação MD. As flutuações estruturais observadas, numa região {r.a.} do sistema molecular, ao longo de uma simulação deste tipo podem ser representadas pelo desvio quadrático médio ($rms(t_j)$) em relação a uma estrutura de referência:

$$rms(t_j) = \left\{ \frac{\sum_{i}^{\{r.a.\}} \left\{ \left[\mathbf{r}_x(i,j) - \mathbf{r}_x^{ref}(i)\right]^2 + \left[\mathbf{r}_y(i,j) - \mathbf{r}_y^{ref}(i)\right]^2 + \left[\mathbf{r}_z(i,j) - \mathbf{r}_z^{ref}(i)\right]^2 \right\}}{3\,N_{r.a.} - 1} \right\}^{1/2} \tag{2.121}$$

Na equação (2.121), {r.a.} pode representar todo o sistema molecular ou uma região particular que se pretenda estudar separadamente, $N_{r.a.}$ é o número de átomos na região {r.a.} e a estrutura de referência é escolhida tendo em conta o tipo de flutuações que se pretendem estudar. As estruturas média (equação 2.119) e de raios X (quando disponível) são as mais utilizadas para este efeito.

O desvio quadrático médio total ($\overline{rms}$) definido como,

$$\overline{rms} = \left\{ \frac{\sum_{j=1}^{L} \sum_{i}^{\{r.a.\}} \left\{ \left[\mathbf{r}_x(i,j) - \mathbf{r}_x^{ref}(i)\right]^2 + \left[\mathbf{r}_y(i,j) - \mathbf{r}_y^{ref}(i)\right]^2 + \left[\mathbf{r}_z(i,j) - \mathbf{r}_z^{ref}(i)\right]^2 \right\}}{3\,N_{r.a.} \times L - 1} \right\}^{1/2} , \tag{2.122}$$

permite avaliar as flutuações globais da região {r.a.}.

O método da Dinâmica Molecular é bastante mais eficiente que o método de Monte Carlo na amostragem de sistemas moleculares com muitos graus de liberdade internos [66]. Consequentemente, o método da Dinâmica Molecular foi utilizado em todas as simulações efectuadas neste trabalho.

2.5- Compressão de ficheiros de trajectórias obtidos por Dinâmica Molecular

Na implementação computacional de uma simulação MD as matrizes de trajectórias (2.119) são habitualmente guardadas em ficheiros binários, sendo cada coordenada representada por um número real de 4 *bytes*. Estes ficheiros são de grande dimensão e o seu armazenamento em disco provoca muitos problemas.

Os algoritmos conservativos de compressão clássicos, como os de Huffman [125] utilizado no utilitário *pack*, Huffman adaptado [126] utilizado no *compact*, LZW [127] utilizado no *compress* e LZ77 [128] utilizado no *gzip*, não permitem grande eficiência na compressão destes ficheiros. Consequentemente, algoritmos específicos não conservativos, que aumentem significativamente a eficiência de compressão preservando um alto grau de precisão, apresentam grande interesse para uma melhor abordagem deste problema.

Neste trabalho, é proposto um novo algoritmo específico chamado "*Byte Structure Variable Length Coding* (BS-VLC)", que apresenta eficiências de compressão significativamente melhores que os algoritmos conservativos clássicos mantendo uma boa precisão.

Uma compressão BS-VLC inclui os seguintes passos:

Pré processamento: Os dados iniciais (2.119) são convertidos em matrizes de trajectórias diferenciais definidas como,

$$\Delta \mathbf{r}_u(i,j) = \mathbf{r}_u(i,j) - \mathbf{r}_u^{ref}(i,j) \qquad (2.121)$$

onde $\mathbf{r}_u^{ref}(i,j)$ é a coordenada de referência associada a $\mathbf{r}_u(i,j)$.

Na equação (2.121) e no resto desta secção **u=x,y** ou **z**.

Neste tipo de metodologia, é necessário conservar L_0 conformações (conformações padrão $[r_u^{\Theta}(i,k)]$, i=1,...,N ; k=1,...,L_0) na sua forma original de forma a permitir a reconstrução das matrizes de trajectórias $[\mathbf{r}_u(i,j)]$ a partir das correspondentes matrizes diferenciais $[\Delta\mathbf{r}_u(i,j)]$, no processo de descompressão. O número de conformações padrão conservadas depende dos critérios usados na escolha das coordenadas de referência $\mathbf{r}_u^{ref}(i,j)$.

Quantização: A informação resultante do pré-processamento é posteriormente transformada nas matrizes de trajectórias diferenciais inteiras $[\mathbf{u}(i,j)]$, cujas componentes são obtidas convertendo as coordenadas diferenciais para inteiros através da multiplicação pelo factor de escala (Scale):

$$\mathbf{u}(i,j) = \Delta\mathbf{r}_u(i,j) \times \text{Scale} \tag{2.124}$$

Quanto maior for o factor de escala, maior será a precisão e menor será a eficiência do processo de compressão. Pelo contrário, uma diminuição do factor de escala aumentará a eficiência de compressão mas diminuirá a precisão. É importante encontrar um compromisso que permita compressões significativas com erros razoáveis.

Codificação de comprimento variável (VLC): Neste passo, as matrizes diferenciais inteiras $[\mathbf{u}(i,j)]$ são comprimidas usando uma codificação de comprimento variável, onde as trajectórias são subdivididas em partes (estruturas) que são codificadas separadamente.

Os números inteiros, componentes das matrizes diferenciais inteiras $[\mathbf{u}(i,j)]$, podem ser representados usando uma codificação implícita do sinal ou uma codificação explícita do sinal. A primeira alternativa corresponde à forma convencional de decompor números inteiros em *bytes*, ou seja 1 *bit* é reservado para representar o sinal e os restantes *bits* representam o respectivo valor absoluto. Na segunda alternativa, os *bits* de sinais são explicitamente codificados e posteriormente agrupados em *bytes* de sinais. Esta última representação pode ser mais favorável se os dados estiverem centrados em certas regiões.

Num computador de 32 *bits*, o número mínimo de *bytes* necessário para representar valores absolutos de números inteiros, depende dos intervalos em que eles se encontram:

$$
\begin{aligned}
&1\ byte \Leftrightarrow [0 : 2^8 - 1] \\
&2\ bytes \Leftrightarrow [2^8 : 2^{16} - 1] \\
&3\ bytes \Leftrightarrow [2^{16} : 2^{24} - 1] \\
&4\ bytes \Leftrightarrow [2^{24} : 2^{32} - 1]
\end{aligned}
\tag{2.125}
$$

No entanto, quando se usam números inteiros com sinal, metade do espaço tem de ser reservado para representar os números negativos. Consequentemente, pode ser efectuada a seguinte associação entre número mínimo de *bytes* e intervalos:

$$
\begin{aligned}
&1\ byte \Leftrightarrow [0 : 2^7 - 1] \cup [-1 : -2^7] \\
&2\ bytes \Leftrightarrow [2^7 : 2^{15} - 1] \cup [-2^7 - 1 : -2^{15}] \\
&3\ bytes \Leftrightarrow [2^{15} : 2^{23} - 1] \cup [-2^{15} - 1 : -2^{23}] \\
&4\ bytes \Leftrightarrow [2^{23} : 2^{31} - 1] \cup [-2^{23} - 1 : -2^{31}]
\end{aligned}
\tag{2.126}
$$

Por outro lado, os sinais podem ser codificados explicitamente permitindo que os valores absolutos sejam representados de acordo com a equação (2.125). Uma forma possível de implementar esta codificação consiste em subdividir os dados em grupos de oito elementos e associar aos seus sinais o valor 1 quando são positivos e o valor 0 quando são negativos,

Por exemplo: + - + + - - + + <=> 1 0 1 1 0 0 1 1 ,

cada grupo de 8 *bits* é então guardado num *byte* de sinais e a sequência completa de *bytes* num vector de sinais. A dimensão (nsbyte) deste vector é igual ao número de coordenadas a dividir por 8.

Associando as equações (2.125) e (2.126), o espaço acessível a números inteiros num computador de 32 *bits* pode ser subdividido em sete intervalos, aos quais estão associados números mínimos de *bytes* diferentes conforme a codificação de sinal efectuada. Toda esta informação é apresentada na tabela 2.1.

Tabela 2.1. Número mínimo de *bytes* utilizados na codificação de diferentes intervalos de inteiros.

Identificação do intervalo	Intervalo	Número mínimo de *bytes*	
		Codificação implícita do sinal	Codificação explícita do sinal
1	$[0:2^7-1]\cup[-1:-2^7]$	1	1
2	$[2^7:2^8-1]\cup[-2^7-1:-2^8+1]$	2	1
3	$[2^8:2^{15}-1]\cup[-2^8:-2^{15}]$	2	2
4	$[2^{15}:2^{16}-1]\cup[-2^{15}-1:-2^{16}+1]$	3	2
5	$[2^{16}:2^{23}-1]\cup[-2^{16}:-2^{23}]$	3	3
6	$[2^{23}:2^{24}-1]\cup[-2^{23}-1:-2^{24}+1]$	4	3
7	$[2^{24}:2^{32}-1]\cup[-2^{24}:-2^{31}]$	4	4

Através da análise da tabela 2.1, pode concluir-se que a codificação explícita do sinal é mais favorável quando os dados estão concentrados nos intervalos 2, 4 ou 6.

O algoritmo BS-VLC é baseado na possibilidade de um conjunto de *estruturas*, escolhidas de forma adequada, poder ser representado por um número de *bytes* inferior ao das matrizes de trajectórias originais. Neste contexto, as *estruturas* são definidas como partes das matrizes de trajectórias diferencias cujos elementos exibem algum tipo de correlação. Todos os elementos de uma estrutura S_k são então representados por um número de *bytes* (nsbyte(S_k)) necessário para codificar o seu elemento de maior valor absoluto. No entanto, é necessário reservar um *byte* extra para cada estrutura para indicar o número de *bytes* utilizado na sua representação.

Considerando uma matriz de trajectórias dividida L_0 conformações padrão, n_1 *estruturas* que utilizam uma codificação implícita do sinal e n_2 *estruturas* que utilizam uma codificação explícita do sinal, o tamanho do ficheiro comprimido em *bytes* (CFBL) é dado por:

$$CFBL=ovh+(n_1+n_2)+nsbyte+\sum_{k=1}^{n1} n(S_k)\times nbmax1(S_k)+\sum_{m=1}^{n2} n(S_m)\times nbmax1(S_m)+ns_0 \quad (2.128)$$

Na equação (2.128), ovh representa a quantidade de informação necessária para descrever a forma como o ficheiro de trajectórias foi comprimido. Esta informação é fundamental para o processo de descompressão, mas representa uma parte insignificante (menor do que 0.01%) de CFBL. Na mesma equação, $n(S_k)$ e $n(S_m)$ são o número de elementos das *estruturas* S_k e S_m, respectivamente; $nbmax1(S_k)$ é o número de *bytes* necessário para representar o elemento da *estrutura* S_k com maior valor absoluto determinado usando a terceira coluna da tabela 2.1; $nbmax2(S_j)$ é o número de *bytes* necessário para representar o elemento da *estrutura* S_m com maior valor absoluto determinado usando a quarta coluna da tabela 2.1, ns_0 é o número de *bytes* necessário para guardar as conformações padrão e nsbyte é o número total de elementos representados com codificação explícita do sinal a dividir por 8:

$$ns_0 = 4 \times 3N \times L_0 \tag{2.129}$$

$$nsbyte = \sum_{m=1}^{n2} n(S_m)/8 \tag{2.130}$$

O algoritmo BS-VLC considera dois tipos de *estruturas*: *estruturas* associadas a correlações temporais (blocos e conformações dentro de um bloco) e *estruturas* associadas a correlações espaciais (átomos dentro de um bloco e coordenadas cartesianas atómicas dentro de um bloco).

O bloco é um conjunto sequencial de conformações e representa a *estrutura* de maior dimensão considerada no algoritmo BS-VLC.

Utilizando uma codificação implícita do sinal, o tamanho de um bloco B_k em *bytes* ($BBL(B_k)$) é dado por,

$$BBL(B_k) = 3N \times nconf(B_k) \times nbmax1(B_k)\,, \tag{2.131}$$

onde $nconf(B_k)$ é o número de conformações do bloco B_k.

Utilizando uma codificação implícita do sinal, $BBL(B_i)$ é calculado como,

$$BBL(B_k)= nsbyte(B_k) + 3N \times nconf(B_k) \times nbmax2(B_k)\,, \tag{2.132}$$

onde nsbyte (B_k) é dado por,

$$nsbyte(B_k)=nconf(B_k) \times 3N / 8 \tag{2.133}$$

A conformação dentro de um bloco é a *estrutura* temporal de menor dimensão considerada no algoritmo BS-VLC. Neste caso, BBL(B_k) é dado por,

$$BBL(B_k)= nconf(B_k) + 3N \sum_{j=1}^{nconf(B_k)} nbmax(conf(j))\,, \tag{2.134}$$

se for utilizada uma codificação implícita do sinal e por

$$BBL(B_k)= nconf(B_k) + nsbyte(B_k) + 3N \sum_{j=1}^{nconf(B_k)} nbmax2(conf(j)) \tag{2.135}$$

se for utilizada uma codificação explícita do sinal.

Nas equações (2.134) e (2.135), conf(j) representa a conformação número j de um bloco B_k.

Numa trajectória MD é também possível estabelecer correlações espaciais. A *estrutura* átomo dentro de um bloco reflecte este tipo de correlação. Por exemplo, numa proteína as coordenadas diferenciais dos átomos das cadeias laterais são normalmente maiores que as dos átomos da cadeia principal.

Para esta *estrutura*, BBL(B_k) é calculado como

$$BBL(B_k)= N + 3\ nconf(B_k) \sum_{i=1}^{N} nbmax1(at(i))\,, \tag{2.136}$$

se for utilizada uma codificação implícita do sinal e por

$$BBL(B_k)= N + nsbyte(B_k) + 3\ nconf(B_k) \sum_{i=1}^{N} nbmax2(at(i)) \qquad (2.137)$$

se for utilizada uma codificação explícita do sinal.

Nas equações (2.136) e (2.137), at(i) representa as posições do átomo i dentro do bloco B_k.

As coordenadas cartesianas dentro de um bloco constituem uma subdivisão da *estrutura* anterior e são a escolha apropriada quando a energia potencial exibe algum tipo de anisotropia. $BBL(B_k)$ é agora calculado pela equação

$$BBL(B_k)=3N+nconf(B_k)\left\{ \sum_{i=1}^{N} nbmax1(x(i)) + \sum_{i=1}^{N} nbmax1(y(i)) + \sum_{i=1}^{N} nbmax1(z(i)) \right\} \quad (2.138)$$

se for utilizada uma codificação implícita do sinal e por

$$BBL(B_k)= 3N +nsbyte(B_k) + \\ + nconf(B_k) \left\{ \sum_{i=1}^{N} nbmax2(x(i)) + \sum_{i=1}^{N} nbmax2(y(i)) + \sum_{i=1}^{N} nbmax2(z(i)) \right\} \qquad (2.139)$$

se for utilizada uma codificação explícita do sinal.

Nas equações (2.138) e (2.139), x(i), y(i) e z(i) representam, respectivamente, o conjunto de coordenadas cartesianas x, y e z assumidas pelo átomo i no bloco B_k.

Uma descompressão dos ficheiros obtidos pelo algoritmo BS-VLC incluí os seguintes passos:

Descodificação de comprimento variável (VLD): Neste passo, o ficheiro comprimido é descodificado reconstruindo as matrizes diferenciais inteiras [**u**(i,j)].

Quantização inversa: As matrizes [**u**(i,j)] são usadas para determinar as matrizes diferenciais descomprimidas $[\Delta \tilde{\mathbf{r}}_u (i,j)]$:

$$\Delta \tilde{\mathbf{r}}_u(i,j) = \mathbf{u}(i,j)/\text{Scale} \tag{2.140}$$

Os elementos das matrizes diferenciais descomprimidas diferem dos elementos das matrizes diferenciais originais por uma quantidade $e(\Delta \mathbf{r}_u(i,j))$:

$$\Delta \tilde{\mathbf{r}}_u(i,j) = \Delta \mathbf{r}_u(i,j) + e(\Delta \mathbf{r}_u(i,j)) \tag{2.141}$$

O valor do erro $e(\Delta \mathbf{r}_u(i,j)$ depende do factor de escala utilizado.

Pós-processamento: As matrizes de trajectórias descomprimidas $[\tilde{\mathbf{r}}_u(i,j)]$ são então determinadas a partir das matrizes diferenciais descomprimidas $[\Delta \tilde{\mathbf{r}}_u(i,j)]$ e das coordenadas de referência descomprimidas $\tilde{\mathbf{r}}_u^{ref}(i,j)$:

$$\tilde{\mathbf{r}}_u(i,j) = \tilde{\mathbf{r}}_u^{ref}(i,j) + \Delta \tilde{\mathbf{r}}_u(i,j) \tag{2.142}$$

Os elementos das matrizes de trajectórias descomprimidas diferem dos elementos das matrizes de trajectórias originais por uma quantidade $e(\mathbf{r}_u(i,j))$:

$$\tilde{\mathbf{r}}_u(i,j) = \mathbf{r}_u(i,j) + e(\mathbf{r}_u(i,j)) \tag{2.143}$$

As características dos erros $e(\mathbf{r}_u(i,j))$ dependem do tipo de coordenadas de referência $\mathbf{r}_u^{ref}(i,j)$ utilizadas.

Podem utilizar-se vários critérios na escolha $\mathbf{r}_u^{ref}(i,j)$. As coordenadas imediatamente anteriores às coordenadas presentes $\mathbf{r}_u(i,j)$ são a escolha mais óbvia:

$$\mathbf{r}_u^{ref}(i,j) = \mathbf{r}_u(i,j-1) \tag{2.144}$$

Neste contexto, a conformação padrão é a primeira conformação $\mathbf{r}_u(i,1)$ das trajectórias e o número de *bytes* (ns_0) necessário para a guardar é dado por:

$$ns_0 = 4 \times 3na \tag{2.145}$$

A equação (2.142) é alterada para:

$$\tilde{\mathbf{r}}_u(i,j) = \tilde{\mathbf{r}}_u(i,j-1) + \Delta\tilde{\mathbf{r}}_u(i,j) \tag{2.146}$$

Substituindo a equação (2.141) na equação (2.146) obtém-se:

$$\tilde{\mathbf{r}}_u(i,j) = \tilde{\mathbf{r}}_u(i,j-1) + \Delta\mathbf{r}_u(i,j) + e(\Delta\mathbf{r}_u(i,j)) \tag{2.147}$$

Repetindo esta substituição para todos os valores de $\tilde{\mathbf{r}}_u(i,j-1)$ anteriores chega-se à equação,

$$\tilde{\mathbf{r}}_u(i,j) = \mathbf{r}_u(i,1) + \sum_{n=2}^{j} \Delta\mathbf{r}_u(i,n) + \sum_{n=2}^{j} e\left(\Delta\mathbf{r}_u(i,n)\right), \tag{2.148}$$

que pode ser reescrita como:

$$\tilde{\mathbf{r}}_u(i,j) = \mathbf{r}_u(i,j) + \sum_{n=2}^{j} e\left(\Delta\mathbf{r}_u(i,n)\right) \tag{2.149}$$

Comparando as equações (2.149) e (2.143), os erros $e(\mathbf{r}_u(i,j)$ podem ser calculados como:

$$e(\mathbf{r}_u(i,j)) = \sum_{n=2}^{j} e\left(\Delta\mathbf{r}_u(i,n)\right) \tag{2.150}$$

Se os sucessivos valores $e(\Delta\mathbf{r}_u(i,j))$ não se compensarem mutuamente, isto originará erros $e(\mathbf{r}_u(i,j))$ cumulativos.

Uma metodologia alternativa, que previne parcialmente os inconvenientes da escolha anterior, consiste em truncar a acumulação do erro no fim de cada bloco. Nesta situação, é necessário guardar, não só a primeira conformação, mas também as primeiras conformações dos restantes blocos. Consequentemente, os erros são ainda cumulativos dentro de cada bloco mas tornam-se nulos na transição entre a última conformação de um bloco e a primeira conformação do bloco seguinte. As eficiências de compressão são também ligeiramente reduzidas por este procedimento,

$$ns_0 = 4 \times 3N \times nblo\ , \tag{2.151}$$

onde nblo é o número total de blocos usado.

As coordenadas atómicas descomprimidas são contudo a escolha mais apropriada para coordenadas de referência:

$$\mathbf{r}_u^{ref}(i,j) = \tilde{\mathbf{r}}_u(i,j-1) \tag{2.152}$$

Neste caso, a equação (2.123) é alterada para,

$$\Delta\mathbf{r}_u(i,j) = \mathbf{r}_u(i,j) - \tilde{\mathbf{r}}_u(i,j-1)\ , \tag{2.153}$$

e a equação (2.141) pode ser reescrita como,

$$\Delta\tilde{\mathbf{r}}_u(i,j) = \mathbf{r}_u(i,j) - \tilde{\mathbf{r}}_u(i,j-1) + e(\Delta\mathbf{r}_u(i,j)) \tag{2.154}$$

A equação (2.154) pode ser rearranjada como,

$$\tilde{\mathbf{r}}_u(i,j-1) + \Delta\tilde{\mathbf{r}}_u(i,j) = \mathbf{r}_u(i,j) + e(\Delta\mathbf{r}_u(i,j)) \tag{2.155}$$

ou

$$\tilde{\mathbf{r}}_u(i,j) = \mathbf{r}_u(i,j) + e(\Delta\mathbf{r}_u(i,j)) \tag{2.156}$$

Comparando as equações (2.156) e (2.143), pode concluir-se que os erros $e(\mathbf{r}_u(i,j))$ podem ser calculados como:

$$e(\mathbf{r}_u(i,j)) = e(\Delta\mathbf{r}_u(i,j)) \tag{2.157}$$

Nesta aproximação, os erros tornam-se não cumulativos preservando a eficiência de compressão obtida pela primeira escolha aqui apresentada para as coordenadas de referência. Esta formulação corresponde ao esquema clássico DPCM [129] (Differential Pulse Code Modulation) que é a base de vários algoritmos de compressão audio, imagem e vídeo como o JPEG [130] e o MPEG [131].

A melhor maneira de avaliar a natureza cumulativa dos erros associados aos processos de compressão/descompressão é representar o desvio quadrático médio ($rms(t_j)$) entre as coordenadas originais $\mathbf{r}_u(i,j)$ e descomprimidas $\tilde{\mathbf{r}}_u(i,j)$,

$$rms(t_j) = \left\{ \frac{\sum_{i=1}^{N} \left\{ \left[\mathbf{r}_x(i,j) - \tilde{\mathbf{r}}_x(i,j)\right]^2 + \left[\mathbf{r}_y(i,j) - \tilde{\mathbf{r}}_y(i,j)\right]^2 + \left[\mathbf{r}_z(i,j) - \tilde{\mathbf{r}}_z(i,j)\right]^2 \right\}}{3N-1} \right\}^{1/2}, \tag{2.158}$$

em função do tempo de simulação.

Neste contexto, o desvio quadrático médio total,

$$\overline{rms} = \left\{ \frac{\sum_{j=1}^{L}\sum_{i=1}^{N} \left\{ \left[\mathbf{r}_x(i,j) - \tilde{\mathbf{r}}_x(i,j)\right]^2 + \left[\mathbf{r}_y(i,j) - \tilde{\mathbf{r}}_y(i,j)\right]^2 + \left[\mathbf{r}_z(i,j) - \tilde{\mathbf{r}}_z(i,j)\right]^2 \right\}}{3N \times L-1} \right\}^{1/2}, \tag{2.159}$$

permite uma avaliação da precisão global dos processos de compressão/descompressão.

A implementação computacional do algoritmo BS-VLC utiliza o seguinte procedimento para a compressão de um ficheiro de trajectórias:

Pré-processamento

1) Ler os dados iniciais $[\mathbf{r}_u(i,j)]$.

2) Escolher o número de blocos (nblo).

3) Seleccionar as coordenadas de referência.

4) Guardar as conformações padrão $[r_u^{\Theta}(i,k)]$.

5) Determinar as matrizes de trajectórias diferenciais $[\Delta\mathbf{r}_u(i,j)]$.

Quantização

1) Escolher o factor de escala (Scale).

2) Calcular as matrizes diferenciais inteiras $[\mathbf{u}(i,j)]$.

VLC

Para cada bloco:

1) Escolher a estrutura (bloco completo, conformação dentro de um bloco, átomo dentro de um bloco ou coordenadas cartesianas atómicas dentro de um bloco) e a codificação (explícita ou implícita) do sinal que conduzam a uma melhor eficiência de compressão do bloco (equações 2.131 a 2.139).

2) Comprimir o bloco usando a estrutura e a codificação do sinal escolhidas em 1).

O tamanho do ficheiro comprimido em *bytes* é dado por:

$$CFBL = ovh + \sum_{k=1}^{nblo} BBL(B_k) + ns_0 \qquad (2.160)$$

A descompressão de um ficheiro obtido através do algoritmo BS-VLC utiliza o seguinte procedimento:

VLD

Para cada bloco

1) Ler a estrutura (bloco completo, conformação dentro de um bloco, átomo dentro de um bloco ou coordenadas cartesianas atómicas dentro de um bloco) e a codificação (explícita ou implícita) do sinal usados na compressão do bloco.

2) Descomprimir o bloco usando a estrutura e a codificação do sinal escolhidas em 1) (as matrizes diferenciais inteiras **[u**(i,j)] são reconstruídas).

Quantização inversa

1) Ler o factor de escala (Scale).

2) Calcular as matrizes diferenciais descomprimidas $[\Delta \tilde{\mathbf{r}}_u (i,j)]$.

Pós-processamento

1) Ler as conformações padrão $[\mathbf{r}_u^{\Theta}(i,k)]$.

2) Calcular as matrizes de trajectórias descomprimidas $[\tilde{\mathbf{r}}_u (i,j)]$

Uma representação esquemática do algoritmo BS-VLC é apresentado na figura (2.3).

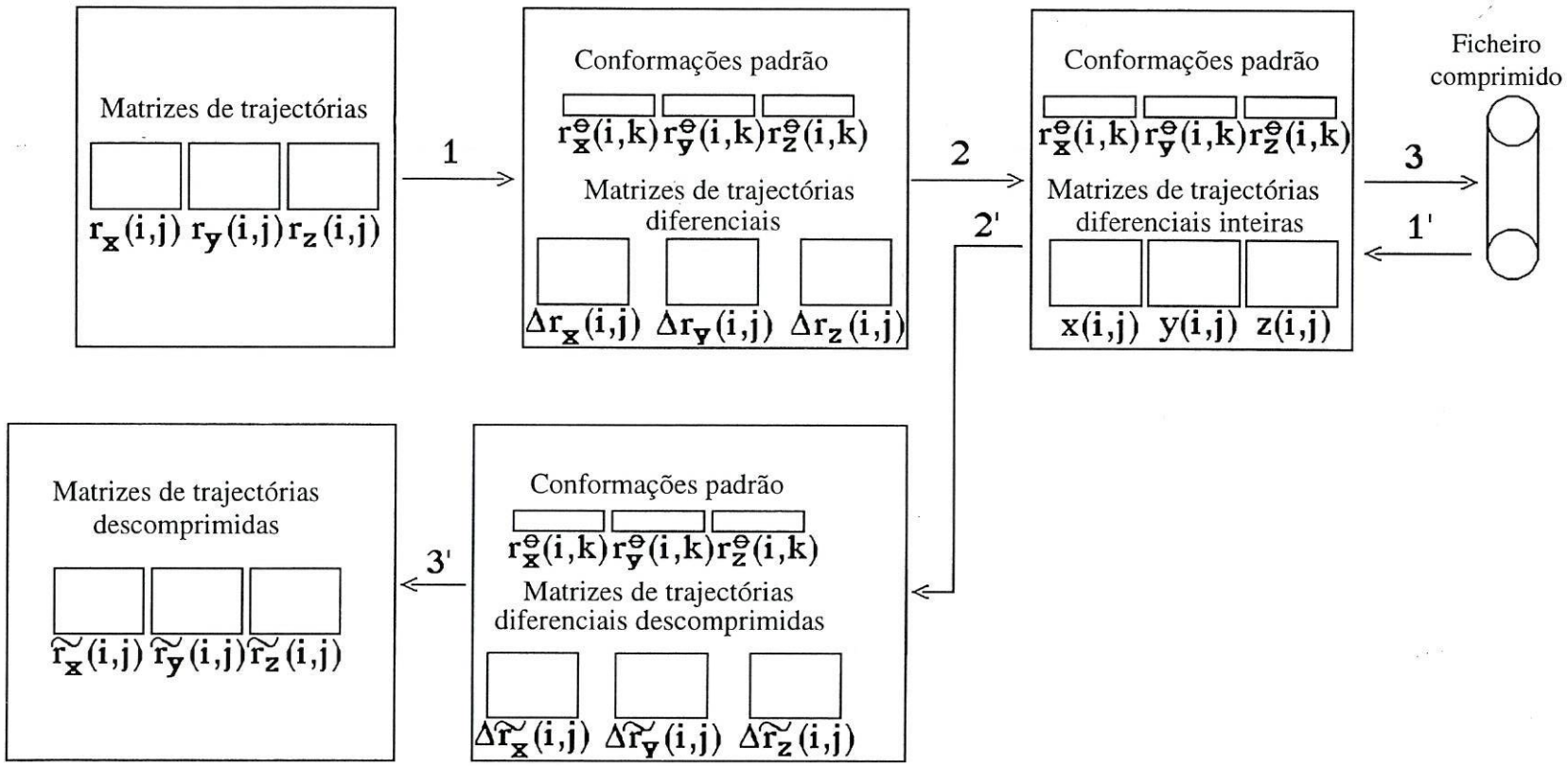


Figura 2.3. Representação esquemática do algoritmo BS-VLC. Devem ser considerados os seguintes passos: Compressão - (1) Pré-processamento; (2) Quantização; (3) Codificação de comprimento variável. Descompressão - (1') Descodificação de comprimento variável; (2') Quantização inversa; (3') Pós-processamento.

Referências

[1] D. Frenkel e B. Smith, *Understanding Molecular Simulations, From Algorithms to Applications*, Academic Press, San Diego (1996).

[2] J. P. Valleau e G. M. Torrie, em *Statistical Mechanics, Part A: Equilibrium Techniques*, Ed. B. J. Berne, Plenum Press, New York (1977).

[3] B. R. Gelin, em *Computer Simulations of Biomolecular Systems, Theoretical and Experimental Applications*, Vol. 2, Ed. W. F. van Gunsteren, P. K. Weiner e A. J. Wilkinson, ESCON, Leiden (1993).

[4] J. P. Bowen e N. L. Allinger, em *Reviews in Computational Chemistry*, Vol. 2, Ed. K. B. Lipkowitz e D. B. Boyd, VCH Publishers Inc., New York (1991).

[5] V. Dinur e A. T. Hagler, em *Reviews in Computational Chemistry*, Vol. 2, Ed. K. B. Lipkowitz e D. B. Boyd, VCH Publishers Inc., New York (1994).

[6] I. Petterson e T. Liljefors, em *Reviews in Computational Chemistry*, Vol. 9, Ed. K. B. Lipkowitz e D. B. Boyd, VCH Publishers Inc., New York (1996).

[7] N. L. Allinger, *J. Am. Chem. Soc.*, **99**, 8127 (1977).

[8] N. L. Allinger, Y. H. Yuh e J.-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551 (1989).

[9] J.-H. Lii e N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8566 (1989).

[10] J.-H. Lii e N. L. Allinger, *J. Am. Chem. Soc.*, **111**, 8576 (1989).

[11] N. L. Allinger, M. Rahman e J.-H. Lii, *J. Am. Chem. Soc.*, **112**, 8293 (1990).

[12] R. Schmitz e N. L. Allinger, *J. Am. Chem. Soc.*, **112**, 8307 (1990).

[13] M. Clark, R. D. Cramer III e N. van Optenbosch, *J. Comp. Chem.*, **10**, 982 (1989).

[14] S. L. Mayo, B. D. Olafson e W. A. Goddard III, *J. Phys. Chem.*, **94**, 8897 (1990).

[15] S. D. Mosley, R. J. Abraham, I. S. Haworth, D. E. Jackson, M.R. Saunders e J. G. Vister, *J. Comp.-Aided Mol. Design*, **5**, 475 (1991).

[16] B. R. Brooks, R. F. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan e M. Karplus, *J. Comp. Chem.*, **4**, 187 (1983).

[17] A. D. MacKerell Jr., J. Wiórkiwicz-Kuczera e M. Karplus, *J. Am. Chem. Soc.*, **117**, 11946 (1995).

[18] P. K. Weiner e P. A. Kollman, *J. Comp. Chem.*, **2**, 287 (1981).

[19] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta e P. K. Weiner, *J. Am. Chem. Soc.*, **106**, 765 (1984).

[20] F. A. Momany, L. M. Carrutheres, R. F. McGuire e H. A. Scheraga, *J. Phys. Chem.*, **78**, 1595 (1974).

[21] F. A. Momany, L. M. Carrutheres e H. A. Scheraga, *J. Phys. Chem.*, **78**, 1621 (1974).

[22] F. A. Momany, R. F. McGuire, A. W. Burgess e H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).

[23] G. Némethy, M. S. Pottle e H. A. Scheraga, *J. Phys. Chem.*, **87**, 1883 (1983).

[24] M.J. Sippe, G. Némethy e H. A. Scheraga, *J. Phys. Chem.*, **88**, 6231 (1984).

[25] W. L. Jorgensen e J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).

[26] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren e J. P. M. Postma, *Biopolymers*, **23**, 1513 (1984).

[27] J. R. Maple, M. J. Hwang, T. P. Stockfisch, U. Dinur, M. Waldman, C. S. Ewig e A. T. Hagler, *J. Comp. Chem.*, **15**, 62 (1994).

[28] M. J. Hwang, T. P. Stockfisch e A. T. Hagler, *J. Am. Chem. Soc.*, **116**, 2515 (1994).

[29] President and Fellows of Harvard College, CHARMM (versão 2.2), Molecular Simulations Inc., Burlington, USA.

30] T. Schlick, em *Reviews in Computational Chemistry*, Vol. 3, Ed. K. B. Lipkowitz e D. B. Boyd, VCH Publishers Inc., New York (1992).

[31] J. Cao e B. J. Berne, *J. Chem. Phys.*, **91**, 6359 (1989).

[32] D. L: Beveridge e F. M. DiCapua, em *Computer Simulations of Biomolecular Systems, Theoretical and Experimental Applications*, Vol. 1, Ed. W. F. van Gunsteren e P. K. Weiner, ESCON, Leiden (1989).

[33] D. L: Beveridge e F. M. DiCapua, *Annu. Rev. Biophys. and Biophys. Chem.* , **18**, 431 (1989).

[34] R. W. Zwanzig, *J. Chem. Phys.*, **22**, 1420 (1954).

[35] J. G. Kirkwood, *J. Chem. Phys.*, **3**, 300 (1935).

[36] W. L. Jorgensen e C. Ravimohan, *J. Chem. Phys.*, **83**, 3050 (1985).

[37] D. A. Pearlman, *J. Chem. Phys.*, **98**, 8946 (1993).

[38] D. A. Pearlman, *J. Comp. Chem.*, **15**, 105 (1994).

[39] M. Mezei e D. L. Beveridge, *Ann. N. Y. Acad. Sci.*, **482**, 1 (1986).

[40] T. C. Beuler, A. E. Mark, R. C. van Schaik, P. R. Gerber e W. F. van Gunsteren, *Chem. Phys. Lett.*, **222**, 529 (1994).

[41] G. A. Worth e W. G. Richards, *J. Am. Chem. Soc.*, **116**, 239 (1994).

[42] A. Melo e M. J. Ramos, *J. Peptide Res.*, **50**, 239 (1997).

[43] A. J. Cross, *Ann. N. Y. Acad. Sci.*, **482**, 89 (1986)

[44] D. A. Pearlman e P. A. Kollman, em *Computer Simulations of Biomolecular Systems, Theoretical and Experimental Applications*, Vol. 1, Ed. W. F. van Gunsteren e P. K. Weiner, ESCON, Leiden (1989).

[45] M. Mezei, *J. Comp. Chem.*, **13**, 651 (1992).

[46] D. A. Pearlman, *J. Phys. Chem.*, **98**, 1487 (1994).

[47] C. Chipot, P. A. Kollman e D. A. Pearlman, *J. Comp. Chem.*, **17**, 1112 (1996).

[48] X. J. Kong e C. L. Brooks, *J. Chem. Phys.*, **105**, 2414 (1996).

[49] D. K. Joneshertzog e W. L. Jorgensen, *J. Med. Chem.*, **40**, 1539 (1997).

[50] N. A. McDonald, H. A. Carlson e W. L. Jorgensen, *J. Org. Phys. Chem.*, **10**, 563 (1997).

[51] F. Fogolari e J. M. Briggs, *Chem. Phys. Lett.*, **281**, 135 (1997).

[52] I. Kolossvary, *J. Phys. Chem. A*, **101**, 9900 (1997).

[53] C. Bartels e M. Karplus, *J. Comp. Chem.*, **18**, 1450 (1997).

[54] M. Schaefer, M. Sommer e M. Karplus, *J. Phys. Chem.*, **101**, 9900 (1997).

[55] A. Amadei, M. E. F. Apol e H. J. C. Berendsen, *J. Phys. Chem.*, **106**, 1893 (1997).

[56] R. J. Radmer e P. A. Kollman, *J. Comp. Chem.*, **18**, 902 (1997).

[57] C. Jarque e B. Tidor, *J. Phys. Chem. B*, **101**, 9362 (1997).

[58] M. K. Gilson, J. A. Given e M. S. Head, *Chem. Biol.*, **4**, 87 (1997).

[59] N. Ota e A. T. Brünger, *Theor. Chem. Acc.*, **98**, 171 (1997).

[60] M. L. Lamb e W. L. Jorgensen, *Curr. Opin. Struct. Biol.*, **1**, 449 (1998).

[61] Z. Guo, C. L. Brooks III e K. Kong, *J. Phys. Chem. B*, **102**, 2032 (1998).

[62] C. Bartels e M. Karplus, *J. Phys. Chem. B*, **102**, 865 (1998).

[63] C. Bartels e M. Karplus, *J. Am. Chem. Soc.*, **120**, 1920 (1998).

[64] J. A. McCammon, *Curr. Op. Struct. Biol.*, **1**, 1 (1991).

[65] C. A. Reynolds, P. M. King e W. G. Richards, *Mol. Phys.*, **76**, 251 (1992).

[66] W. F. van Gunsteren, em *Computer Simulations of Biomolecular Systems, Theoretical and Experimental Applications*, Vol. 1, Ed. W. F. van Gunsteren e P. K. Weiner, ESCON, Leiden (1989).

[67] C. L. Brooks III, *J. Chem. Phys.*, **90**, 6680 (1986).

[68] W. F. van Gunsteren e H. J. C. Berendsen, *Angew. Chem., Int. Ed. Eng.*, **29**, 98 (1990).

[69] W. H. Press, B. P. Flannery, S. A. Teukolsky e W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge (1986).

[70] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood (1976).

[71] D. Beeman, *J. Comp. Phys.*, **20**, 130 (1976).

[72] L. Verlet, *Phys. Rev..*, **159**, 98 (1967).

[73] J. P. Ryckaert, G. Ciccoti e H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327 (1977).

[74] M. P. Allen e D. J. Tildesley, *Computer Simulations of Liquids*, Oxford University Press, Oxford (1990).

[75] A. T. Brünger, C. L. Brooks III e M. Karplus, *Proc. Natl. Acad. Sci. USA*, **82**, 8458 (1985).

[76] M. C. Menziani, C. A. Reynolds e W. G. Richards, *J. Chem. Soc., Chem. Commun.*, 853 (1989).

[77] J. Gao, K. Kuczera, B. Tidor e M. Karplus, *Science*, **244**, 1069 (1989).

[78] K. Kuczera, J. Gao, B. Tidor e M. Karplus, *Proc. Natl. Acad. Sci. USA*, **87**, 8481 (1990).

[79] M. Prevost, S. J. Wodak, B. Tidor e M. Karplus, *Proc. Natl. Acad. Sci. USA*, **88**, 10880 (1991).

[80] B. Tidor e M. Karplus, *Biochemistry*, **30**, 3217 (1991).

[81] B. Prod'hom e M. Karplus, *Protein Eng.*, **6**, 585 (1993).

[82] F. T. K. Lau e M. Karplus, *J. Mol. Biol.*, **236**, 1049 (1994).

[83] D. J. Tobias, G. J. Martyna e M. L. Klein, *J. Phys. Chem.*, **97**, 12959 (1993).

[84] S. Nosé, *J. Chem. Phys.*, **81**, 511 (1984).

[85] W. G. Hoover, *Phys. Rev. A*, **31**, 1695 (1985).

[86] S. Toxvaerd e O. H. Olson, *Ber. Bunsen-Ges. Phys. Chem.*, **94**, 274 (1990)

[87] G. J. Martyna, M. L. Klein e M. Tuckerman, *J. Chem. Phys.*, **97**, 2635 (1992).

[88] M. Tuckerman, B. Berne, G. J. Martyna e M. L. Klein, *J. Chem. Phys.*, **99**, 2796 (1993).

[89] L. V. Woodcock, *Chem. Phys. Lett.*, **10**, 257 (1971).

[90] H. C. Andersen, *J. Chem. Phys.*, **72**, 2384 (1980).

[91] T. Schneider e E. Stoll, *Phys. Rev. B*, **17**, 1302 (1978).

[92] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola e J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).

[93] I. G. Tironi, R. Sperb, P.E. Smith e W. F. van Gunsteren, *J. Chem. Phys.*, **102**, 5451 (1995).

[94] R. W. Dixon e P. A. Kollman, *J. Comp. Chem.*, **18**, 1632 (1997).

[95] B. J. Berne e J. E. Straub, *Curr. Opin. Struct. Biol.*, **7**, 181 (1997).

[96] T. Huber, A. E. Torda e W. F. van Gunsteren, *J. Phys. Chem. A*, **101**, 5926 (1997).

[97] W. F. van Gunsteren e A. E. Mark, *J. Chem. Phys.*, **108**, 6109 (1998).

[98] P. H. Hunenberger e W. F. van Gunsteren, *J. Phys. Chem.*, **108**, 6117 (1997).

[99] L. S. D. Caves, J. D. Evanseck e M. Karplus, *Prot. Sci.*, **7**, 649 (1998).

[100] T. Yamane, Y. Inoue e M. Sakurai, *Chem. Phys. Lett.*, **291**, 137 (1998).

[101] O. Edholm, P. I. Ohlson, M. L. Smith e J. Paula, *Chem. Phys. Lett.*, **291**, 501 (1998).

[102] P. A. Bopp, A. Kohlmeyer e E. Spohr, *Electrochimica Acta*, **43**, 2911 (1998).

[103] P. Nambi, A. Wierzbicki e S. A. Allison, *J. Phys. Chem.*, **95**, 9595 (1991).

[104] T. Solmajer e E. L. Mehler, *Int. J. Quantum Chem.*, **44**, 291 (1992).

[105] R. M. Brunne, E. Liepinsh, G. Otting, K. Wutrich e W. F. van Gunsteren, *J. Mol. Biol.*, **231**, 1040 (1993).

[106] E. W. Knapp e I. Muegge, *J. Phys. Chem.*, **97**, 11339 (1993).

[107] J. Shimada, H. Kaneto e T. Takada, *J. Comp. Chem.*, **14**, 867 (1993).

[108] A. S. Yang, M. R. Gunner, M. Sampogna, K. Sharp e B. Honig, *Prot. Struct. Func. Gen.*, **15**, 252 (1993).

[109] P. F. W. Stouten, C. Frommmel, H. Nakamura e C. Sander, *Mol. Simul.*, **10**, 97 (1993).

[110] Z. X. Xiang, F. H. Huang e Y. Y. Shi, *J. Phys. Chem.*, **98**, 12782 (1994).

[111] P. E. Smith, R. C. van Shaik, T. Szyperski, K. Wuthrich e e W. F. van Gunsteren, *J. Mol. Biol.*, **246**, 356 (1995).

[112] M. Watanabe e M. Karplus, *J. Phys. Chem.*, **99**, 5680 (1995).

[113] R. M. Brunne, K. Berndt, P. Guntert, K. Wutrich e W. F. van Gunsteren, *Prot. Struct. Funct. Gen.*, **23**, 49 (1995).

[114] J. Antosiewicz, J. A. McCammon e M. K. Gilson, *Biochemistry*, **35**, 7819 (1996).

[115] J. P. Kocher, M. Prevost, S. J. Wodak e B. Lee, *Structure*, **4**, 1517 (1996).

[116] H. Kakajima e O. Kikuchi, *J. Mol. Struct. (Theochem)*, **365**, 39 (1996).

[117] D. Xu, J. C. Philips e K. Schulten, *J. Phys. Chem.*, **100**, 12108 (1996).

[118] B. Cheng, A. Nayeen e H. A. Scheraga, *J. Comp. Chem.*, **17**, 1453 (1996).

[119] E. L. Mehler, *J. Phys. Chem.*, **100**, 16006 (1996).

[120] A. Buncsura, I. Y. Enyedy e I. M. Kovach, *J. Am. Chem. Soc.*, **118**, 8531 (1996).

[121] J. D. Augspurger e H. A. Scheraga, *J. Comp. Chem.*, **18**, 1072 (1997).

[122] J. Woenckaus, R. R. Hudgins e M. F. Jarrold, *J. Am. Chem. Soc.*, **119**, 9586 (1997).

[123] V. P. Denisov, K. Venu, J. Peters, H. D. Horlein e B. Halle, *J. Phys. Chem. B*, **101**, 9380 (1997).

[124] P. Fastman e S. Doniach, *Prot. Struct. Funct. Gen.*, **39**, 215 (1998).

[125] D. V. Huffman, *Proc. I.R.E.*, **40**, 1098 (1952).

[126] R. G. Gallager, *IEEE Trans. Inform. Theory*, **IT-24**, 668 (1978).

[127] T. A. Welch, *IEEE Computer*, **17**, 8 (1984).

[128] J. Ziv e A. Lempel, *IEEE Trans. Inform. Theory*, **IT-23**, 337 (1977).

[129] N. S. Jayan, P. Noll, *Digital Coding of Waveforms; Principles and Applications to Speech and Video*, Prentice-Hall, Signal Processing Series, Englewood Cliffs, 1984.

[130] G. K. Wallace, *Comm. ACM*, **34**, 30 (1991).

[131] ISO/IEC JTC 1/SC 29/WG 11, *ISO/IEC CD 13818-3: Information Technology - Generic Coding of Moving Pictures and Associated Audio Information*, (1994).

Capítulo 3: Métodos Quânticos

A meta da investigação já não é, por conseguinte, o conhecimento dos átomos e dos seus movimentos << em si >>, prescindindo da problemática suscitada pelos nossos processos de experimentação; desde o princípio nos encontramos no centro da disputa entre homem e natureza, da qual a ciência não é senão uma parte. As vulgares divisões do universo em sujeito e objecto, mundo interior e mundo exterior, alma e corpo, só servem para suscitar equívocos. De modo que, na ciência, o objecto da investigação não é a natureza em si mesma, mas a natureza subordinada à maneira humana de pôr o problema. Também neste domínio, o homem se encontra sozinho frente a si próprio.

Werner Heisenberg, *A imagem da Natureza na Física Moderna*, Livros do Brasil, Lisboa.

3.1 - Introdução

Os diversos agregados moleculares, que ocorrem na natureza, são compostos por um conjunto de partículas microscópicas elementares que se combinam entre si de diferentes maneiras. A estrutura e propriedades daqueles agregados são, essencialmente, condicionadas pelas interacções que se estabelecem os seus constituintes e o seu estudo constitui um problema de multipartículas de resolução bastante complexa.

No estudo de sistemas moleculares de grande dimensão, atendendo às limitações actuais dos meios de cálculo computacional, métodos baseados na mecânica clássica podem constituir uma primeira aproximação razoável. No entanto, as partículas microscópicas só são correctamente descritas pela mecânica quântica e há determinados fenómenos, como a permuta e partilha electrónicas, que não apresentam qualquer correspondência clássica. Atendendo ao princípio de incerteza de Heisenberg, não se pode determinar simultaneamente, com inteiro rigor, a posição e o momento de uma partícula elementar. Consequentemente, o modelo clássico de órbita não é adequado para descrever sistemas microscópicos. Isto implica que, considerando um sistema de partículas idênticas, é, em princípio, impossível distingui-las umas das outras visto um observador exterior não as poder seguir em nenhum tipo de órbita. Em sistemas quânticos, duas partículas idênticas podem permutar as suas posições sem que isto possa ser detectado por qualquer medição das propriedades do sistema efectuada por um observador exterior. De um ponto de vista matemático, este princípio de indistinguibilidade implica que o resultado da medição de uma observável, que dependa unicamente das coordenadas de partículas idênticas, é invariante a qualquer tentativa de etiquetar estas partículas. No caso particular de sistemas electrónicos, isto conduz ao bem conhecido princípio de exclusão de Pauli que estabelece as propriedades de antissimetria da correspondente função de estado. Os fenómenos de permuta introduzidos desta forma dão origem a uma nova componente da energia (a energia de permuta) que é tipicamente quântica. Esta componente é de grande importância para a compreensão da formação de agregados moleculares covalentes a partir dos seus fragmentos isolados. De facto, a permuta electrónica está sempre associada com a partilha electrónica e, consequentemente, com a formação de ligações covalentes [1].

Num sistema molecular contendo N núcleos (X=1, 2, ... , N) de massas M_A e cargas Z_A e n electrões (i=1, 2, ... , n), a respectiva função de estado, $\Psi_I(Q_1,Q_2,...,Q_N;q_1,q_2,...,q_n)$, pode ser obtida a partir da resolução da equação de Schrödinger independente do tempo,

$$\hat{H}\,\Psi_I(Q_1,Q_2,...,Q_N;q_1,q_2,...,q_n) = E_I\,\Psi_I(Q_1,Q_2,...,Q_N;q_1,q_2,...,q_n)\ , \qquad (3.1)$$

onde Q_W representa as coordenadas do núcleo W ($Q_W=x_Wy_Wz_W$), q_i representa as coordenadas do electrão i ($q_i=x_iy_iz_is_i$), E_k é a energia do estado k e $\hat{H}$ é o hamiltoniano não relativístico do sistema expresso em unidades atómicas:

$$\hat{H} = -\frac{1}{2}\sum_{X=1}^{N}\frac{1}{m_X}\nabla_X^2 + \sum_{X=1}^{N}\sum_{Y=1}^{X-1}\frac{Z_XZ_Y}{\varepsilon r_{XY}} - \frac{1}{2}\sum_{i=1}^{n}\nabla_i^2 - \sum_{X=1}^{N}\sum_{i=1}^{n}\frac{Z_X}{\varepsilon r_{iX}} + \sum_{i=1}^{n}\sum_{j=1}^{i-1}\frac{1}{\varepsilon r_{ij}} \qquad (3.2)$$

Na equação (3.2), ε é a constante dieléctrica do meio e r_{XY}, r_{iX} e r_{ij} representam as distâncias entre os núcleos X e Y, o electrão i e o núcleo X e os electrões i e j, respectivamente.

O hamiltoniano (3.2) contém termos que descrevem não só o movimento dos electrões, mas também o movimento dos núcleos uns em relação aos outros e da molécula como um todo no espaço. A correspondente equação de Schrödinger, equação (3.2), é de muito difícil resolução e só apresenta soluções rigorosas para sistemas moleculares muito simples como o H_2^+ e o H_2 [2]. No entanto, o problema pode ser consideravelmente simplificado introduzindo novas aproximações que conduzam ao tratamento separado dos movimentos electrónicos e nucleares.

As características dos movimentos nucleares e electrónicos, num sistema molecular, não são idênticas. A massa do protão é 1836 vezes maior que a massa do electrão e, consequentemente, as massas nucleares têm uma ordem de grandeza três a quatro vezes maior que a massa electrónica.

Com base em simples argumentos clássicos, é de esperar que o movimento dos electrões seja muito mais rápido do que o movimento dos núcleos, porque a massa dos últimos é muito maior. Consequentemente, a escala de tempo dos dois movimentos pode ser separada. Os núcleos lentos apenas sentem um potencial médio correspondente ao movimento

dos electrões. Pelo contrário, os electrões rápidos conseguem ajustar-se instantaneamente a qualquer alteração das posições nucleares. Baseados neste facto, Born e Oppenheimer [3] formularam uma aproximação segundo a qual se desprezam quaisquer contribuições dos movimentos nucleares quando se pretende estudar unicamente a estrutura electrónica. Esta aproximação implica que a função de estado ($\Psi_I(Q_1,Q_2,...,Q_N;q_1,q_2,...,q_n)$) pode ser expressa como um produto das respectivas componentes nuclear ($\phi_c(Q_1,Q_2,...,Q_N)$) e electrónica ($\psi_k^{\Gamma}(q_1,q_2,\ ...\ ,q_n)$) :

$$\Psi_I(Q_1,Q_2,...,Q_N;q_1,q_2,...,q_n) = \phi_c(Q_1,Q_2,...,Q_N) \times \psi_k^{\Gamma}(q_1,q_2,\ ...\ ,q_n) \qquad (3.3)$$

A notação (Γ) é utilizada para indicar o tipo de estado electrónico do agregado molecular em relação aos respectivos fragmentos constituintes. Neste contexto, os fragmentos são entidades absolutamente genéricas (átomos, grupo de átomos, iões, monómeros, etc). Neste trabalho, é utilizada a seguinte classificação relativamente aos estados electrónicos de agregados moleculares:

Estados não ligantes ($\Gamma = \infty$)- Os fragmentos encontram-se a distância infinita e por isso não existe qualquer interacção entre eles.

Estados não covalentes ligantes ($\Gamma = \mathrm{no}$)- Os fragmentos encontram-se a distância finita mas não existe compartilhamento de carga electrónica entre eles.

Estados covalentes ligantes ($\Gamma = \mathrm{ov}$)- Os fragmentos encontram-se a distância finita mas não existe compartilhamento de carga electrónica entre eles.

Para uma dada configuração de núcleos fixos, a função de estado pode ser obtida através da resolução da equação de Schrödinger electrónica,

$$\hat{H}_{el.}\ \psi_k^{\Gamma}(q_1,q_2,\ ...\ ,q_n) = E_k^{\Gamma,el.}\ \psi_k^{\Gamma}(q_1,q_2,\ ...\ ,q_n)\ , \qquad (3.4)$$

onde $E_k^{\Gamma,el.}$ é a energia do estado k do tipo Γ e $\hat{H}_{el.}$ é o hamiltoniano electrónico definido como:

$$\hat{H}_{el.} = -\frac{1}{2}\sum_{i=1}^{n} \nabla_i^2 - \sum_{X=1}^{N}\sum_{i=1}^{n} \frac{Z_X}{\varepsilon r_{iX}} + \sum_{i=1}^{n}\sum_{j=1}^{i-1} \frac{1}{\varepsilon r_{ij}} \qquad (3.5)$$

A energia total do sistema (E_k^{Γ}), para a configuração de núcleos fixos considerada, é então dada por:

$$E_k^{\Gamma} = E_k^{\Gamma,el.} + \sum_{X=1}^{N}\sum_{Y=1}^{X-1} \frac{Z_X Z_Y}{\varepsilon r_{XY}} \qquad (3.6)$$

Habitualmente, a função de estado nuclear ($\phi_c(Q_1, Q_2, ..., Q_N)$) pode ser expressa como o produto das suas componentes translacional ($\phi_t(C_x, C_y, C_z)$) , rotacional ($\phi_r(r_1, ..., r_q)$) e vibracional ($\phi_v(f_1, f_2, ..., f_s)$),

$$\phi_c(Q_1, Q_2, ..., Q_N) = \phi_t(C_x, C_y, C_z) \times \phi_r(r_1, r_2, ..., r_q) \times \phi_v(f_1, f_2, ..., f_s) , \qquad (3.7)$$

onde (C_x, C_y, C_z), ($r_1, r_2, ..., r_q$) e ($f_1, f_2, ..., f_s$) representam os modos normais de translação, rotação e vibração, respectivamente. Em moléculas lineares, só existem dois modos normais de rotação (q=2) e o número de modos normais de vibração (s) é 3N-5. Em moléculas não lineares, existem três modos normais de rotação e 3N-6 modos normais de vibração. A energia nuclear ($E_c^{nuc.}$) pode então ser escrita como soma das suas componentes translacional ($E_c^{trans.}$), rotacional ($E_r^{rot.}$) e vibracional ($E_v^{vib.}$):

$$E_c^{nuc.} = E_t^{trans.} + E_r^{rot.} + E_v^{vib.} \qquad (3.8)$$

Em termos de uma interpretação molecular correcta das propriedades físicas, a determinação da estrutura vibracional desempenha, habitualmente, um papel muito mais importante do que a determinação das correspondentes estruturas rotacional e translacional. As energias e funções de estado vibracionais são obtidas através da resolução da equação de valores próprios:

$$\hat{H}_{vib.}\ \phi_v(f_1, f_2, \ldots, f_s) = E_v^{vib.}\ \phi_v(f_1, f_2, \ldots, f_s) \tag{3.9}$$

Aplicando a aproximação harmónica a uma configuração nuclear estacionária (e) de um estado electrónico (k), o hamiltoniano vibracional pode ser escrito como um somatório de s componentes ($\hat{H}_l$), cada uma associada a um modo normal de vibração (f_l),

$$\hat{H}_{vib.} = \sum_{l=1}^{s} \hat{H}_l \ , \tag{3.10}$$

onde,

$$\hat{H}_l = -\frac{1}{2}\frac{\partial^2}{\partial f_l^2} + \frac{1}{2}\lambda_l f_l^2 \tag{3.11}$$

Na equação (3.11), λ_l é a constante de força associada ao modo normal f_l definida como,

$$\lambda_l = \left(\frac{\partial^2 E_k}{\partial f_l^2}\right)_e \tag{3.12}$$

Neste formalismo, E_k é encarado como o potencial sentido pelos núcleos nos seus movimentos vibracionais em torno da configuração estacionária (e) do estado electrónico (k).

A função de estado vibracional ($\phi_v(f_1, f_2, \ldots, f_s)$) pode então ser escrita como produto das funções de onda ($\Theta_{v_l}(f_l)$), cada uma associada a um modo normal de vibração (f_l):

$$\phi_v(f_1, f_2, \ldots, f_s) = \prod_{l=1}^{s} \Theta_{v_l}(f_l) \tag{3.13}$$

As funções de onda ($\Theta_{v_l}(f_l)$) são as soluções das equações de valores próprios:

$$\hat{H}_l \, \Theta_{v_l}(f_l) = \varepsilon_{v_l} \, \Theta_{v_l}(f_l) \; ; \; l=1, \, 2, \, \ldots \, , \, s \tag{3.14}$$

A energia vibracional ($E_v^{vib.}$) é então dada por:

$$E_v^{vib.} = \sum_{l=1}^{s} \varepsilon_{v_l} \tag{3.15}$$

As equações (3.14), diferem umas das outras por envolverem modos normais de vibração (f_l) diferentes e constantes de força (λ_l) diferentes. Estas equações são do tipo da equação de Schrödinger para o oscilador harmónico, cujos valores próprios são dados por:

$$\varepsilon_{v_l} = (v_l + 1/2) \, h\upsilon_l \; ; \; v_l = 0, \, 1, \, 2, \, \ldots \, , \tag{3.16}$$

onde,

$$\upsilon_l = (\lambda_l)^{1/2} / 2\pi \tag{3.17}$$

O movimento vibracional de um sistema molecular pode então ser decomposto no movimento de uma colecção de osciladores harmónicos independentes, cada um dos quais está associado a um modo normal (f_l). As quantidades (υ_l), dadas pela equação (3.17), podem agora ser encaradas como as frequências harmónicas associadas a cada modo normal (f_l).

A aproximação de Born-Oppenheimer permite definir o conceito de superfície de energia potencial, em que a energia total (E_k) é representada em termos de coordenadas correspondentes a configurações relevantes de núcleos fixos, que é fundamental na descrição quântica de estruturas e reacções químicas. Nestas superfícies existem dois tipos de estados estacionários relevantes:

Os mínimos (ou estruturas de equilíbrio), para os quais todas as constantes de força (λ_l) e frequências (υ_l) são positivas, e os estados de transição onde existem uma ou mais

constantes de força negativas associadas a frequências imaginárias. As coordenadas associadas a estas frequências designam-se habitualmente por coordenadas de reacção. Um estado de transição corresponde a um máximo de energia relativamente às coordenadas de reacção e a um mínimo relativamente às restantes coordenadas nucleares.

A localização de estados de transição apresenta grande interesse no estudo da reactividade química, porque permite a determinação teórica de energias de activação associadas a processos químicos cujo progresso possa ser expresso em termos de uma ou mais coordenadas de reacção.

Na figura 3.1, é apresentado um exemplo de superfície de energia potencial unidimensional.

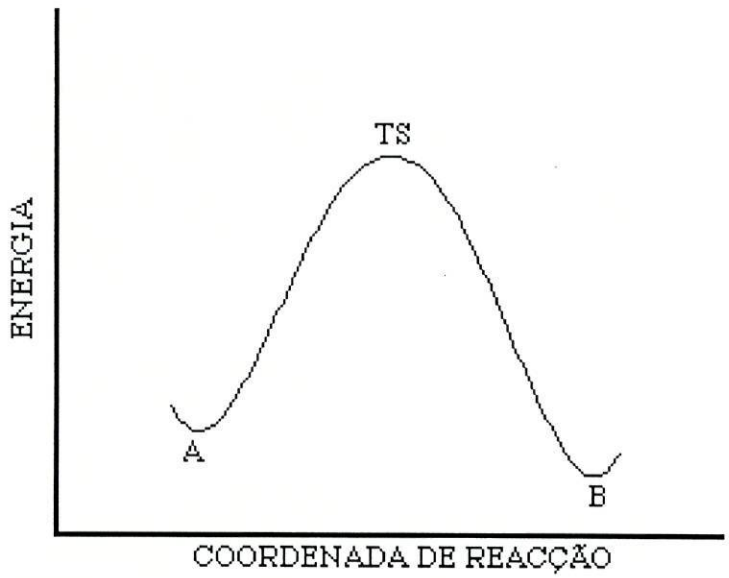


Figura 3.1. Representação esquemática de uma superfície de energia potencial unidimensional. A e B representam mínimos e TS representa o estado de transição.

3.2 - Formalismo das matrizes densidade

O formalismo das matrizes densidade foi introduzido por von Neumann [4] e Dirac [5-7]. Este formalismo foi inicialmente aplicado em mecânica estatística e só posteriormente em mecânica quântica. Nos anos quarenta, Husimi [8] introduziu o conceito de matrizes densidade reduzidas, de importância fundamental para o tratamento adequado de sistemas de

N partículas. No entanto, a formulação consistente de uma teoria geral das matrizes densidade e a sua aplicação generalizada a sistemas quânticos deve-se essencialmente aos trabalhos de Löwdin [9-10] e McWeeny [11-13].

Para um estado electrónico k do tipo Γ de um sistema molecular com N átomos e n electrões, a matriz densidade reduzida é definida como,

$$\rho_k^{\Gamma}(X_1'|X_1) = n \int \psi_k^{\Gamma *}(q_1', q_2, \ldots, q_n)\, \psi_k^{\Gamma}(q_1, q_2, \ldots, q_n)\, ds_1 d\tau_2 \ldots d\tau_n , \qquad (3.18)$$

e a matriz densidade de par reduzida é dada pela expressão,

$$\pi_k^{\Gamma}(X_1'X_2'|X_1X_2) = n(n-1) \int \psi_k^{\Gamma *}(q_1', q_2', \ldots, q_n)\, \psi_k^{\Gamma}(q_1, q_2, \ldots, q_n)\, ds_1 ds_2 d\tau_3 \ldots d\tau_n , \qquad (3.19)$$

onde

$$d\tau_i = dx_i\, dy_i\, dz_i\, ds_i \qquad (3.20)$$

e

$$X_i = x_i\, y_i\, z_i \qquad (3.21)$$

No presente formalismo, os elementos diagonais das matrizes densidade reduzidas são de especial importância e têm o seguinte significado físico:

$\rho_k^{\Gamma}(X_1|X_1)dV_1 = \rho_k^{\Gamma}(X_1)dV_1$: Densidade electrónica provável no elemento infinitesimal de volume $dV_1=(dx_1dy_1dz_1)$ centrado no ponto X_1. (3.22)

$\pi_k^{\Gamma}(X_1X_2|X_1X_2)\, dV_1\, dV_2 = \pi_k^{\Gamma}(X_1X_2)\, dV_1\, dV_2$: Número provável de pares electrónicos em que um dos constituintes se encontra no elemento infinitesimal de volume dV_1 centrado no ponto X_1 e o outro no elemento infinitesimal de volume dV_2 centrado no ponto X_2. (3.23)

As matrizes densidade reduzidas estão associadas com as seguintes relações de conservação:

$$\int \rho_k^{\Gamma}(X_1)\, dV_1 = \int_{X_1'=X_1} \rho_k^{\Gamma}(X_1'|X_1)\, dV_1 = n \tag{3.24}$$

$$\frac{1}{2}\int\int \pi_k^{\Gamma}(X_1X_2)\, dV_1\, dV_2 = \frac{1}{2}\iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \pi_k^{\Gamma}(X_1'X_2'|X_1X_2)\, dV_1\, dV_2 = \frac{n(n-1)}{2} \tag{3.25}$$

$$\int \pi_k^{\Gamma}(X_1'X_2|X_1X_2)\, dV_2 = (n-1)\, \rho_k^{\Gamma}(X_1'|X_1) \tag{3.26}$$

Considerando uma observável (G) do sistema molecular associada a um operador ($\hat{G}$) que possa ser expresso como a soma de operadores elementares independentes do tipo,

$$\hat{G} = G_0 + \sum_{i=1}^{n} \hat{G}_i + \sum_{i=1}^{n}\sum_{j=1}^{i-1} \hat{G}_{ij}\,, \tag{3.27}$$

o correspondente valor esperado para um estado descrito pela função de onda $\psi_k^{\Gamma}(q_1, q_2, \ldots, q_n)$ pode ser inteiramente determinado através das suas matrizes densidade e densidade de par reduzidas,

$$\begin{aligned} \langle G \rangle &= \int \psi_k^{\Gamma *}(q_1, q_2, \ldots, q_n)\, \hat{G}\, \psi_k^{\Gamma}(q_1, q_2, \ldots, q_n)\, d\tau = \left\langle \psi_k^{\Gamma} \middle| \hat{G} \middle| \psi_k^{\Gamma} \right\rangle \\ &= G_k^{\Gamma(0)} + G_k^{\Gamma(1)} + G_k^{\Gamma(2)} \end{aligned} \tag{3.28}$$

com

$$d\tau = d\tau_1\, d\tau_2 \ldots d\tau_n\,, \tag{3.29}$$

$$G_k^{\Gamma(0)} = G_0\,, \tag{3.30}$$

$$G_k^{\Gamma(1)} = = \int_{X_1'=X_1} \hat{G}_1 \, \rho_k^{\Gamma}(X_1'|X_1)\, dV_1 , \tag{3.31}$$

e

$$G_k^{\Gamma(2)} = \frac{1}{2} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{G}_{12} \, \pi_k^{\Gamma}(X_1'X_2'|X_1X_2)\, dV_1\, dV_2 \tag{3.32}$$

O significado da notação *"prima"*, no formalismo das matrizes densidade, pode ser compreendido comparando as equações (3.31) e (3.32) com a equação (3.28). Um operador apenas actua sobre a função de onda ($\psi_k^{\Gamma}(q_1,q_2,...,q_n)$) transformando-a, por exemplo, através de uma derivação. No entanto, para que a posterior integração se possa fazer adequadamente, esta notação permite uma apropriada mudança de variável na função não afectada pelo operador, igualando X_1' a X_1 (nas equações 3.31 e 3.32) e X_2' a X_2 (na equação 3.32).

Para a configuração de núcleos fixos, o hamiltoniano ($\hat{H}$) pode ser expresso de uma forma consistente com a equação (3.27),

$$\hat{H} = H_0 + \sum_{i=1}^{n} \hat{h}_i + \sum_{i=1}^{n} \sum_{j=1}^{i-1} \hat{h}_{ij} , \tag{3.33}$$

onde

$$H_0 = \sum_{X=1}^{N} \sum_{Y=1}^{X-1} \frac{Z_X \, Z_Y}{\varepsilon r_{XY}} , \tag{3.34}$$

$$\hat{h}_i = -\frac{1}{2} \nabla_i^2 - \sum_{X=1}^{N} \frac{Z_X}{\varepsilon r_{iX}} \tag{3.35}$$

e

$$\hat{h}_{ij} = \frac{1}{\varepsilon r_{ij}} \tag{3.36}$$

O correspondente valor esperado da energia total do sistema (E_k), para um estado descrito pela função de onda $\psi_k^{\Gamma}(q_1, q_2, \ldots, q_n)$, pode então ser calculado pela equação :

$$E_k^{\Gamma} = E_k^{\Gamma(0)} + E_k^{\Gamma(1)} + E_k^{\Gamma(2)} \tag{3.37}$$

Na equação (3.37), $E_k^{\Gamma(0)}$ é a energia de repulsão nuclear que é independente das matrizes densidade,

$$E_k^{\Gamma(0)} = H_0 = \sum_{X=1}^{N} \sum_{Y=1}^{X-1} \frac{Z_X Z_Y}{\varepsilon r_{XY}} , \tag{3.38}$$

$E_k^{\Gamma(1)}$ é a energia monoelectrónica que depende unicamente da matriz densidade reduzida,

$$E_k^{\Gamma(1)} = = \int_{X_1'=X_1} \hat{h}_1 \rho_k^{\Gamma}(X_1'|X_1)\, dV_1 \tag{3.39}$$

e $E_k^{\Gamma(2)}$ é a energia bielectrónica que é calculada em termos da matriz densidade de par reduzida

$$E_k^{\Gamma(2)} = \frac{1}{2} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12}\, \pi_k^{\Gamma}(X_1'X_2'|X_1X_2)\, dV_1\, dV_2 \tag{3.40}$$

3.3 Métodos aproximados

A aproximação de Born-Oppenheimer representa uma simplificação considerável na determinação de estados electrónicos em sistemas moleculares. No entanto, só podem ser

obtidas soluções analíticas da equação de Schrödinger electrónica para sistemas monoelectrónicos [14-15]. No estudo de sistemas polielectrónicos, de maior interesse químico, é necessário introduzir métodos numéricos que permitam encontrar soluções aproximadas da referida equação.

A função de onda electrónica ($\psi_k^{\Gamma}(q_1,q_2,...,q_n)$) apresenta uma grande complexidade que dificulta a sua interpretação física e o desenvolvimento de metodologias específicas que permitam a sua determinação aproximada. Uma forma de simplificar a forma desta função consiste na sua expansão em termos de um conjunto ortonormal de funções associadas às diversas configurações electrónicas do sistema:

$$\psi_k^{\Gamma}(q_1,q_2,\ \cdots\ ,q_n) = \sum_{L=1}^{M} C_L \Phi_k^{\Gamma}(q_1,q_2,\ \cdots\ ,q_n) \tag{3.41}$$

Na equação (3.41), $\Phi_k^{\Gamma}(q_1,q_2,\ \cdots\ ,q_n)$ é o determinante de Slater associado à configuração electrónica L do sistema expresso em termos de funções spin-orbital monoelectrónicas (orbitais moleculares),

$$\Phi_k^{\Gamma}(q_1,q_2,\ \cdots\ ,q_n) = (n!)^{-\frac{1}{2}} \left|\varphi_{L_1}(q_1)\,\varphi_{L_2}(q_2)\cdots\varphi_{L_n}(q_n)\right| , \tag{3.42}$$

onde

$$\varphi_i(q_i) = \eta_i(X_i)\,\omega_i(s_i) \tag{3.43}$$

Na equação (3.43), $\eta_i(X_i)$ e $\omega_i(s_i)$ representam respectivamente as componentes espacial e de spin da função $\varphi_i(q_i)$.

$\Phi_k^{\Gamma}(q_1,q_2,\ \cdots\ ,q_n)$ deve ser interpretada como a função obtida colocando um electrão em cada uma das orbitais moleculares $\varphi_{L_1},\varphi_{L_2}\cdots\varphi_{L_n}$, enquanto as restantes orbitais moleculares (φ_N, $N \notin \{L_i\}$) permanecem desocupadas. A introdução destas funções monoelectrónicas permite uma racionalização da distribuição electrónica de uma dada configuração L através da atribuição de cada electrão a uma orbital molecular φ_{L_i}.

No formalismo anteriormente apresentado, as orbitais moleculares φ_i são funções absolutamente genéricas. No entanto, a interpretação da distribuição electrónica num sistema molecular em termos das características dos seus átomos constituintes apresenta grande interesse em aplicações da mecânica quântica ao estudo de problemas químicos. Uma abordagem deste tipo foi sugerida por Roothaan [16] e consiste em exprimir as funções orbitais (η_i) por uma combinação linear de m funções de base atómicas $\{\chi_p\}$:

$$\eta_i = \sum_{p=1}^{m} c_{pi}\,\chi_p \tag{3.44}$$

As equações (3.18) e (3.19) podem agora ser escritas como:

$$\rho_k^{\Gamma}(X_1'|X_1) = \sum_{p=1}^{m}\sum_{q=1}^{m} P_{pq}^{\Gamma,k}\;\chi_p^*(X_1')\;\chi_q(X_1) \tag{3.45}$$

$$\pi_k^{\Gamma}(X_1'X_2'|X_1X_2) = \sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m} D_{pqrs}^{\Gamma,k}\;\chi_p^*(X_1')\;\chi_q(X_1)\,\chi_r^*(X_2')\;\chi_s(X_2) \tag{3.46}$$

Nas equações anteriores, $P_{pq}^{\Gamma,k}$ e $D_{pqrs}^{\Gamma,k}$ são elementos das matrizes densidade e densidade de par reduzidas, do estado electrónico k do tipo Γ, expressas em termos do conjunto de funções de base atómicas. Para simplificar a notação, $P_{pq}^{\Gamma,k}$ e $D_{pqrs}^{\Gamma,k}$ passarão a ser designados como elementos das matrizes densidade e densidade de par, respectivamente.

As relações de conservação (3.24 a 3.26) podem ser expressas no novo espaço por,

$$\sum_{p=1}^{m}\sum_{q=1}^{m} P_{pq}^{\Gamma,k}\,S_{pq} = n\ , \tag{3.47}$$

$$\frac{1}{2}\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m} D_{pqrs}^{\Gamma,k}\,S_{pq}\,S_{rs} = \frac{n(n-1)}{2} \tag{3.48}$$

$$\sum_{r=1}^{m}\sum_{s=1}^{m} D_{pqrs}^{\Gamma,k}\, S_{rs} \;=\; (n-1)\, P_{pq}^{\Gamma,k} \tag{3.49}$$

onde S_{pq} é o integral de sobreposição entre as funções de base χ_p e χ_q :

$$S_{pq} = \langle \chi_p | \chi_q \rangle \tag{3.50}$$

As relações de conservação (3.47) e (3.48) podem ser interpretadas, respectivamente, em termos da análise de populações electrónicas e de populações electrónicas de par no espaço discreto das funções de base atómicas.

As componentes monoelectrónica e bielectrónica da energia podem agora ser expressas em termos de quantidades associadas às funções de base atómicas,

$$E_k^{\Gamma(1)} = \sum_{p=1}^{m}\sum_{q=1}^{m} P_{pq}^{\Gamma,k}\, h_{pq} \tag{3.51}$$

e

$$E_k^{\Gamma(2)} = \frac{1}{2}\sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m} D_{pqrs}^{\Gamma,k}\,(pq|rs) \tag{3.52}$$

onde h_{pq} é o integral monoelectrónico associado às funções de base χ_p e χ_q,

$$h_{pq} = \langle \chi_p | \hat{h}_1 | \chi_q \rangle \;, \tag{3.53}$$

e (pq|rs) é o integral bielectrónico associado às funções de base χ_p, χ_q, χ_r e χ_s:

$$(pq|rs) = \langle \chi_p\, \chi_r | \hat{h}_{12} | \chi_q\, \chi_s \rangle \tag{3.54}$$

Numa primeira abordagem, o desenvolvimento anteriormente apresentado nesta secção não introduz qualquer tipo de aproximação adicional relativamente à forma da função de onda electrónica. De facto, a função obtida em termos de conjuntos completos de funções apropriadas ($M = \infty$ na equação 3.41 e $m = \infty$ na equação 3.44) é exactamente equivalente à função de onda original. No entanto, a dupla expansão infinita utilizada é virtualmente impossível na prática e é necessário recorrer a métodos aproximados para encontrar soluções razoáveis para o problema.

Os métodos aproximados habitualmente mais utilizados baseiam-se no princípio variacional. Consequentemente, num cálculo deste tipo estão envolvidos os seguinte passos:

. Escrever explicitamente o operador $\hat{H}_{el}$ para o sistema em estudo.

. Propor uma função tentativa, representativa da função de onda do sistema em estudo, escrita em termos de parâmetros ajustáveis.

. Minimizar o valor esperado da energia em relação aos parâmetros variacionais escolhidos.

A forma sugerida para a função de onda permite uma aplicação mais fácil de métodos aproximados deste tipo. De facto, truncando a expansão desta função a um número limitado de configurações relevantes para o problema em estudo e escrevendo a componente espacial das orbitais moleculares como uma combinação linear de um conjunto finito de funções de base atómicas, obtém-se uma função tentativa especialmente adequada a um tratamento deste tipo. Os parâmetros variacionais a optimizar serão, em princípio, os coeficientes C_L da expansão (3.41) e c_{pi} da equação (3.44). No entanto, dependendo do método aproximado utilizado, poder-se-ão optimizar os dois conjuntos de coeficientes ou apenas um deles.

Estes métodos são normalmente agrupados em métodos *ab initio* e métodos *semi-empíricos*. A designação *ab initio* está associada aos métodos computacionais em que o procedimento variacional acima descrito é efectuado rigorosamente sem recorrer a nenhum tipo de parametrização empírica. Por sua vez, os métodos *semi-empíricos* introduzem aproximações neste procedimento e recorrem a parâmetros empíricos obtidos com base em dados experimentais ou cálculos *ab initio*.

Os métodos *ab initio* de melhor qualidade preservam a forma multideterminante da função de onda ($M>1$ na equação 3.41), permitindo uma maior flexibilidade da distribuição electrónica obtida a partir do procedimento variacional utilizado [17]. Neste contexto, os métodos mais vulgarmente utilizados são o *método do campo autoconsistente*

multiconfiguracional (MCSCF) [18] e o *método de interacção de configurações* (CI) [19]. No primeiro caso, os coeficientes C_L da expansão (3.41) e os coeficientes c_{pi} da equação (3.44) são optimizados de uma forma simultânea. No segundo caso, esta optimização é efectuada separadamente. Os coeficientes c_{pi} são obtidos previamente utilizando cálculos monoconfiguracionais e os coeficientes C_L são em seguida determinados variacionalmente.

Um cálculo variacional associado a uma expansão do tipo da equação (3.41) é de difícil convergência e apresenta significativas dificuldades em termos computacionais. Por isso, apresenta grande interesse que a função de onda possa ser aproximada por um único determinante de Slater. Uma das alternativas fisicamente aceitável para abordar este problema é a chamada aproximação das partículas independentes ou aproximação monoelectrónica de Hartree-Fock [20-21]. De acordo com este modelo, admite-se que o movimento de um electrão é condicionado pelo campo criado pelos núcleos e pela distribuição média dos restantes electrões desprezando-se, portanto, as repulsões interelectrónicas instantâneas. Assume-se, assim, que o operador $\hat{H}_{el.}$ pode ser aproximado por um operador hamiltoniano electrónico efectivo,

$$\hat{H}^0_{el.} = \sum_{i=1}^{n} (\hat{h}_i + \hat{u}_i) = \sum_{i=1}^{n} \hat{F}_i \ , \qquad (3.55)$$

construído como um somatório de operadores monoelectrónicos (operadores de Fock) $\hat{F}_i$. Na equação (3.55), $\hat{u}_i$ é o operador que representa a repulsão média exercida sobre o electrão i pelos restantes n-1 electrões.

Existem duas formulações da aproximação das partículas independentes: o modelo *não restrito de Hartree-Fock* (UHF) e o modelo *restrito de Hartree-Fock* (RHF). Os modelos UHF e RHF são aproximações particulares dentro do formalismo multiconfiguracional mais geral. Consequentemente, as equações (3.45) a (3.52) são ainda válidas nestes casos particulares. No entanto, é possível derivar expressões específicas para calcular os elementos da matrizes densidade e densidade de par.

No modelo UHF, a função de onda assume a seguinte forma,

$$\psi_k^{\Gamma}(q_1, q_2, \ldots, q_n) = \Phi_k^{\Gamma}(q_1, q_2, \ldots, q_n) =$$
$$= (n!)^{-\frac{1}{2}} \left| \varphi_{k_1}^{\alpha}(q_1)\, \varphi_{k_1}^{\beta}(q_2) \ldots \varphi_{k_v}^{\alpha}(q_{2v-1})\, \varphi_{k_v}^{\beta}(q_{2v})\, \varphi_{k_{v+1}}^{\alpha}(q_{2v+1}) \ldots \varphi_{k_{v+u}}^{\alpha}(q_n) \right| \tag{3.56}$$

onde,

$$\varphi_i^{\alpha}(q_j) = \eta_i^{\alpha}(X_j)\,\alpha_j \text{ e } \varphi_i^{\beta}(q_j) = \eta_i^{\beta}(X_j)\,\beta_j \tag{3.57}$$

As n orbitais moleculares do sistema são subdivididas em dois conjuntos: v+u orbitais (φ_i^{α}) associadas a um spin α e v orbitais associadas a um spin β:

$$\eta_i^{\alpha} = \sum_{p=1}^{m} c_{p_i}^{\alpha} \chi_p \quad , i=k_1, \ldots, k_{v+u} \tag{3.58}$$

$$\eta_i^{\beta} = \sum_{p=1}^{m} c_{p_i}^{\beta} \chi_p \quad , i=k_1, \ldots, k_v \tag{3.59}$$

Similarmente, os elementos das matrizes das matrizes densidade podem ser expressos em termos de componentes α e β,

$$P_{pq}^{\Gamma,k} = P_{pq}^{\Gamma,k\alpha} + P_{pq}^{\Gamma,k\beta} \tag{3.60}$$

onde,

$$P_{pq}^{\Gamma,k\alpha} = \sum_{i=1}^{v+u} c_{p_{k_i}}^{\alpha *} c_{q_{k_i}}^{\alpha} \quad , \tag{3.61}$$

e

$$P_{pq}^{\Gamma,k\beta} = \sum_{i=1}^{v} c_{p_{k_i}}^{\beta *} c_{q_{k_i}}^{\beta} \tag{3.62}$$

Dentro do formalismo UHF, os elementos da matriz densidade de par podem ser calculados pela equação:

$$D^{\Gamma,k}_{pqrs} = P^{\Gamma,k}_{pq} P^{\Gamma,k}_{rs} - (P^{\Gamma,k\alpha}_{ps} P^{\Gamma,k\alpha}_{rq} + P^{\Gamma,k\beta}_{ps} P^{\Gamma,k\beta}_{rq}) \tag{3.63}$$

No modelo RHF, aplicado a sistemas de camada fechada, assume-se que a mesma função orbital (η_i) descreve um par electrónico:

$$\eta_i = \sum_{p=1}^{m} c_{p_i} \chi_p \quad , i=k_1, .. ,k_{n/2} \tag{3.64}$$

A correspondente função de onda pode então ser escrita como,

$$\begin{aligned} \psi^{\Gamma}_{k}(q_1,q_2, \ldots ,q_n) &= \Phi^{\Gamma}_{k}(q_1,q_2, \ldots ,q_n) = \\ &= (n!)^{-\frac{1}{2}} \; |\varphi^{\alpha}_{k_1}(q_1)\, \varphi^{\beta}_{k_1}(q_2)\, \varphi^{\alpha}_{k_2}(q_3)\; \varphi^{\beta}_{k_2}(q_4) \; \ldots \; \varphi^{\alpha}_{k_{n/2}}(q_{n-1})\, \varphi^{\beta}_{k_{n/2}}(q_n)\,| \end{aligned} \tag{3.65}$$

onde,

$$\varphi^{\alpha}_{i}(q_j) = \eta_i(X_j)\,\alpha_j \quad \text{e} \quad \varphi^{\beta}_{i}(q_{j+1}) = \eta_i(X_{j+1})\,\beta_{j+1} \tag{3.66}$$

Dentro do formalismo RHF, podem derivar-se as seguintes relações relativamente aos elementos das matrizes densidade e densidade de par:

$$P^{\Gamma,k\alpha}_{pq} = P^{\Gamma,k\beta}_{pq} = \frac{1}{2} P^{\Gamma,k}_{pq} \tag{3.67}$$

$$P^{\Gamma,k}_{pq} = 2 \sum_{i=1}^{n/2} c^{*}_{p_{k_i}} c_{q_{k_i}} \tag{3.68}$$

$$D^{\Gamma,k}_{pqrs} = P^{\Gamma,k}_{pq} P^{\Gamma,k}_{rs} - \frac{1}{2} P^{\Gamma,k}_{ps} P^{\Gamma,k}_{rq} \tag{3.69}$$

Um cálculo variacional aplicado a uma função de onda monoconfiguracional conduz às chamadas equações de Hartree-Fock. No formalismo RHF estas equações tomam a forma,

$$\hat{F}_i \, \eta_i(X_1) = e_i \, \eta_i(X_1) \; (i = 1, \ldots, n/2) \; , \tag{3.70}$$

onde os operadores de Fock ($\hat{F}_i$) são definidos como:

$$\hat{F}_i = \hat{h}_i + \sum_{j=k_1}^{k_{n/2}} 2\hat{J}_j - \hat{K}_j \; , \tag{3.71}$$

Na equação (3.71), $\hat{J}_j$ e $\hat{K}_j$ são os operadores de Coulomb e permuta dados por:

$$\hat{J}_j \, \eta_i(X_1) = \left\langle \eta_j(X_2) \middle| \hat{h}_{12} \middle| \eta_i(X_1) \, \eta_j(X_2) \right\rangle \tag{3.72}$$

$$\hat{K}_j \, \eta_i(X_1) = \left\langle \eta_j(X_2) \middle| \hat{h}_{12} \middle| \eta_j(X_1) \, \eta_i(X_2) \right\rangle \tag{3.73}$$

Os valores próprios (e_i) dos operadores de Fock podem ser calculados em termos dos integrais monoelectrónicos (h_{ii}) e bielectrónicos coulômbicos (J_{ij}) e de permuta (K_{ij}),

$$e_i = h_{ii} + \sum_{j=k_1}^{k_{n/2}} G_{ij} = h_{ii} + \sum_{j=k_1}^{k_{n/2}} 2J_{ij} - K_{ij} \; , \tag{3.74}$$

onde

$$h_{ii} = \left\langle \eta_i(X_1) \middle| \hat{h}_1 \middle| \eta_i(X_1) \right\rangle \tag{3.75}$$

$$J_{ij} = \left\langle \eta_i(X_1) \, \eta_j(X_2) \middle| \hat{h}_{12} \middle| \eta_i(X_1) \, \eta_j(X_2) \right\rangle \tag{3.76}$$

$$K_{ij} = \left\langle \eta_i(X_1)\, \eta_j(X_2) \left| \hat{h}_{12} \right| \eta_j(X_1)\, \eta_i(X_2) \right\rangle \tag{3.77}$$

A energia total pode então ser calculada em termos das componentes h_{ii} e G_{ij}:

$$E_k^{\Gamma} = E_k^{\Gamma\,(0)} + 2 \sum_{i=k_1}^{k_{n/2}} h_{ii} + \sum_{i=k_1}^{k_{n/2}} \sum_{j=k_1}^{k_{n/2}} G_{ij} \tag{3.78}$$

Em sistemas moleculares é extremamente difícil encontrar soluções para as equações de Hartree-Fock sem recorrer a aproximações que restrinjam o universo de funções η_i a serem pesquisadas. A aplicação da aproximação de Roothaan (equação 3.64) fornece uma boa metodologia para simplificar estas equações, pois basta agora encontrar o conjunto de coeficientes c_{pi}, referentes à base de funções atómicas previamente estabelecidas $\{ \chi_p \}$, que conduza a um valor mínimo da energia. Chega-se deste modo às equações de Hartree-Fock-Roothaan [16],

$$\sum_{q=1}^{m} c_{pi} \, (F_{pq}^{\Gamma,k} - e_i \, S_{pq}) = 0 \text{ ou } (\mathbf{F}^{\Gamma,k} - \mathbf{E}\,\mathbf{S})\,\mathbf{C} = 0 \,, \; p = 1, 2, \ldots, m \,, \tag{3.79}$$

onde $F_{pq}^{\Gamma,k}$ são os elementos da matriz de Fock definidos como:

$$F_{pq}^{\Gamma,k} = h_{pq} + \sum_{r=1}^{m} \sum_{s=1}^{m} P_{rs}^{\Gamma,k} \, G_{pqrs} = h_{pq} + \sum_{r=1}^{m} \sum_{s=1}^{m} P_{rs}^{\Gamma,k} \left\{ 2(pq|rs) - (ps|rq) \right\} \tag{3.80}$$

A energia pode então ser calculada em termos dos integrais monoelectrónicos e dos elementos da matriz de Fock:

$$E_k^{\Gamma} = E_k^{\Gamma\,(0)} + \frac{1}{2} \sum_{p=1}^{m} \sum_{q=1}^{m} P_{pq}^{\Gamma,k} \, (h_{pq} + F_{pq}^{\Gamma,k}) \tag{3.81}$$

As equações de Hartree-Fock-Roothaan só podem ser resolvidas iterativamente uma vez que a forma das orbitais moleculares (ou mais concretamente o valor dos coeficientes que as determinam) depende do potencial médio sentido pelos respectivos electrões e este potencial é uma função das orbitais moleculares escolhidas. O método do *campo autoconsistente* (SCF) [22] é o método mais usado para resolver estas equações.

No formalismo UHF as equações de Hartree-Fock são subdivididas em dois conjuntos:

$$\hat{F}_i^{\alpha}\ \eta_i^{\alpha}(X_1) = e_i^{\alpha}\ \eta_i^{\alpha}(X_1)\ ,\ i = k_1,\ \ldots\ ,\ k_{v+u} \tag{3.82}$$

e

$$\hat{F}_i^{\beta}\ \eta_i^{\beta}(X_1) = e_i^{\beta}\ \eta_i^{\beta}(X_1)\ ,\ i = k_1,\ \ldots\ ,\ k_v \tag{3.83}$$

Nas equações (3.82) e (3.83), $\hat{F}_i^{\alpha}$ e $\hat{F}_i^{\beta}$ são, respectivamente, os operadores de Fock associados a um electrão α e a um electrão β, definidos como,

$$\hat{F}_i^{\alpha} = \hat{h}_i + \sum_j^{(\alpha+\beta)} \hat{J}_j - \sum_j^{(\alpha)} \hat{K}_j \tag{3.84}$$

e

$$\hat{F}_i^{\beta} = \hat{h}_i + \sum_j^{(\alpha+\beta)} \hat{J}_j - \sum_j^{(\beta)} \hat{K}_j\ , \tag{3.85}$$

onde as notações $(\alpha+\beta)$, (α) e (β) significam que os somatórios são extensíveis a todas as funções orbitais, só às funções orbitais α e só às funções orbitais β, respectivamente.

Os valores próprios dos operadores de Fock podem ser calculados como:

$$e_i^{\alpha} = h_{ii} + \sum_{j}^{(\alpha+\beta)} J_{ij} - \sum_{j}^{(\alpha)} K_{ij} \ , \tag{3.86}$$

e

$$e_i^{\beta} = h_{ii} + \sum_{j}^{(\alpha+\beta)} J_{ij} - \sum_{j}^{(\beta)} K_{ij} \tag{3.87}$$

A energia total pode então ser calculada em termos das componentes h_{ii}, J_{ij} e K_{ij}:

$$E_k^{\Gamma} = E_k^{\Gamma(0)} + \sum_{i}^{(\alpha+\beta)} h_{ii} + \sum_{i}^{(\alpha+\beta)}\sum_{j}^{(\alpha+\beta)} J_{ij} - \sum_{i}^{(\alpha)}\sum_{j}^{(\alpha)} K_{ij} - \sum_{i}^{(\beta)}\sum_{j}^{(\beta)} K_{ij} \tag{3.88}$$

A aplicação da aproximação de Roothaan ao formalismo UHF(equações 3.58 e 3.59) conduz também a dois conjuntos de equações,

$$\sum_{q=1}^{m} c_{pi}^{\alpha} \left(F_{pq}^{\Gamma,k\alpha} - e_i^{\alpha} S_{pq}\right) = 0 \text{ ou } \mathbf{C}^{\alpha} \left(\mathbf{F}^{\Gamma,k\alpha} - \mathbf{E}^{\alpha} \mathbf{S}\right) = 0 \ , \ p = 1, 2, \ldots , m \ , \tag{3.89}$$

e

$$\sum_{q=1}^{m} c_{pi}^{\beta} \left(F_{pq}^{\Gamma,k\beta} - e_i^{\beta} S_{pq}\right) = 0 \text{ ou } \mathbf{C}^{\beta} \left(\mathbf{F}^{\Gamma,k\beta} - \mathbf{E}^{\beta} \mathbf{S}\right) = 0 \ , \ p = 1, 2, \ldots , m \tag{3.90}$$

onde os elementos da matriz de Fock são definidos como:

$$F_{pq}^{\Gamma,k\alpha} = h_{pq} + \sum_{r=1}^{m}\sum_{s=1}^{m} \left\{ P_{rs}^{\Gamma,k} (pq|rs) - P_{rs}^{\Gamma,k\alpha} (ps|rq) \right\} \tag{3.91}$$

e

$$F_{pq}^{\Gamma,k\beta} = h_{pq} + \sum_{r=1}^{m}\sum_{s=1}^{m}\left\{P_{rs}^{\Gamma,k}\,(pq|rs) - P_{rs}^{\Gamma,k\beta}\,(ps|rq)\right\} \qquad (3.92)$$

As equações (3.89) e (3.90) são também habitualmente resolvidas pelo método SCF.

A forma antissimétrica das funções de onda características dos métodos de partículas independentes (UHF e RHF) assegura que dois electrões com spins paralelos não podem ocupar simultaneamente a mesma posição espacial. Este fenómeno é habitualmente designado por *buraco de Fermi*. Dentro do formalismo das matrizes densidade, isto significa que os elementos diagonais da matriz densidade de par reduzida $\pi_k(X\,X)$ (com $X_1 = X_2 = X$) sejam nulos sempre que os spins dos electrões envolvidos sejam paralelos. Por outro lado, os correspondentes elementos da matriz densidade reduzida $\rho_k(X)$ são, em geral, não nulos. Isto significa que a probabilidade de encontrar dois electrões com spins paralelos na mesma posição do espaço X é nula, mesmo quando existe uma densidade electrónica considerável neste ponto. Consequentemente, a antissimetria da função de onda introduz alguma correlação nos movimentos de electrões com spins paralelos.

Outra correlação entre os movimentos dos electrões é representada pelo designado *buraco de Coulomb*. As repulsões coulômbicas asseguram que estas partículas não se possam aproximar demasiadamente umas das outras mesmo quando têm spins antiparalelos. No modelo das partículas independentes, o *buraco de Coulomb* não é convenientemente tratado porque as repulsões instantâneas sentidas por um dado electrão são substituídas por um potencial efectivo médio que reflecte a interacção com os restantes electrões. A energia exacta do sistema ($E_k^{\Gamma,\,ext.}$) é obrigatoriamente menor que a obtida com o referido modelo, porque as repulsões interelectrónicas instantâneas obrigam a que os electrões permaneçam mais afastados uns dos outros.

A energia de correlação ($E_k^{\Gamma,correl.}$) é então definida como [23],

$$E_k^{\Gamma,\text{correl.}} = E_k^{\Gamma,\text{ext.}} - E_k^{\Gamma,\text{lim. HF}} , \tag{3.93}$$

onde $E_k^{\Gamma,\text{lim. HF}}$ é a energia do estado obtida no chamado limite de Hartree-Fock. Esta energia é obtida a partir de uma função de onda monoconfiguracional onde as orbitais moleculares associadas são completamente flexíveis. Isto corresponde, dentro da aproximação de Roothaan, a representar a parte espacial de cada orbital molecular como uma expansão infinita de funções de base atómicas (m=∞ na equação 3.44).

A energia de correlação representa habitualmente menos de 1% da energia total do sistema [22]. No entanto, em aplicações químicas são habitualmente mais usadas grandezas do tipo diferenças de energia (por exemplo, energias de estabilização e energias de ligação) cujo valor é da mesma ordem de grandeza de $E_k^{\Gamma,\text{correl.}}$. Nestes casos, o uso do método das partículas independentes só é razoável se a energia de correlação não variar significativamente nos processos químicos considerados.

O tratamento da correlação electrónica pode ser efectuado introduzindo uma maior flexibilidade na função de onda que permita uma redução das repulsões interelectrónicas. Os métodos multiconfiguracionais (MCSCF e CI) atrás mencionados são especialmente apropriados para este efeito. Uma alternativa a estes métodos variacionais é o chamado método de Møller-Plesset [24-27] baseado na teoria das perturbações. De acordo com este método, o hamiltoniano electrónico é escrito como,

$$\hat{H}_{el.} = \hat{H}_{el.}^0 + \hat{V} , \tag{3.94}$$

onde,

$$\hat{V} = \sum_{i=1}^{n} \sum_{j=1}^{i-1} \hat{h}_{ij} - \sum_{i=1}^{n} \hat{u}_i \tag{3.95}$$

O hamiltoniano não perturbado escolhido é o hamiltoniano electrónico efectivo ($\hat{H}_{el.}^0$) e as funções de onda não perturbadas são os correspondentes determinantes de Slater $\Phi_k^{\Gamma}(q_1, q_2, \ldots, q_n)$:

$$\hat{H}^{0}_{el.}\ \Phi^{\Gamma}_{k}(q_1,q_2,\ \dots\ ,q_n) = \varepsilon^{\Gamma,el.}_{k}\ \Phi^{\Gamma}_{k}(q_1,q_2,\ \dots\ ,q_n) \tag{3.96}$$

A correlação electrónica é então tratada em termos da perturbação $\hat{V}$. A função de onda ($\psi^{\Gamma}_{k}(q_1,q_2,\dots,q_n)$) e energia electrónica ($E^{\Gamma,el.}_{k}$) são escritas na forma de expansões perturbacionais:

$$\psi^{\Gamma}_{k}(q_1,q_2,\dots,q_n) = \Phi^{\Gamma}_{k}(q_1,q_2,\dots,q_n) + \psi^{\Gamma\,1}_{k}(q_1,q_2,\ \dots\ ,q_n) + \psi^{\Gamma\,2}_{k}(q_1,q_2,\dots,q_n) + \dots \tag{3.97}$$

$$E^{\Gamma,el.}_{k} = \varepsilon^{\Gamma,el.}_{k} + E^{\Gamma,el.\,1}_{k} + E^{\Gamma,el.\,2}_{k} + \dots \tag{3.98}$$

As correcções perturbacionais na função de onda são calculadas em termos de configurações excitadas a partir da configuração de referência $\Phi^{\Gamma}_{k}(q_1,q_2,\dots,q_n)$. A ordem da correcção determina as configurações a introduzir em $\psi^{\Gamma}_{k}(q_1,q_2,\dots,q_n)$. De acordo com o teorema de Brillouin [28], as configurações monoexcitadas não contribuem para melhorar as funções monodeterminante $\Phi^{\Gamma}_{k}(q_1,q_2,\dots,q_n)$ Consequentemente, a correcção de 1ª ordem na função de onda ($\psi^{\Gamma\,1}_{k}(q_1,q_2,\ \dots\ ,q_n)$) é expressa em termos de configurações duplamente excitadas,

$$\psi^{\Gamma\,1}_{k}(q_1,q_2,\ \dots\ ,q_n) = \sum_{\substack{i>j \\ a>b}} C^{ab\,1}_{ij}\ \Phi^{ab}_{ij}(q_1,q_2,\ \dots\ ,q_n)\ , \tag{3.99}$$

onde,

$$C^{ab\,1}_{ij} = \frac{< \Phi^{ab}_{ij} | \hat{V} | \Phi_k >}{\varepsilon^{\Gamma,el.}_{k} - \varepsilon^{ab}_{ij}} \tag{3.100}$$

Nas equações anteriores, $\Phi_{ij}^{ab}(q_1, q_2, \ldots, q_n)$ é o determinante de Slater obtido substituindo as orbitais moleculares ocupadas $\varphi_i(q_i)$ e $\varphi_j(q_j)$ pelas orbitais moleculares não ocupadas $\varphi_a(q_j)$ e $\varphi_b(q_j)$ e ε_{ij}^{ab} é a correspondente energia electrónica efectiva.

As correcções de ordem superior na função de onda são expressas em termos de configurações obtidas através da excitação de um número crescente de electrões:

A correcção de 2ª ordem é expressa como uma combinação linear de configurações triplamente excitadas. A correcção de 3ª ordem é expressa como uma combinação linear de configurações tetraexcitadas , etc.

As correcções perturbacionais na energia podem ser calculadas através das equações:

$$E_k^{\Gamma,el.\,1} = \langle \Phi_k^\Gamma | \hat{V} | \Phi_k^\Gamma \rangle \tag{3.101}$$

$$E_k^{\Gamma,el.\,2} = \langle \Phi_k^\Gamma | \hat{V} | \psi_k^{\Gamma\,1} \rangle = \sum_{\substack{i>j \\ a>b}} |C_{ij}^{ab\,1}|^2 \, (\varepsilon_k^{\Gamma,el.} - \varepsilon_{ij}^{ab}) \tag{3.102}$$

$$E_k^{\Gamma,el.\,3} = \langle \Phi_k^\Gamma | \hat{V} | \psi_k^{\Gamma\,2} \rangle \tag{3.103}$$

.....

$$E_k^{\Gamma,el.\,n} = \langle \Phi_k^\Gamma | \hat{V} | \psi_k^{\Gamma\,n-1} \rangle \tag{3.104}$$

A correcção de 2ª ordem na energia ($E_k^{\Gamma,el.\,2}$) está associada à correcção de 1ª ordem na função de onda ($\psi_k^{\Gamma\,1}(q_1, q_2, \ldots, q_n)$) e corresponde ao nível mais elementar (MP2) a que se pode aplicar o método de Møller-Plesset para calcular energias de correlação. Níveis mais altos de aplicação desta teoria perturbacional não são conceptualmente muito diferentes:

Ao nível MP3, são introduzidas correcções de 3ª ordem na energia e de 2ª ordem na função de onda. Ao nível MP4, são introduzidas correcções de 4ª ordem na energia e de 3ª ordem na função de onda. Ao nível MPn, são introduzidas correcções de nª ordem na energia e de n-1ª ordem na função de onda.

A introdução de uma maior flexibilidade na função de onda, utilizando os métodos *ab initio* multiconfiguracionais ou o método de Møller-Plesset, permite um tratamento

aproximado da correlação electrónica associada ao *buraco de Coulomb*. Este tratamento será tanto mais rigoroso quanto maior for o número de configurações incluídas na expansão da função de onda $\psi_k^{\Gamma}(q_1, q_2, ..., q_n)$. No limite $(M = \infty)$, esta função é obtida através de uma expansão completa de configurações a que corresponde o melhor tratamento da correlação electrónica para a base de funções utilizada. Por outro lado, esta base pode ser melhorada aumentando o número de funções utilizadas na equação (3.44). Isto permite uma melhor descrição das orbitais moleculares, que no limite de Hartree-Fock $(m = \infty)$, são funções absolutamente flexíveis.

As funções de base habitualmente utilizadas neste tipo de cálculos são as gaussianas contraídas $(\chi_{n,l,m}(q))$ [29-31]. Estas funções são expressas como combinações lineares, com coeficientes fixos c_t, de P funções gaussianas primitivas $(G_{n,l,m}(\zeta, q))$:

$$\chi_{n,l,m}(q) = \sum_{t=1}^{P} c_t \, G_{n,l,m}(\zeta_t, q) \, , \tag{3.105}$$

onde,

$$G_{n,l,m}(\zeta_t, q) = N(\zeta_t, n, l, m) \, x^l \, y^m \, z^n \exp(-\zeta_t (x^2 + y^2 + z^2)) \tag{3.106}$$

Na equação (3.106), $N(\zeta_t, n, l, m)$ é uma constante de normalização, n, l e m são números quânticos e ζ_t é uma constante habitualmente designada por expoente da gaussiana primitiva.

Este tipo de funções de base consegue conjugar uma boa descrição da distribuição electrónica nas regiões vizinhas dos núcleos com uma forma matemática apropriada para uma rápida determinação dos integrais bielectrónicos policêntricos [32].

O nível mais elementar de cálculos *ab initio* envolve o uso de bases mínimas, onde cada orbital atómica dos níveis ocupados é representada por uma gaussiana contraída. As bases mínimas mais usadas são as STO-3G [33], onde cada gaussiana contraída é expandida como uma combinação linear de três primitivas e pretende reproduzir uma função do tipo Slater.

As bases mínimas não têm capacidade de descrever adequadamente a distribuição anisotrópica da carga electrónica em sistemas moleculares. Para uma descrição mais correcta desta distribuição de carga é necessário recorrer a bases em que cada orbital atómica seja representada por mais de uma função. Neste contexto, as bases do tipo duplo zeta, em que cada orbital atómica é representada por duas gaussianas contraídas, são as mais usadas. Neste trabalho, é utilizada uma base deste tipo proposta por Dunning e Hay [34] e habitualmente designada por D95. As bases do tipo triplo zeta, em que cada orbital atómica é representada por três gaussianas contraídas, são muito menos comuns.

Na descrição da ligação química em sistemas moleculares, as orbitais atómicas de valência desempenham um papel muito mais importante que as orbitais atómicas do cerne. Por isso, foram propostas bases em que se usa uma representação diferenciada dos dois tipos de orbitais. Habitualmente, as orbitais atómicas do cerne são representadas por uma base mínima e as orbitais atómicas de valência são representadas por bases duplo ou triplo zeta. Neste trabalho, são utilizadas duas bases deste tipo: a 6-31G [35-36] e a 6-311G [37]. Em ambas as bases, as orbitais atómicas do cerne são representadas por uma gaussiana contraída obtida como combinação linear de seis gaussianas primitivas. As orbitais atómicas de valência são representadas de forma diferente nas duas bases. Na base 6-31G, cada orbital deste tipo é representada por uma gaussiana e por uma gaussiana contraída obtida como combinação linear de três gaussianas primitivas. Na base 6-311G, cada orbital deste tipo é representada por duas gaussianas e por uma gaussiana contraída obtida como combinação linear de três gaussianas primitivas.

As bases anteriormente apresentadas apresentam dificuldades em descrever sistemas moleculares onde existam electrões fracamente ligados, o que implica a existência de uma densidade electrónica considerável em regiões distantes dos núcleos. Uma das formas de abordar este problema consiste no uso de bases contendo funções difusas [38], que são apropriadas para descrever distribuições electrónicas com energias próximas do limite de ionização. Neste trabalho, é utilizada uma base deste tipo, a 6-31+G [39-42], que é obtida a partir da base 6-31G por adição, para todos os átomos à exclusão do hidrogénio, de um conjunto de gaussianas difusas do tipo s e p.

As bases anteriormente referidas apresentam dificuldades no tratamento adequado de sistemas moleculares fortemente polarizáveis e de fenómenos de transferência de carga. Nestes casos, a adição de gaussianas com número quântico do momento angular (l) mais

elevado permite uma melhor descrição da distorção de simetria resultante da colocação de um átomo no ambiente molecular. Estas funções são habitualmente designadas por funções de polarização. Neste contexto, a notação (*) significa a adição, a uma base comum, de um conjunto de gaussianas do tipo d para todos os átomos à excepção do hidrogénio e a notação (**) significa a adição suplementar de um conjunto de gaussianas do tipo p para todos os átomos de hidrogénio. Neste trabalho são utilizadas oito bases deste tipo: 6-31G* [43], 6-31G** [43], 6-311G* [37], 6-311G** [37], D95* [34], D95** [34], 6-31+G* [39-42] e 6-31+G** [39-42].

A figura 3.2 apresenta um diagrama indicando as duas direcções em que podem ser introduzidas refinamentos num cálculo *ab initio*:

Método de Hartree-Fock monoconfiguracional com uma base mínima	Refinamentos no tratamento da correlação electrónica →	Expansão completa de configurações
Melhoramento da base de funções ↓		
Limite de Hartree-Fock		Solução exacta da equação de Schrödinger electrónica

Figura 3.2. Representação esquemática dos tipos de refinamentos que podem ser introduzidos num cálculo *ab initio*.

Um cálculo monoconfiguracional utilizando uma base mínima corresponde ao nível mais baixo a que estes métodos podem ser aplicados e é representado no canto inferior esquerdo do diagrama. As duas direcções representadas na figura correspondem aos dois tipos de refinamentos que podem ser introduzidos no modelo original. Refinamentos introduzidos

na direcção vertical correspondem a um aumento da flexibilidade das orbitais moleculares obtido através do melhoramento da base de funções atómicas utilizada na sua expansão. O limite de Hartree-Fock, correspondente a orbitais moleculares completamente flexíveis, é representado no canto inferior esquerdo do diagrama. Refinamentos introduzidos na direcção horizontal correspondem a uma melhoria no tratamento da correlação electrónica obtida através do aumento do número de configurações utilizadas na expansão da função de onda. O limite, correspondente a uma expansão completa de configurações, é representado no canto superior direito do diagrama. A solução exacta da equação de Schrödinger electrónica, representada no canto inferior direito do diagrama, pode ser obtida conjugando uma expansão completa de configurações com orbitais moleculares completamente flexíveis.

O tempo de computação gasto num cálculo *ab initio* cresce muito rapidamente com a dimensão do sistema molecular. Este facto limita a aplicação destes métodos a sistemas de pequena e média dimensão. O tratamento quântico de moléculas de maior dimensão, de grande interesse no estudo de sistemas biológicos, só é possível recorrendo a métodos *semi-empíricos*. Estes métodos partilham o formalismo variacional com os métodos *ab initio*, mas, de forma a reduzir drasticamente o tempo necessário à sua implementação computacional, introduzem as seguintes simplificações:

(i) Admite-se que os electrões internos constituem um cerne não polarizável, que não intervém directamente na ligação química, só afectando o potencial nuclear sentido pelos electrões de valência. Consequentemente, o hamiltoniano electrónico (equação 3.5) toma a forma,

$$\hat{H}_{el.} = -\frac{1}{2}\sum_{i=1}^{n_{val.}} \nabla_i^2 + \sum_{X=1}^{N}\sum_{i=1}^{n_{val.}} V_X(i) + \sum_{i=1}^{n_{val.}}\sum_{j=1}^{i-1} \frac{1}{\varepsilon r_{ij}}, \qquad (3.107)$$

onde $n_{val.}$ é o número de electrões de valência do sistema molecular e $V_X(i)$ representa a energia potencial de interacção entre o cerne X e o electrão de valência i.

Na equação (3.6), o primeiro termo refere-se agora à energia electrónica de valência ($E_k^{\Gamma,el.\,val.}$) e o segundo é substituído por outro representando a energia de repulsão entre os cernes X e Y (C_{XY}):

$$E_k^{\Gamma} = E_k^{\Gamma,\text{el. val.}} + \sum_{X=1}^{N}\sum_{Y=1}^{X-1} C_{XY} \qquad (3.108)$$

(ii) Utiliza-se uma base mínima de funções de Slater, cujos expoentes são tratados como parâmetros empíricos.

(iii) Utiliza-se, em maior ou menor escala, a *aproximação da sobreposição diferencial nula* (ZDO) [44], que assume que todos os integrais envolvendo sobreposições do tipo $\chi_p(X_1)$ $\chi_r(X_1)$, com p≠r, são nulos.

Em nenhum destes métodos, a aproximação ZDO é aplicada no cálculo dos integrais monoelectrónicos (h_{pr}), sendo estes calculados semi-empiricamente de forma a reflectir o efeito ligante da sobreposição. Para todos estes métodos é aplicada a referida aproximação ao cálculo dos integrais de sobreposição (S_{pr}). Consequentemente, a base de funções atómicas é considerada ortonormal,

$$S_{pr} = \delta_{pr}\,, \qquad (3.109)$$

o que implica que as relações de conservação (3.47 a 3.49) podem ser simplificadas:

$$\sum_{p=1}^{m} P_{pp}^{\Gamma,k} = n\,, \qquad (3.110)$$

$$\frac{1}{2}\sum_{p=1}^{m}\sum_{r=1}^{m} D_{ppr r}^{\Gamma,k} = \frac{n(n-1)}{2}\,, \qquad (3.111)$$

e

$$\sum_{r=1}^{m} D_{pqr r}^{\Gamma,k} = (n-1)\, P_{pq}^{\Gamma,k} \qquad (3.112)$$

A extensão com que a aproximação ZDO é utilizada no cálculo dos integrais bielectrónicos (pq|rs) permite classificar os métodos *semi-empíricos* nas seguintes famílias:

Métodos CNDO (Complete Neglect of Differential Overlap)

A aproximação ZDO é aplicada a todos os integrais bielectrónicos,

$$(p_X q_Y|r_W s_Z) = (p_X p_X|r_W r_W)\, \delta_{XY}\, \delta_{WZ}\, \delta_{pq}\, \delta_{rs}\,, \tag{3.113}$$

onde os índices utilizando letras minúsculas estão associados a funções de base atómicas e os índices utilizando letras maiúsculas estão associados a cernes atómicos.

Métodos INDO (Intermediate Neglect of Differential Overlap)

A aproximação ZDO é aplicada a todos os integrais bielectrónicos bicêntricos, tricêntricos e tetracêntricos:

$$(p_X q_Y|r_W s_Z) = (p_X p_X|r_W r_W)\, \delta_{XY}\, \delta_{WZ}\, \delta_{pq}\, \delta_{rs}\,, \text{ com } X \neq W \tag{3.114}$$

No entanto, todos os integrais bielectrónicos monocêntricos são conservados.

Métodos NDDO (Neglect of Diatomic Differential Overlap)

A aproximação ZDO é aplicada a todos os integrais bielectrónicos que envolvam sobreposições entre funções de base localizadas em átomos diferentes:

$$(p_X q_Y|r_W s_Z) = (p_X q_X|r_W s_W)\, \delta_{XY}\, \delta_{WZ} \tag{3.115}$$

Dentro de cada família, podem distinguir-se vários métodos através da parametrização efectuada nos integrais e termos da energia total que não são anulados pela aplicação da aproximação ZDO. Esta parametrização pode pretender reproduzir resultados de cálculos *ab initio* (métodos CNDO/1 [45], CNDO/2 [46], INDO [47] e NDDO [48]) ou valores experimentais de certas propriedades moleculares (métodos MINDO/1 [49], MINDO/2 [50], MINDO/3 [51], PNDO [52], MNDO [53], AM1 [54] e PM3 [55]).

O método AM1, utilizado neste trabalho, baseia-se nas aproximações NDDO e é especialmente apropriado para o estudo de sistemas moleculares onde as pontes de hidrogénio desempenham um papel importante.

A aplicação das aproximações NDDO ao cálculo dos integrais monoelectrónicos conduz às seguintes equações,

$$h_{p_X p_X} = U_{p_X p_X} + \sum_{Y \neq X} V^Y_{p_X p_X} , \tag{3.116}$$

$$h_{p_X q_X} = \sum_{Y \neq X} V^Y_{p_X q_X} , \tag{3.117}$$

e

$$h_{p_X q_Y} = T_{p_X q_Y} + V^Y_{p_X q_Y} + V^X_{p_X q_Y} = \beta_{p_X q_Y} , \tag{3.118}$$

onde,

$$U_{p_X p_X} = T_{p_X p_X} + V^X_{p_X p_X} , \tag{3.119}$$

$$T_{pq} = \left\langle \chi_p \left| -\frac{1}{2}\nabla^2_1 \right| \chi_q \right\rangle , \tag{3.120}$$

e

$$V^Y_{pq} = - \left\langle \chi_p \left| \hat{V}_Y \right| \chi_q \right\rangle \tag{3.121}$$

A parametrização do método AM1 segue a seguinte metodologia:

a) Os integrais monocêntricos, monoelectrónicos ($U_{p_Xp_X}$) e bielectrónicos ($p_Xq_X|r_Xs_X$), são calculados pelas fórmulas de Oleari [56].

b) Os integrais monoelectrónicos bicêntricos coulômbicos ($V^Y_{p_Xq_X}$) são calculados assumindo que o efeito de blindagem exercido pelas orbitais atómicas do cerne B pode ser simulado pela respectiva distribuição de carga de valência s,

$$V^Y_{p_Xq_X} = - Z'_Y (p_Xq_X|s_Ys_Y) , \tag{3.122}$$

onde Z'_Y representa a carga do cerne Y e s_Y a sua orbital s de valência.

c) Os integrais monoelectrónicos bicêntricos de ressonância ($\beta_{p_Xq_Y}$) são considerados proporcionais aos correspondentes integrais de sobreposição ($S_{p_Xq_Y}$):,

$$\beta_{p_Xq_Y} = \frac{1}{2}(\beta^X_{p_X} + \beta^Y_{q_Y}) S_{p_Xq_Y} , \tag{3.123}$$

onde $\beta^X_{p_X}$ e $\beta^Y_{q_Y}$ são parâmetros empíricos característicos das funções de base χ_{p_X} e χ_{q_Y}, respectivamente.

d) Os expoentes das funções de base de Slater são tratados como parâmetros empíricos característicos de cada orbital atómica de valência a que estão associados. No entanto, para o caso dos átomos de carbono, azoto e oxigénio os expoentes das orbitais do tipo s e p tomam os mesmos valores.

e) Os integrais bielectrónicos bicêntricos ($p_Xq_X|r_Ys_Y$) são calculados através de uma expansão multipolar de distribuições electrónicas centradas nos átomos X e Y.

f) As energias de repulsão cerne-cerne (C_{XY}) são dadas por,

$$C_{XY} = Z'_X Z'_Y (s_Xs_X|s_Ys_Y) + f_{XY}(R_{XY}) , \tag{3.124}$$

onde $f_{XY}(R_{XY})$ é uma função empírica que depende dos átomos X e Y e da respectiva distância R_{XY}.

3.4 - Energias de estabilização

As reacções de formação de agregados moleculares a partir dos seus fragmentos apresentam um grande interesse em aplicações químicas e biológicas [57-73]. Numa reacção deste tipo a variação de energia envolvida (ΔE_k^{Γ}) é dada por,

$$\Delta E_k^{\Gamma} = E_k^{\Gamma} - \sum_{A=1}^{t} E_{k_A}^{\infty}(A) \ , \tag{3.125}$$

e é habitualmente designada por energia de estabilização do agregado molecular.

Na equação (3.125), t é o número de fragmentos, E_k^{Γ} é a energia do agregado molecular no estado electrónico k do tipo Γ e $E_{k_A}^{\infty}(A)$ é a energia do fragmento A isolado no estado electrónico k_A.

A determinação teórica de energias de estabilização não pode ser efectuada de uma forma exacta. Os métodos perturbacionais e variacionais, utilizados na resolução aproximada da equação de Schrödinger, são também os mais usados no cálculo destas grandezas físicas.

Os métodos perturbacionais são especialmente apropriados para descrever associações fracas onde a energia de estabilização representa uma fracção pequena da energia do agregado molecular [74-78]. De acordo com este formalismo, o hamiltoniano ($\hat{H}$) é escrito como,

$$\hat{H} = \hat{H}^0 + \hat{V}_{int.} \ , \tag{3.126}$$

onde $\hat{H}^0$ é o hamiltoniano não perturbado e $\hat{V}_{int.}$ é o operador de perturbação que contém todos os termos de interacção coulômbica que se podem estabelecer entre os diversos fragmentos.

O hamiltoniano não perturbado ($\hat{H}^0$) é definido como a soma dos hamiltonianos ($\hat{H}(A)$) correspondentes a cada fragmento isolado:

$$\hat{H}^0 = \sum_{A=1}^{t} \hat{H}(A) \tag{3.127}$$

onde ,

$$\hat{H}(A) = -\frac{1}{2}\sum_{i=1_A^{in}}^{1_A^f} \nabla_i^2 - \sum_{X=N_A}^{M_A}\sum_{i=1_A^{in}}^{1_A^f} \frac{Z_X}{\varepsilon r_{iX}} + \sum_{i=1_A^{in}}^{1_A^f}\sum_{j=1_A^{in}}^{i-1} \frac{1}{\varepsilon r_{ij}} + \sum_{X=N_A}^{M_A}\sum_{Y=N_A}^{X-1} \frac{Z_X Z_Y}{\varepsilon r_{XY}} \tag{3.128}$$

Na equação (3.128), $1_A^{in.}$ e 1_A^f representam os primeiro e último electrões e N_A e M_A os primeiro e último núcleos do fragmento A.

O operador de perturbação ($\hat{V}_{int.}$) é definido como,

$$\hat{V}_{int.} = \sum_{A=1}^{t}\sum_{B=1}^{A-1} \hat{V}(A,B) \ , \tag{3.129}$$

onde $\hat{V}(A,B)$ é o operador associado às interacções que se podem estabelecer entre os fragmentos A e B:

$$\hat{V}(A,B) = -\sum_{X=N_A}^{M_A}\sum_{j=1_B^{in}}^{1_B^f} \frac{Z_X}{\varepsilon r_{jX}} - \sum_{Y=N_B}^{M_B}\sum_{i=1_A^{in}}^{1_A^f} \frac{Z_Y}{\varepsilon r_{iY}} + \sum_{i=1_A^{in}}^{1_A^f}\sum_{j=1_B^{in}}^{1_B^f} \frac{1}{\varepsilon r_{ij}} + \sum_{X=N_A}^{M_A}\sum_{Y=N_B}^{M_B} \frac{Z_X Z_Y}{\varepsilon r_{XY}} \tag{3.130}$$

Num estado do tipo não ligante do agregado molecular, os fragmentos encontram-se a distância infinita e não há qualquer sobreposição entre as suas densidades electrónicas. Nestas condições, a função de onda ($\psi_k^{\infty}(q_1, q_2, \ldots, q_n)$) pode ser escrita como o produto de t funções de onda ($\psi_{k_A}^{\infty}(q_{1_A^{in}}, \ldots, q_{1_A^f})$) associadas a cada fragmento isolado A:

$$\psi_k^{\infty}(q_1,q_2,\ldots,q_n) = \prod_{A=1}^{t} \psi_{k_A}^{\infty}(q_{1_A^{in}},\ldots,q_{1_A^{f}}) \tag{3.131}$$

Nesta aproximação, os electrões associados a um fragmento A ($1_A^{in}, \ldots, 1_A^{f}$) são considerados distinguíveis em relação aos electrões associados a um fragmento B ($1_B^{in}, \ldots, 1_B^{f}$). Isto significa que a função de onda total ($\psi_k^{\infty}(q_1,q_2,\ldots,q_n)$) não obedece ao princípio da antissimetria em relação à permuta electrónica entre fragmentos, embora cada uma das suas componentes ($\psi_{k_A}^{\infty}(q_{1_A^{in}},\ldots,q_{1_A^{f}})$) o satisfaça em relação aos electrões que descreve ($1_A^{in}, \ldots, 1_A^{f}$).

Quando se pretende descrever agregados moleculares em estados do tipo não covalente ligante, uma função do tipo (3.131) constitui uma boa escolha para função de onda não perturbada:

$$\psi_k^{no\ 0}(q_1,q_2,\ldots,q_n) = \psi_k^{\infty}(q_1,q_2,\ldots,q_n) = \prod_{A=1}^{t} \psi_{k_A}^{\infty}(q_{1_A^{in}},\ldots,q_{1_A^{f}}) \tag{3.132}$$

O agregado molecular é então tratado em função da perturbação ($\hat{V}_{int.}$), sendo as sua função de onda e energia escritas na forma de expansões perturbacionais:

$$\psi_k^{no}(q_1,q_2,\ldots,q_n) = \psi_k^{no\ 0}(q_1,q_2,\ldots,q_n) + \psi_k^{no\ 1}(q_1,q_2,\ldots,q_n) + \ldots \tag{3.133}$$

$$\begin{aligned} E_k^{no} &= E_k^{no\ 0} + E_k^{no\ 1} + E_k^{no\ 2} + \ldots \\ &= \sum_{A=1}^{t} E_{k_A}^{\infty} + E_k^{no\ 1} + E_k^{no\ 2} + \ldots \end{aligned} \tag{3.134}$$

As séries perturbacionais (3.133) e (3.134) são geralmente truncadas a partir da 1ª ordem na função de onda e 2ª ordem na energia. As correcções correspondentes podem ser obtidas utilizando o formalismo da teoria de perturbações de Rayleigh-Schrödinger [74-77].

As correcções de 1ª ordem na função de onda são dadas por:

$$\psi_k^{no\ 1}(q_1, q_2, \ldots, q_n) = \sum_{j \neq k} c_{jk}^1 \, \psi_j^{\infty}(q_1, q_2, \ldots, q_n) \tag{3.135}$$

onde,

$$c_{jk}^1 = \left\langle \psi_j^{\infty} \middle| \hat{V}_{int.} \middle| \psi_k^{\infty} \right\rangle \tag{3.136}$$

A correcção de 1ª ordem na energia corresponde à chamada energia electrostática resultante das interacções entre os multipolos permanentes dos diversos fragmentos não perturbados. Estas correcções podem ser calculadas pela equação:

$$E_k^{no\ 1} = E_{k,es}^{no\ 1} = \left\langle \psi_k^{\infty} \middle| \hat{V}_{int.} \middle| \psi_k^{\infty} \right\rangle \tag{3.137}$$

Substituindo as equações (3.129) e (3.131) na equação (3.137) obtém-se,

$$E_k^{no\ 1} = E_{k,es}^{no\ 1} = \sum_{A=1}^{t} \sum_{B=1}^{A-1} E_k^{no\ 1}(A,B) = \sum_{A=1}^{t} \sum_{B=1}^{A-1} E_{k,es}^{no\ 1}(A,B) \ , \tag{3.138}$$

onde,

$$E_k^{no\ 1}(A,B) = E_{k,es}^{no\ 1}(A,B) = \left\langle \psi_{k_A}^{\infty} \, \psi_{k_B}^{\infty} \middle| \hat{V}(A,B) \middle| \psi_{k_A}^{\infty} \, \psi_{k_B}^{\infty} \right\rangle \tag{3.139}$$

A correcção de 2ª ordem na energia é dada por:

$$E_k^{no\ 2} = \sum_{j \neq k} \frac{\left| \left\langle \psi_j^{\infty} \middle| \hat{V}_{int.} \middle| \psi_k^{\infty} \right\rangle \right|^2}{E_k^{\infty} - E_j^{\infty}} \tag{3.140}$$

Substituindo as equações (3.129) e (3.131) na equação (3.140) obtém-se,

$$E_k^{no\ 2} = \sum_{A=1}^{t} \sum_{B=1}^{A-1} E_k^{no\ 2}(A,B) \ , \tag{3.141}$$

onde,

$$\begin{aligned} E_k^{no\ 2}(A,B) = & \sum_{j_A \neq k_A} \frac{\left| \left\langle \psi_{j_A}^{\infty} \psi_{k_B}^{\infty} \middle| \hat{V}(A,B) \middle| \psi_{k_A}^{\infty} \psi_{k_B}^{\infty} \right\rangle \right|^2}{E_{k_A}^{\infty} - E_{j_A}^{\infty}} \\ & + \sum_{j_B \neq k_B} \frac{\left| \left\langle \psi_{k_A}^{\infty} \psi_{j_B}^{\infty} \middle| \hat{V}(A,B) \middle| \psi_{k_A}^{\infty} \psi_{k_B}^{\infty} \right\rangle \right|^2}{E_{k_B}^{\infty} - E_{j_B}^{\infty}} \\ & + \sum_{j_A \neq k_A} \sum_{j_B \neq k_B} \frac{\left| \left\langle \psi_{j_A}^{\infty} \psi_{j_B}^{\infty} \middle| \hat{V}(A,B) \middle| \psi_{k_A}^{\infty} \psi_{k_B}^{\infty} \right\rangle \right|^2}{E_{k_A}^{\infty} + E_{k_B}^{\infty} - E_{j_A}^{\infty} - E_{j_B}^{\infty}} \end{aligned} \tag{3.142}$$

Na equação (3.142), o primeiro termo representa a energia de polarização do fragmento A induzida pelo fragmento B, o segundo termo representa a energia de polarização do fragmento B induzida pelo fragmento A e o terceiro termo representa a energia de dispersão associada aos fragmentos A e B. Consequentemente, as correcções de 2ª ordem na energia podem ser decompostas nas suas componentes de polarização e dispersão:

$$E_k^{no\ 2} = E_{k,pol}^{no\ 2} + E_{k,dis}^{no\ 2} \tag{3.143}$$

Comparando as equações (3.134), (3.137), (3.143) e (3.125), pode exprimir-se a energia de estabilização do agregado molecular em termos das respectivas correcções perturbacionais na energia:

$$\Delta E_k^{no} = E_k^{no\ 1} + E_k^{no\ 2} + \ldots = E_{k,el}^{no\ 1} + E_{k,pol}^{no\ 2} + E_{k,dis}^{no\ 2} + \ldots \tag{3.144}$$

A teoria de perturbações de Rayleigh-Schrödinger é inadequada para descrever agregados moleculares em estados covalentes ($\Gamma = ov$), visto assumir que a sobreposição

electrónica inter-fragmentos é nula [76-77]. Neste tipo de estados, não é razoável utilizar uma função de onda de ordem zero do tipo (3.132) e diversas metodologias foram propostas para incluir a permuta electrónica inter-fragmentos no formalismo perturbacional [78-89]. Estas teorias, embora sejam genericamente designadas por *teorias de perturbações com permuta (EPT)*, diferem bastante umas das outras e a comparação de resultados é dificultada por utilizarem a mesma designação para componentes que não são calculadas do mesmo modo. A título ilustrativo, apresenta-se em seguida o método proposto por Daudey *et al* [88], por este apresentar um formalismo especialmente apropriado para estabelecer comparações com os métodos variacionais [90].

O método de Daudey é aplicado a agregados moleculares com apenas dois fragmentos, dentro de um formalismo de Hartree-Fock monoconfiguracional, e a função de onda não perturbada,

$$\psi_k^{ov\,0}(q_1,q_2,\ldots,q_n) = \hat{A}(\psi_{k_A}^{\infty}(q_1,\ldots,q_{1_A^f})\,\psi_{k_B}^{\infty}(q_1,\ldots,q_{1_B^f}) , \tag{3.145}$$

é uma função completamente antissimetrizada, onde as orbitais moleculares dos fragmentos A e B não são sujeitas a ortogonalização mútua e o operador $\hat{A}$ está associado unicamente a permutas electrónicas entre os dois fragmentos.

Uma função do tipo (3.145) já não é função própria de um operador hamiltoniano não perturbado do tipo (3.127). O método de Daudey utiliza a partição de Epstein-Nesbet [90-91], em que o hamiltoniano não perturbado é definido como a parte diagonal do hamiltoniano perturbado na base $\left\{\psi_k^{ov\,0}\right\}$:

$$H_{lk}^0 = \left\langle \psi_l^{ov\,0} \middle| \hat{H}^0 \middle| \psi_k^{ov\,0} \right\rangle = \delta_{lk} \left\langle \psi_l^{ov\,0} \middle| \hat{H} \middle| \psi_k^{ov\,0} \right\rangle = \delta_{lk} H_{lk} \tag{3.146}$$

Os elementos do operador de perturbação podem ser obtidos conjugando as equações (3.126) e (3.146).

$$V_{lk} = \left\langle \psi_l^{ov\,0} \middle| \hat{V}_{int.} \middle| \psi_k^{ov\,0} \right\rangle = (1-\delta_{lk}) \left\langle \psi_l^{ov\,0} \middle| \hat{H} \middle| \psi_k^{ov\,0} \right\rangle = (1-\delta_{lk})\, H_{lk} \tag{3.147}$$

As correcções de 1ª ordem na função de onda são então calculadas em termos das configurações excitadas a partir da configuração de referência ($\psi_l^{ov\,0}$):

$$\psi_k^{ov\,1}(q_1,q_2,\ldots,q_n) = \sum_{l\neq k} c_{lk}^1\, \psi_l^{ov\,1}(q_1,q_2,\ldots,q_n) \tag{3.148}$$

onde,

$$c_{lk}^1 = \frac{V_{lk}}{H_{kk}^0 - H_{ll}^0} = \frac{H_{lk}}{H_{kk} - H_{ll}} \tag{3.149}$$

A energia não perturbada deixa de ser igual à soma da energia dos fragmentos isolados,

$$E_k^{ov\,0} = H_{kk}^0 = H_{kk}\,, \tag{3.150}$$

e pode ser decomposta como:

$$E_k^{ov\,0} = E_{k_A}^{\infty} + E_{k_B}^{\infty} + E_{k,es}^{no\,1}(A,B) + E_{k,ex}^{ov\,0}(A,B) \tag{3.151}$$

Os primeiros três termos da equação (3.151) são equivalentes aos obtidos pela teoria das perturbações de Rayleigh-Schrödinger, o quarto termo é a energia de permuta de ordem 0 e corresponde a uma correcção à energia electrostática ($E_{k,es}^{no\,1}(A,B)$) que deriva da sobreposição electrónica entre os fragmentos A e B ser não nula.

Neste formalismo, as correcções de 1ª ordem na energia são nulas:

$$E_k^{ov\,1} = V_{kk} = (1-\delta_{kk})\,H_{kk} = 0 \tag{3.152}$$

As correcções de 2ª ordem na energia são dadas pela equação:

$$E_k^{ov\,2} = \sum_{l \neq k} \frac{V_{lk}\, V_{kl}}{H_{kk}^0 - H_{ll}^0} = \sum_{l \neq k} \frac{|H_{lk}|^2}{H_{kk} - H_{ll}} = \sum_{l \neq k} \frac{\left\langle \psi_l^{ov\,0} \middle| \hat{H} \middle| \psi_k^{ov\,0} \right\rangle}{H_{kk} - H_{ll}} \qquad (3.153)$$

As correcções de 2ª ordem na energia podem ser decompostas em componentes associadas às configurações ($\psi_l^{ov\,1}(q_1, q_2, \ldots, q_n)$) excitadas a partir da configuração de referência:

$$E_k^{ov\,2} = E_{k,pol}^{ov\,2} + E_{k,ct}^{ov\,2} + E_{k,dis}^{ov\,2} + E_{k,corr}^{ov\,2} + E_{k,ex}^{ov\,2} \qquad (3.154)$$

A energia de polarização ($E_{k,pol}^{ov\,2}$) está associada a configurações resultantes de excitações simples dentro dos fragmentos A ou B ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{i_A^*}$ ou $\varphi_{j_B} \rightarrow \varphi_{j_B^*}$).

A energia de transferência de carga ($E_{k,ct}^{ov\,2}$) está associada a configurações resultantes de excitações de orbitais moleculares ocupadas de A para orbitais moleculares vazias de B ou vice-versa ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{j_B^*}$ ou $\varphi_{j_B} \rightarrow \varphi_{i_A^*}$).

A energia de dispersão ($E_{k,dis}^{ov\,2}$) está associada a configurações resultantes de excitações simultâneas de um electrão de A e de um electrão de B para as suas orbitais moleculares não ocupadas ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{i_A^*}$ e $\varphi_{j_B} \rightarrow \varphi_{j_B^*}$).

A energia de correlação ($E_{k,corr}^{ov\,2}$) está associada a configurações resultantes de duplas excitações num fragmento ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{i_A^*}$ e $\varphi_{i_A'} \rightarrow \varphi_{i_A'^*}$ ou $\varphi_{j_B} \rightarrow \varphi_{j_B^*}$ e $\varphi_{j_B'} \rightarrow \varphi_{j_B'^*}$). Outras contribuições para a energia de correlação estão associadas a configurações do tipo ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{i_A^*}$ e $\varphi_{i_A'} \rightarrow \varphi_{j_B^*}$) e ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{j_B^*}$ e $\varphi_{i_A'} \rightarrow \varphi_{j_B'^*}$) e termos análogos envolvendo a excitação de dois electrões de B.

A energia de permuta de 2ª ordem ($E_{k,ex}^{ov\,2}$) está associada a configurações resultantes de duplas excitações cruzadas do tipo ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{j_B^*}$ e $\varphi_{j_B} \rightarrow \varphi_{i_A^*}$) e termos análogos envolvendo excitações simultâneas de um electrão de A e de um electrão de B para orbitais moleculares não ocupadas do mesmo sub-espaço ($l \leftrightarrow \varphi_{i_A} \rightarrow \varphi_{i_A^*}$ e $\varphi_{j_B} \rightarrow \varphi_{i_A'^*}$ ou

$\varphi_{i_A} \rightarrow \varphi_{j_B^*}$ e $\varphi_{j_B} \rightarrow \varphi_{j_B^*}$).

A decomposição perturbacional acima descrita é apresentada de uma forma esquemática na figura (3.3).

Comparando as equações (3.151), (3.154) e (3.125), pode exprimir-se a energia de estabilização do agregado molecular em termos das respectivas correcções perturbacionais na energia do agregado molecular:

$$\Delta E_k^{ov} = E_{k,es}^{no\ 1} + E_{k,ex}^{ov\ 0} + E_{k,pol}^{ov\ 2} + E_{k,ct}^{ov\ 2} + E_{k,dis}^{ov\ 2} + E_{k,corr}^{ov\ 2} + E_{k,ex}^{ov\ 2} \qquad (3.155)$$

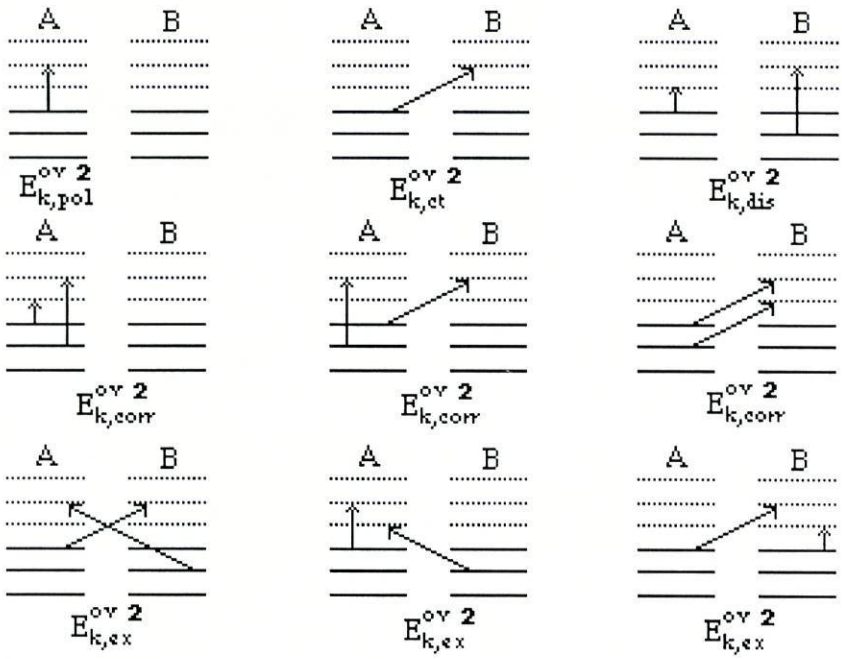


Figura 3.3. Representação esquemática da partição das correcções de 2ª ordem na energia de acordo com o formalismo de Daudey [88] ($E_{k,pol}^{ov\ 2}$, energia de polarização ; $E_{k,ct}^{ov\ 2}$, energia de transferência de carga ; $E_{k,dis}^{ov\ 2}$, energia de dispersão ; $E_{k,corr}^{ov\ 2}$, energia de correlação e $E_{k,ex}^{ov\ 2}$, energia de permuta).

Os métodos variacionais constituem uma alternativa ao formalismo perturbacional anteriormente apresentado. Uma vez estabelecida a forma do hamiltoniano do agregado molecular , equações (3.126 a 3.130), a energia de estabilização do mesmo é agora calculada como,

$$\Delta E_k^{\Gamma} = \left\langle \psi_k^{\Gamma} \middle| \hat{H} \middle| \psi_k^{\Gamma} \right\rangle - \sum_{A=1}^{t} \left\langle \psi_{k_A}^{\Gamma} \middle| \hat{H}(A) \middle| \psi_{k_A}^{\Gamma} \right\rangle , \qquad (3.156)$$

onde os valores esperados do hamiltoniano total ($\hat{H}$) e das suas componentes ($\hat{H}(A)$) são calculados variacionalmente utilizando métodos *ab initio* ou *semi-empíricos*.

Os métodos variacionais utilizam o chamado *modelo da super-molécula* porque o agregado global é tratado como se fosse constituído por uma única molécula.

As duas metodologias anteriormente descritas apresentam vantagens e inconvenientes quando aplicadas ao cálculo de energias de estabilização de agregados moleculares.

Os métodos perturbacionais apresentam as seguintes vantagens:

(i) As energias de estabilização são calculadas directamente evitando os inconvenientes do cálculo por diferença utilizado nos métodos variacionais.

(ii) A análise das diversas componentes da energia de estabilização pode ser feita em termos das propriedades dos fragmentos isolados.

(iii) Podem ser derivadas, como consequência de (ii), formas simplificadas para as diversas componentes da energia de estabilização com correspondência em termos clássicos.

Os principais inconvenientes destes métodos são:

(1) No formalismo perturbacional, admite-se que se conhecem as funções de onda exactas associadas aos fragmentos isolados ($\psi_{k_A}^{\infty}(q_{l_A^i}, \ldots, q_{l_A^f})$, $A = 1, \ldots, t$). No entanto, estas funções são normalmente aproximações das verdadeiras soluções da equação de Schrödinger para os respectivos sub-sistemas, cuja qualidade condiciona os resultados obtidos pelas expansões perturbacionais.

(2) Não existe um limite claramente definido entre as interacções fortes (dominantes em estados covalentes) e interacções fracas (típicas de estados não covalentes ligantes). Este facto introduz uma certa arbitrariedade em relação aos limites de aplicabilidade das diferentes metodologias perturbacionais.

(3) Os agregados moleculares em estados covalentes não são correctamente descritos pela teoria das perturbações de Rayleigh-Schrödinger. Nestes casos, os métodos EPT, cujo formalismo inclui explicitamente o tratamento da permuta electrónica inter-fragmentos, são muito mais apropriados. No entanto, estes métodos diferem bastante uns dos outros e a

comparação de resultados é dificultada por se utilizarem muitas vezes as mesmas designações para componentes que não são calculadas do mesmo modo.

(4) Na prática, a maior parte das expansões perturbacionais é truncada a partir da 2ª ordem, o que exclui termos de ordem superior que podem dar contribuições importantes para a energia de estabilização.

(5) No formalismo perturbacional, os fragmentos são considerados entidades rígidas. Isto significa que fenómenos como a relaxação da geometria interna dos fragmentos ou a troca de átomos entre eles não podem ser adequadamente descritos.

(6) Quando os estados electrónicos dos fragmentos isolados e do agregado molecular diferem muito entre si podem ocorrer problemas de convergência nas séries perturbacionais. Isto significa que estes métodos não são apropriados para descrever a formação de agregados moleculares que envolvam a alteração significativa na distribuição de carga electrónica dos respectivos fragmentos. Habitualmente, os métodos perturbacionais são utilizados no estudo de agregados moleculares e de fragmentos nos respectivos estados fundamentais.

(7) Na prática, é difícil aplicar os métodos perturbacionais a agregados moleculares com mais de dois fragmentos.

Os métodos variacionais apresentam as seguintes vantagens:

(i) A equação (3.156) é absolutamente geral, o que permite que seja aplicada para qualquer geometria do agregado molecular.

(ii) Os termos da equação (3.156) são calculados independentemente, o que significa que não existem quaisquer tipo de restrições em relação aos estados electrónicos em que se encontram o agregado molecular e respectivos fragmentos.

(iii) Os fragmentos podem ser entidades absolutamente flexíveis. Isto significa que fenómenos como a relaxação da geometria interna dos fragmentos ou a troca de átomos entre eles não podem agora ser incorporados no modelo.

(iv) Podem ser descritos, sem grandes dificuldades, agregados moleculares com mais de dois fragmentos.

Os principais inconvenientes destes métodos são:

(1) O uso da metodologia variacional garante que os diferentes termos da equação (3.156) sejam majorantes da energia exacta dos respectivos sistemas associados. O valor de ΔE_k^{Γ}, visto ser calculado por diferença, não corresponde obrigatoriamente a um majorante da

verdadeira energia de estabilização do agregado molecular. Além disso, ΔE_k^{Γ} é muito menor em valor absoluto do que os referidos termos, o que significa que estes têm de ser calculados com erros muito pequenos para que a precisão obtida na determinação da energia de estabilização seja razoável.

(2) As funções de onda $\left\{\psi_k^{\Gamma}(q_1, q_2, \ldots, q_n) \text{ e } (\psi_k^{\Gamma}(q_{1_A^i}, \ldots, q_{1_A^f}), A = 1, \ldots, t)\right\}$ utilizadas condicionam, em grande medida, a qualidade dos resultados obtidos. O uso de uma boa base de funções e, em certos casos, o tratamento explícito da correlação electrónica são muito importantes para obter valores precisos da energia de estabilização.

(3) As funções de onda são expandidas em termos de um conjunto incompleto de funções de base. Isto origina o chamado *erro de sobreposição de bases (BSSE)*, que resulta de o agregado molecular ser melhor descrito do que os respectivos fragmentos [93]. Consequentemente, os valores ΔE_k^{Γ} obtidos pela equação (3.156) reflectem uma estabilização artificial do agregado molecular. Boys e Bernardi [94].propuseram um método, habitualmente designado de *método de counterpoise (CP)*, para estimar este erro. Neste método, as energias dos fragmentos são calculadas utilizando a base de funções do agregado molecular na configuração para a qual se pretende calcular a energia de estabilização. Deste modo, aumenta-se a flexibilidade na descrição dos fragmentos baixando a sua energia e aumentando o valor de ΔE_k^{Γ}. O método CP não tem uma base física muito consistente e a sua aplicação prática conduz frequentemente a que o erro BSSE seja avaliado por excesso. Mais recentemente, diversos autores sugeriram métodos alternativos para calcular este erro [95-96].

(4) A energia de estabilização obtida a partir da equação (3.156) não pode ser decomposta, de uma forma imediata, em componentes associadas com termos clássicos.

3.5 - Decomposição das energias molecular e de estabilização

3.5.1- Considerações gerais

A interpretação física de fenómenos como a ligação química, a associação molecular e as reacções de transferência de protão, requer a partição da energia (ou da energia de

estabilização) em componentes associadas aos diversos tipos de interacções envolvidas. Nesta secção, apresentam-se várias metodologias que foram propostas para efectuar este tipo de partições.

Uma das metodologias propostas consiste na expansão perturbacional da energia de estabilização do agregado molecular em função das suas componentes. Na secção (3.4) foram apresentados métodos perturbacionais deste tipo diferentes. Uma alternativa, dentro do formalismo RHF, foi sugerida por Morokuma *et al* [97-98]. De acordo com este método, certos blocos da matriz de Fock do agregado molecular são deliberadamente anulados e a matriz resultante é diagonalizada iterativamente originando as diversas componentes da energia de estabilização. Os métodos perturbacional e variacional de Morokuma são bastante úteis na interpretação física das interacções em termos das propriedades electrónicas dos fragmentos isolados. Contudo, estes métodos apresentam alguma dificuldade em incorporar fenómenos como a relaxação da geometria interna dos fragmentos ou a troca de átomos entre eles que podem ser muito importantes para uma correcta interpretação física da formação de um agregado molecular a partir dos seus fragmentos.

Uma terceira aproximação ao problema é baseada no facto da energia de um agregado molecular ser completamente determinada pelas suas matrizes densidade e densidade de par (equações 3.37 a 3.40). Utilizando a aproximação de Roothaan (equação 3.44), a energia e as populações electrónicas e electrónicas de par podem então ser decompostas em contribuições associadas com as funções de base atómicas (equações 3.47 e 3.54). Esta metodologia de decomposição pode também ser alargada, dentro da aproximação da super-molécula, à energia de estabilização. Neste contexto, foram propostos vários métodos para decomposição das energias molecular e de estabilização em componentes atómicas [99-102].

No seu estudo sobre a natureza da ligação química, Ruedenberg [99] propôs um método para decompor a energia de estabilização combinando componentes espaciais e componentes associadas aos diversos tipos de interacções que envolvem as matrizes densidade e densidade de par.

Outra metodologia para decompor a energia molecular, usando a matriz densidade e os integrais monoelectrónicos e bielectrónicos, foi proposta por Clementi [100]. Dentro deste formalismo, são consideradas componentes intra-atómicas e inter-atómicas (bicêntricas e multicêntricas).

Kollmar [101] sugeriu um método alternativo, utilizando os integrais monoelectrónicos e as matrizes densidade e de Fock, para efectuar este tipo de partições em que apenas são consideradas componentes efectivas intra-atómicas e inter-atómicas bicêntricas.

Bofill *et al* [102] sugeriram um formalismo segundo o qual a energia molecular é decomposta usando a matriz densidade de par e os integrais monoelectrónicos e bielectrónicos. Isto permite que a energia e estrutura de uma molécula sejam analisadas tendo em conta apenas as populações de electrónicas de par.

Neste trabalho, é proposto um novo método de decomposição [103], dentro de um formalismo *ab initio* monoconfiguracional ou multiconfiguracional, baseado na decomposição espacial das matrizes densidade e densidade de par e no formalismo dos integrais indistinguíveis. De acordo com este método, as energias molecular e de estabilização de um agregado são decompostas em componentes associadas com os respectivos fragmentos. Isto permite uma interpretação física flexível da interacção dentro de cada sistema particular.

Os métodos de decomposição dentro do formalismo variacional [97-103] apresentam as características genéricas referidos na secção (3.4). No entanto, cada um destes métodos baseia-se em abordagens diferenciadas do problema em estudo e apresenta vantagens e inconvenientes específicos. Esta problemática será discutida nas próximas secções.

3.5.2- Método de Morokuma

O método de Morokuma [97-98] foi originalmente formulado, dentro da aproximação RHF, para dímeros constituídos por fragmentos rígidos. Neste modelo, as funções orbitais (η_i) de um dímero AB são expandidas em termos das funções orbitais dos fragmentos isolados A (η_i^A) e B (η_i^B),

$$\eta_i = \sum_{l}^{(A)} c_{li}\, \eta_l^A + \sum_{\mu}^{(B)} c_{\mu i}\, \eta_\mu^B \qquad (3.157)$$

onde os símbolos (A) e (B) significam que os correspondentes somatórios são extensíveis a todas as funções orbitais dos fragmentos isolados A e B, respectivamente. Nesta notação, os

índices l, m e g estão associados a funções orbitais do fragmento A isolado, os índices μ, υ e σ estão associados a funções orbitais do fragmento B isolado, os índices p e q estão associados a funções orbitais dos fragmento A ou B isolados e os índices i e j estão associados a funções orbitais do dímero AB.

Neste contexto, as funções orbitais do dímero e respectivas energias podem ser calculadas por equações do tipo Hartree-Fock-Roothaan (equação 3.79), em que os elementos das matrizes envolvidas são agora expressos em termos das funções orbitais dos fragmentos isolados,

$$\sum_{q}^{\{(A)+(B)\}} c_{pi} \left(F_{pq}^{\Gamma,k} - e_i S_{pq}\right) = 0 \text{ ou } (\mathbf{F}^{\Gamma,k} - \boldsymbol{E}\,\mathbf{S})\,\mathbf{C} = 0\,,\ p \in \{(A)+(B)\}\,, \qquad (3.158)$$

onde a notação {(A)+(B)} significa que os correspondentes somatórios e índices são extensíveis ao conjunto de todas as funções orbitais dos dois fragmentos (A e B) isolados.

A energia do dímero pode então ser calculada de uma forma análoga à utilizada na equação (3.81):

$$E_k^{\Gamma} = E_k^{\Gamma(0)} + \frac{1}{2} \sum_{p}^{\{(A)+(B)\}} \sum_{q}^{\{(A)+(B)\}} P_{pq}^{\Gamma,k} \left(h_{pq} + F_{pq}^{\Gamma,k}\right) \qquad (3.159)$$

A energia de estabilização do dímero pode ser calculada comparando as equações (3.159) e (3.156),

$$\Delta E_k^{\Gamma} = \frac{1}{2} \sum_{p}^{\{(A)+(B)\}} \sum_{q}^{\{(A)+(B)\}} P_{pq}^{\Gamma,k} \left(h_{pq} + F_{pq}^{\Gamma,k}\right) + \sum_{X=N_A}^{M_A} \sum_{Y=N_B}^{M_B} \frac{Z_X Z_Y}{\varepsilon\, r_{XY}} - E_k^{\infty(1-2)}, \qquad (3.160)$$

onde $E_k^{\infty(1+2)}$ representa a soma das componentes monoelectrónica e bielectrónica da energia do sistema quando os fragmentos A e B se encontram a distância infinita:

$$E_k^{\infty(1+2)} = E_{k_A}^{\infty(1)}(A) + E_{k_A}^{\infty(2)}(A) + E_{k_B}^{\infty(1)}(B) + E_{k_B}^{\infty(2)}(B) =$$
$$= \left\langle \psi_{k_A}^{\infty} \middle| \hat{H}(A) \middle| \psi_{k_A}^{\infty} \right\rangle + \left\langle \psi_{k_B}^{\infty} \middle| \hat{H}(B) \middle| \psi_{k_B}^{\infty} \right\rangle - E_{k_A}^{\infty(0)}(A) - E_{k_B}^{\infty(0)}(B) \qquad (3.161)$$

A quantidade matricial $(\mathbf{F}^{\Gamma,k} - \boldsymbol{E}\,\mathbf{S})$, utilizada na equação (3.158), é então decomposta da seguinte forma,

$$(\mathbf{F}^{\Gamma,k} - \boldsymbol{E}\,\mathbf{S}) = (\mathbf{F}^{\infty,k} - \boldsymbol{E}\,\mathbf{S}) + \Sigma \ , \qquad (3.162)$$

onde é $\mathbf{F}^{\infty,k}$ a matriz de Fock quando os fragmentos A e B se encontram a distância infinita, **I** é a matriz identidade e Σ é a matriz de interacção. Os elementos destas matrizes são definidos como:

$$F_{lm}^{\infty,k} = e_l^{\infty}\,\delta_{lm} \ ; \ F_{\mu\nu}^{\infty,k} = e_{\mu}^{\infty}\,\delta_{\mu\nu} \ ; \ F_{l\mu}^{\infty,k} = 0 \qquad (3.163)$$

$$\Sigma_{lm} = -\left\langle \eta_l^A \middle| \sum_{X=N_B}^{M_B} \frac{Z_X}{\varepsilon\, r_{lX}} \middle| \eta_m^A \right\rangle + \left\langle \eta_l^A \middle| \sum_{j}^{(AB)_{oc.}} (2\hat{J}_j - \hat{K}_j) \middle| \eta_m^A \right\rangle - \left\langle \eta_l^A \middle| \sum_{g}^{(A)_{oc.}} (2\hat{J}_g - \hat{K}_g) \middle| \eta_m^A \right\rangle \delta_{lm} \qquad (3.164)$$

$$\Sigma_{\mu\nu} = -\left\langle \eta_\mu^B \middle| \sum_{X=N_A}^{M_A} \frac{Z_X}{\varepsilon\, r_{lX}} \middle| \eta_\nu^B \right\rangle + \left\langle \eta_\mu^B \middle| \sum_{j}^{(AB)_{oc.}} (2\hat{J}_j - \hat{K}_j) \middle| \eta_\nu^B \right\rangle - \left\langle \eta_\mu^B \middle| \sum_{\sigma}^{(B)_{oc.}} (2\hat{J}_\sigma - \hat{K}_\sigma) \middle| \eta_\nu^B \right\rangle \delta_{\mu\nu} \qquad (3.165)$$

$$\Sigma_{l\mu} = -(\boldsymbol{E}\mathbf{S})_{l\mu} - \left\langle \eta_l^A \middle| \frac{1}{2}\nabla_1^2 + \sum_{X=N_A}^{M_A} \frac{Z_X}{\varepsilon\, r_{lX}} + \sum_{Y=N_B}^{M_B} \frac{Z_Y}{\varepsilon\, r_{lY}} \middle| \eta_\mu^B \right\rangle + \left\langle \eta_l^A \middle| \sum_{j}^{(AB)_{oc.}} (2\hat{J}_j - \hat{K}_j \middle| \eta_\mu^B \right\rangle \qquad (3.166)$$

Nas equações (3.163) a (3.166), e_l^{∞} e e_{μ}^{∞} são as energias associadas às funções orbitais η_i^A e η_i^B, $\hat{J}_j$ e $\hat{K}_j$ são os operadores de Coulomb e permuta associados a uma função orbital do dímero, $\hat{J}_q$ e $\hat{K}_q$ são os operadores de Coulomb e permuta associados a uma função orbital do fragmento isolado A e $\hat{J}_\sigma$ e $\hat{K}_\sigma$ têm o mesmo significado relativamente ao fragmento B.

As notações $(A)_{oc.}$, $(B)_{oc.}$ e $(AB)_{oc.}$ significam que os respectivos somatórios são extensíveis a todas as funções orbitais ocupadas do fragmento A isolado, do fragmento B isolado e do dímero AB, respectivamente.

A matriz de interacção é então subdividida em blocos tendo em conta as funções orbitais envolvidas (ver figura 3.4).

	A,ocupada	A,vazia	B,ocupada	B,vazia
A,ocupada	$\sum^{esx}$	$\sum^{plx}$	$\sum^{ex'}$	$\sum^{ct}$
A,vazia	$\sum^{plx}$	$\sum^{esx}$	$\sum^{ct}$	$\sum^{ex'}$
B,ocupada	$\sum^{ex'}$	$\sum^{ct}$	$\sum^{esx}$	$\sum^{plx}$
B,vazia	$\sum^{ct}$	$\sum^{ex'}$	$\sum^{plx}$	$\sum^{esx}$

Figura 3.4 Partição da matriz de interacção dentro do formalismo de Morokuma [98] (Σ^{esx}, bloco electrostático e de permuta ; Σ^{plx}, bloco de polarização e de permuta ; $\Sigma^{ex'}$, bloco de permuta ; Σ^{ct}, bloco de transferência de carga).

Os blocos diagonais Σ^{esx} representam as interacções envolvendo funções orbitais ocupadas de um mesmo fragmento. Os termos envolvendo funções orbitais vazias de um mesmo fragmento não contribuem para a energia mas são incluídos no formalismo por consistência. Estes blocos incluem contribuições de permuta (associadas ao operador de permuta $\hat{K}_j$) e contribuições electrostáticas.

Os blocos Σ^{plx} representam as interacções envolvendo funções orbitais ocupadas e vazias de um mesmo fragmento. Estes blocos incluem contribuições de permuta (associadas ao operador de permuta $\hat{K}_j$) e contribuições de polarização.

Os blocos $\Sigma^{ex'}$ representam as interacções envolvendo funções orbitais ocupadas de dois fragmentos diferentes. Os termos envolvendo funções orbitais vazias de dois fragmentos

diferentes não contribuem para a energia mas são incluídos no formalismo por consistência. Estes blocos incluem somente contribuições de permuta.

Os blocos Σ^{ct} representam as interacções envolvendo funções orbitais ocupadas e vazias de dois fragmentos diferentes. Estes blocos incluem apenas contribuições de transferência de carga.

A determinação das diversas componentes da energia envolve a resolução de equações matriciais do tipo (3.162) considerando apenas determinados blocos da matriz de interacção e anulando os restantes.

Os blocos Σ^{esx} são os primeiros a serem considerados dada a sua natureza diagonal. As determinação das contribuições associadas a estes blocos envolve a resolução das seguintes equações matriciais:

$$\left\{\left(\mathbf{F}^{\infty,k} - \boldsymbol{E}\,\mathbf{S}\right) + \Sigma^{esx}\right\}\mathbf{C}^{esx} = 0 \qquad (3.167)$$

A correspondente energia é dada por,

$$E^{\Gamma}_{k,esx} = \sum_{i}^{(AB)_{oc}} \sum_{p}^{(esx)} \sum_{q}^{(esx)} c_{pi}^{esx*} \left(F_{pq}^{\Gamma,k\,esx} - h_{pq}^{esx}\right) c_{pi}^{esx} \,, \qquad (3.168)$$

onde a notação (esx) indica que os somatórios e elementos de matrizes correspondentes apenas incluem as funções orbitais associadas ao bloco Σ^{esx}. Dada a natureza deste bloco, a energia ($E^{\Gamma}_{k,esx}$) é obtida a partir das funções orbitais dos fragmentos isolados (i.e. $c_{pi}^{esx}=1$).

Os restantes blocos não podem ser tratados de uma forma independente em relação ao bloco Σ^{esx} dada a sua natureza diagonal. A determinação das correspondentes contribuições é efectuada através da resolução das seguintes equações matriciais:

$$\left\{\left(\mathbf{F}^{\infty,k} - \boldsymbol{E}\,\mathbf{S}\right) + \Sigma^{esx} + \Sigma^{y}\right\}\mathbf{C}^{esx+y} = 0\,,\ y = \text{(ex', plx ou ct)} \qquad (3.169)$$

A correspondente energia é obtida por diferença em relação a $E^{\Gamma}_{k,esx}$,

$$E^{\Gamma}_{k,y} = E^{\Gamma}_{k,esx+y} - E^{\Gamma}_{k,esx} \text{ , } y = (ex', plx \text{ ou } ct) \text{ ,} \tag{3.170}$$

onde

$$E^{\Gamma}_{k,esx+y} = \sum_{i}^{(AB)_{oc.}} \sum_{p}^{(esx+y)} \sum_{q}^{(esx+y)} c_{pi}^{esx+y*} \, (F_{pq}^{\Gamma,k\,esx+y} - h_{pq}^{esx+y}) \, c_{pi}^{esx+y} \tag{3.171}$$

As contribuições dos diversos blocos para a energia de estabilização podem ser calculados através da equação:

$$\Delta E^{\Gamma}_{k,x} = E^{\Gamma}_{k,x} - E^{\infty(1+2)}_{k} \text{ , } x = (esx, ex', plx \text{ ou } ct) \tag{3.172}$$

No método de Morokuma, as componentes da energia de estabilização são habitualmente rearranjadas de forma a utilizar uma formalismo comparável ao dos métodos perturbacionais apresentados na secção (3.4):

$$\Delta E^{\Gamma}_{k} = \Delta E^{\Gamma}_{k.es} + \Delta E^{\Gamma}_{k,ex} + \Delta E^{\Gamma}_{k,pol} + \Delta E^{\Gamma}_{k,expl} + \Delta E^{\Gamma}_{k,ct} + \Delta E^{\Gamma}_{k,mix} \tag{3.173}$$

Na equação (3.173), $\Delta E^{\Gamma}_{k,es}$ representa a componente electrostática a partir da equação (3.167) ignorando o operador de permuta ($\hat{K}_j$) em Σ^{esx}. Nesta componente deve também ser incluída a energia de repulsão nuclear inter-fragmento que corresponde ao 2º termo da equação (3.160).

A componente de permuta repulsiva ($\Delta E^{\Gamma}_{k,ex}$) é obtida através da equação:

$$\Delta E^{\Gamma}_{k,ex} = \Delta E^{\Gamma}_{k,esx} - \Delta E^{\Gamma}_{k,el} \tag{3.174}$$

A componente de polarização ($\Delta E^{\Gamma}_{k,pol}$) é obtida directamente dos blocos Σ^{plx} ignorando o operador de permuta intermolecular ($\hat{K}_j$) nas equações (3.164) e (3.165). A componente de permuta devida à polarização ($\Delta E^{\Gamma}_{k,expl}$) é então obtida por diferença:

$$\Delta E^{\Gamma}_{k,expl} = \Delta E^{\Gamma}_{k,plx} - \Delta E^{\Gamma}_{k,pol} \qquad (3.175)$$

Esta componente por vezes é considerada separadamente e outra vezes é incorporada nas componentes de polarização ou de permuta repulsiva.

A componente de transferência de carga é $(\Delta E^{\Gamma}_{k,ct})$ obtida directamente dos blocos Σ^{ct}.

A componente residual $(\Delta E^{\Gamma}_{k,mix})$ é calculada por diferença entre a energia de estabilização obtida variacionalmente e a soma das cinco componentes anteriormente apresentadas,

$$\Delta E^{\Gamma}_{k,mix} = \Delta E^{\Gamma}_{k} - \sum_{x} \Delta E^{\Gamma}_{k,x} \ , \ x = (es, ex, pol, expl \text{ e } ct) \ , \qquad (3.176)$$

e resulta do acoplamento entre componentes derivado do uso simultâneo de alguns blocos matriciais.

O método de Morokuma apresenta as seguintes vantagens específicas dentro da família dos métodos variacionais:

(i) A metodologia de decomposição é aplicável tanto à energia molecular como à energia de estabilização.

(ii) Em analogia com os métodos perturbacionais, a energia de estabilização é decomposta em componentes com correspondência em termos clássicos.

(iii) Segundo a metodologia proposta por Cammi e Tomasi [104], as *correcções de counterpoise* podem ser introduzidas no modelo.

(iv) O formalismo é expresso em termos de equações do tipo Hartree-Fock-Roothaan o que facilita a implementação do algoritmo de decomposição.

Os principais inconvenientes deste método são:

(1) O método é formulado dentro da aproximação RHF o que não permita avaliar a contribuição da correlação electrónica para a energia de estabilização.

(2) As componentes obtidas através da equação (3.173) apresentam uma dependência significativa da base de funções utilizada.

(3) O modelo apresenta dificuldades para descrever a relaxação da geometria interna dos fragmentos e a troca de átomos entre eles. Diversas sugestões têm sido apresentadas para introduzir uma componente de deformação no cálculo da energia de estabilização [105-106]. No entanto, esta componente é sempre definida de uma forma basicamente exterior ao formalismo de decomposição de Morokuma que assume os fragmentos como entidades rígidas.

(4) O uso simultâneo de alguns termos matriciais no cálculo das componentes da energia de estabilização origina o termo residual $(\Delta E^{\Gamma}_{k,mix})$. Em estados não covalentes ligantes (Γ=no), este termo é desprezável e a energia de estabilização pode ser decomposta nas restante cinco componentes, que têm uma interpretação física bem definida. No entanto, em estados covalentes (Γ=no), a componente residual não pode ser desprezada o que dificulta a interpretação física da interacção. Nagase *et al* [105] apresentaram uma metodologia de interpretação de $(\Delta E^{\Gamma}_{k,mix})$ com base em diferentes acoplamentos envolvendo as restantes componentes.

(5) O método de decomposição foi apenas formulado para dímeros.

3.5.3- Método de Ruedenberg

O método de Ruedenberg [99] permite efectuar a partição das energias molecular e de estabilização em componentes espaciais e componentes associadas às interacções responsáveis pela formação das ligações químicas.

Neste método utilizam-se as matrizes densidade $\rho^{\Gamma}_{k}(X'_1|X_1)$ (equação 3.45) e densidade de par $\pi^{\Gamma}_{k}(X'_1X'_2|X_1X_2)$ (equação 3.46) no espaço cartesiano e uma base de funções atómicas de valência. Neste contexto, a matriz densidade é decomposta segundo a equação,

$$\rho^{\Gamma}_{k}(X'_1|X_1) = \rho^{\Gamma}_{k,cl}(X'_1|X_1) + \rho^{\Gamma}_{k,I}(X'_1|X_1) \ , \qquad (3.177)$$

onde $\rho^{\Gamma}_{k,cl}(X'_1|X_1)$ representa a matriz densidade quase-clássica cuja integração em todo o espaço cartesiano corresponde à população electrónica do sistema molecular,

$$\int \rho^{\Gamma}_{k,cl}(X_1)\,dV_1 = \int_{X'_1=X_1} \rho^{\Gamma}_{k,cl}(X'_1|X_1)\,dV_1 = n \ , \tag{3.178}$$

e $\rho^{\Gamma}_{k,I}(X'_1|X_1)$, a matriz de interferência, é uma componente quântica que introduz correcções locais na densidade quase-clássica sem alterar a população electrónica total:

$$\int \rho^{\Gamma}_{k,I}(X_1)\,dV_1 = \int_{X'_1=X_1} \rho^{\Gamma}_{k,I}(X'_1|X_1)\,dV_1 = 0 \ , \tag{3.179}$$

Dentro da aproximação de Roothaan, a matriz densidade quase-clássica pode ser decomposta em componentes atómicas,

$$\rho^{\Gamma}_{k,cl}(X'_1|X_1) = \sum_{W=1}^{N} \rho_k^{\Gamma\ Mull.}(W) \ , \tag{3.180}$$

onde $\rho_k^{\Gamma\ Mull.}(W)$ é dado por:

$$\rho_k^{\Gamma\ Mull.}(W) = \rho_k^{\Gamma}(W|W) + \sum_{Z\neq W=1}^{N} \rho_k^{\Gamma}(W|Z) \tag{3.181}$$

Na equação (3.181), $\rho_k^{\Gamma}(W|W)$ e $\rho_k^{\Gamma}(W|Z)$ representam, respectivamente, as componentes intrínseca e de sobreposição com o átomo Z de $\rho_k^{\Gamma\ Mull.}(W)$, dadas por,

$$\rho_k^{\Gamma}(W|W) = \sum_{p=n_W}^{m_W} P_{pp}^{\Gamma,k}\, \chi_p^*(X_1')\, \chi_p(X_1) \,, \tag{3.182}$$

$$\rho_k^{\Gamma}(W|Z) = \sum_{p=n_W}^{m_W}\sum_{q=n_Z}^{m_Z} P_{pq}^{\Gamma,k}\, \chi_p^*(X_1')\, \chi_p(X_1) \,, \tag{3.183}$$

onde n_W e m_W representam a primeira e a última função atómica de valência do átomo W, respectivamente.

As partições (3.180) a (3.183) baseiam-se na chamada aproximação de Mulliken [107], que, de uma forma arbitrária, postula que as populações electrónicas de sobreposição devem ser divididas equalitariamente entre os átomos envolvidos.

A matriz de interferência pode ser decomposta em componentes diatómicas [99],

$$\rho_{k,I}^{\Gamma}(X_1'|X_1) = \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N} \rho_k^{\Gamma}(W/Z) \,, \tag{3.184}$$

onde

$$\rho_k^{\Gamma}(W/Z) = \sum_{p=n_W}^{m_W}\sum_{q=n_Z}^{m_Z} P_{pq}^{\Gamma,k} \left\langle \chi_p, \chi_q \right\rangle \tag{3.185}$$

Na equação (3.185), a quantidade $\left\langle \chi_p, \chi_q \right\rangle$ é definida como:

$$\left\langle \chi_p, \chi_q \right\rangle = \chi_p^*(X_1')\, \chi_q(X_1) - \frac{1}{2} S_{pq} \left\{ \chi_p^*(X_1')\, \chi_p(X_1) + \chi_q^*(X_1')\, \chi_q(X_1) \right\} \tag{3.186}$$

A relação de conservação (3.24) pode agora ser expressa em termos de componentes atómicas:

$$\sum_{W=1}^{N} \int_{X_1'=X_1} \rho_k^{\Gamma\, Mull.}(W)\, dV_1 = n \tag{3.187}$$

Dentro da aproximação de Roothaan, também é possível decompor a matriz densidade de par em componentes atómicas,

$$\pi_k^{\Gamma}(X_1'X_2'|X_1X_2) = 2\sum_{W=1}^{N}\sum_{Z=1}^{N}\sum_{T=1}^{N}\sum_{V=1}^{N} \pi_k^{\Gamma}(WZ|TV)\ , \tag{3.188}$$

onde

$$\pi_k^{\Gamma}(WZ|TV) = \frac{1}{2}\sum_{p=n_W}^{m_W}\sum_{q=n_Z}^{m_Z}\sum_{r=n_T}^{m_T}\sum_{s=n_V}^{m_V} D_{pqrs}^{\Gamma,k}\, \chi_p{}^*(X_1')\, \chi_q(X_1)\, \chi_r{}^*(X_2')\, \chi_s(X_2) \tag{3.189}$$

A relação de conservação (3.25) pode agora ser expressa em termos de componentes atómicas:

$$\sum_{W=1}^{N}\sum_{Z=1}^{N}\sum_{T=1}^{N}\sum_{V=1}^{N} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \pi_k^{\Gamma}(WZ|TV)\, dV_1\, dV_2 = \frac{n(n-1)}{2} \tag{3.190}$$

A equação (3.188) pode ser rearranjada como,

$$\pi_k^{\Gamma}(X_1'X_2'|X_1X_2) = 2\left\{\pi_{k,v}^{\Gamma}(X_1'X_2'|X_1X_2) + \pi_{k,I}^{\Gamma}(X_1'X_2'|X_1X_2) + \pi_{k,II}^{\Gamma}(X_1'X_2'|X_1X_2)\right\}, \tag{3.191}$$

onde $\pi_{k,v}^{\Gamma}(X_1'X_2'|X_1X_2)$, $\pi_{k,I}^{\Gamma}(X_1'X_2'|X_1X_2)$ e $\pi_{k,II}^{\Gamma}(X_1'X_2'|X_1X_2)$ representam respectivamente as matrizes de valência, de interferência primária e de interferência secundária, dadas por:

$$\pi^{\Gamma}_{k,v}(X'_1X'_2|X_1X_2) = \sum_{W=1}^{N}\sum_{Z=1}^{N}\pi^{\Gamma}_{k}(WW|ZZ) \tag{3.192}$$

$$\pi^{\Gamma}_{k,I}(X'_1X'_2|X_1X_2) = 2\sum_{W=1}^{N}\sum_{Z=1}^{N}\sum_{T\neq Z=1}^{N}\pi^{\Gamma}_{k}(WW|ZT) \tag{3.193}$$

$$\pi^{\Gamma}_{k,II}(X'_1X'_2|X_1X_2) = \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\sum_{T=1}^{N}\sum_{V\neq T=1}^{N}\pi^{\Gamma}_{k}(WZ|TV) \tag{3.194}$$

A relação de conservação (3.190) pode então ser reescrita como:

$$\iint_{\substack{X'_1=X_1\\X'_2=X_2}}\left\{\pi^{\Gamma}_{k,v}(X'_1X'_2|X_1X_2)+\pi^{\Gamma}_{k,I}(X'_1X'_2|X_1X_2)+\pi^{\Gamma}_{k,II}(X'_1X'_2|X_1X_2)\right\}\,dV_1dV_2=\frac{n(n-1)}{2} \tag{3.195}$$

Dentro deste formalismo, a energia pode ser calculada através da equação,

$$E^{\Gamma}_{k} = E^{\Gamma}_{k,cl} + E^{\Gamma}_{k,I} + E^{\Gamma}_{k,II}\ , \tag{3.196}$$

onde $E^{\Gamma}_{k,cl}$, $E^{\Gamma}_{k,I}$ e $E^{\Gamma}_{k,II}$ são respectivamente as componentes quase clássica, de interferência primária e de interferência secundária, dadas por:

$$E^{\Gamma}_{k,cl} = \sum_{W=1}^{N}\sum_{Z=1}^{W-1}\frac{Z_W\,Z_Z}{\varepsilon\, r_{WZ}} + \int_{X'_1=X_1}\hat{h}_1\,\rho^{\Gamma}_{k,cl}(X'_1|X_1)\,dV_1 + \iint_{\substack{X'_1=X_1\\X'_2=X_2}}\hat{h}_{12}\,\pi^{\Gamma}_{k,v}(X'_1X'_2|X_1X_2)\,dV_1\,dV_2$$

$$= \sum_{W=1}^{N}\sum_{Z=1}^{W-1}\frac{Z_WZ_Z}{\varepsilon\, r_{WZ}} + \sum_{W=1}^{N}\int_{X'_1=X_1}\hat{h}_1\rho^{\Gamma\ Mull.}_{k}(W)\,dV_1 + \sum_{W=1}^{N}\sum_{Z=1}^{N}\iint_{\substack{X'_1=X_1\\X'_2=X_2}}\hat{h}_{12}\pi^{\Gamma}_{k}(WW|ZZ)\,dV_1\,dV_2 \tag{3.197}$$

$$E^{\Gamma}_{k,I} = \int_{X'_1=X_1} \hat{h}_1 \, \rho^{\Gamma}_{k,I}(X'_1|X_1)\, dV_1 + \iint_{\substack{X'_1=X_1 \\ X'_2=X_2}} \hat{h}_{12} \, \pi^{\Gamma}_{k,I}(X'_1X'_2|X_1X_2)\, dV_1\, dV_2$$

$$= \sum_{W=1}^{N} \sum_{Z\neq W=1}^{N} \int_{X'_1=X_1} \hat{h}_1 \rho^{\Gamma}_k(W/Z)\, dV_1 + 2 \sum_{W=1}^{N} \sum_{Z=1}^{N} \sum_{T\neq Z=1}^{N} \iint_{\substack{X'_1=X_1 \\ X'_2=X_2}} \hat{h}_{12} \pi^{\Gamma}_k(WW|ZT)\, dV_1\, dV_2 \quad (3.198)$$

$$E^{\Gamma}_{k,II} = \iint_{\substack{X'_1=X_1 \\ X'_2=X_2}} \hat{h}_{12} \, \pi^{\Gamma}_{k,II}(X'_1X'_2|X_1X_2)\, dV_1\, dV_2$$

$$= \sum_{W=1}^{N} \sum_{Z\neq W=1}^{N} \sum_{T=1}^{N} \sum_{V\neq T=1}^{N} \iint_{\substack{X'_1=X_1 \\ X'_2=X_2}} \hat{h}_{12} \, \pi^{\Gamma}_k(WZ|TV)\, dV_1\, dV_2 \quad (3.199)$$

No modelo de Ruedenberg, a formação de um agregado molecular a partir dos seus átomos isolados envolve os seguintes estados: estado não ligante fundamental, estado não ligante promovido e estado covalente.

No estado não ligante fundamental, os átomos se encontram-se a distâncias infinitas nos respectivos estados fundamentais. Neste estado, as equações (3.177) e (3.178) podem ser condensadas,

$$\rho^{\infty}_0(X'_1|X_1) = \rho^{\infty}_{0,cl}(X'_1|X_1) = \sum_{W=1}^{N} \rho^{\infty}_0(W|W)\ , \quad (3.200)$$

e as populações electrónicas podem ser calculadas em termos da componente intrínseca da matriz densidade quase clássica,

$$\sum_{W=1}^{N} \int_{X'_1=X_1} \rho^{\infty}_0(W|W)\, dV_1 = n \quad (3.201)$$

A matriz densidade de par pode então ser escrita em termos de componentes de valência,

$$\pi_0^{\infty}(X_1'X_2'|X_1X_2) = 2\,\pi_{0,v}^{\infty}(X_1'X_2'|X_1X_2) = 2\sum_{W=1}^{N}\sum_{Z=1}^{N}\pi_0^{\infty}(WW|ZZ)$$

$$= 2\left\{\sum_{W=1}^{N}\pi_0^{\infty}(WW|WW) + \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\pi_0^{\infty}(WW|ZZ)\right\}, \qquad (3.202)$$

e as populações electrónicas de par podem ser calculadas em termos de componentes monoatómicas e diatómicas:

$$\iint_{\substack{X_1'=X_1\\X_2'=X_2}} \pi_{0,v}^{\infty}(X_1'X_2'|X_1X_2)\,dV_1\,dV_2 = \sum_{W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\pi_0^{\infty}(WW|WW)\,dV_1\,dV_2$$

$$+\sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\pi_0^{\infty}(WW|ZZ)\,dV_1\,dV_2 = \frac{n(n-1)}{2} \qquad (3.203)$$

A energia do estado não ligante fundamental pode ser escrita como a soma de componentes intra-atómicas.

$$E_0^{\infty} = E_{0,cl}^{\infty} = \int_{X_1'=X_1}\hat{h}_1\,\rho_{0,v}^{\infty}(X_1'|X_1)\,dV_1 + \iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\,\pi_{0,v}^{\infty}(X_1'X_2'|X_1X_2)\,dV_1\,dV_2$$

$$= \sum_{W=1}^{N}\int_{X_1'=X_1}\left[-\frac{1}{2}\nabla_1^2 - \frac{Z_W}{\varepsilon\, r_{1W}}\right]\rho_0^{\infty}(W|W)\,dV_1 + \sum_{W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\pi_0^{\infty}(WW|WW)\,dV_1\,dV_2 \qquad (3.204)$$

No estado não ligante promovido, os átomos se encontram-se a distâncias infinitas em estados excitados especialmente apropriados à formação posterior das ligações covalentes. Neste estado, podem definir-se relações do tipo (3.200) e (3.201) para a matriz densidade:

$$\rho_p^{\infty}(X_1'|X_1) = \rho_{p,cl}^{\infty}(X_1'|X_1) = \sum_{W=1}^{N} \rho_p^{\infty}(W|W) \ , \tag{3.205}$$

$$\sum_{W=1}^{N} \int_{X_1'=X_1} \rho_p^{\infty}(W|W)\, dV_1 = n \tag{3.206}$$

Em coerência com a natureza deste estado, as componentes atómicas da matriz densidade quase clássica são definidas de forma a satisfazer um critério de minimização do erro quadrático médio relativamente às correspondentes componentes no estado covalente:

$$\sum_{W=1}^{N} \left\{ \rho_p^{\infty}(W|W) - \rho_k^{ov\ Mull.}(W) \right\}^2 = \min \tag{3.207}$$

Em condições ideais, a equação (3.207) pode ser escrita como:

$$\sum_{W=1}^{N} \left\{ \rho_p^{\infty}(W|W) - \rho_k^{ov\ Mull.}(W) \right\}^2 = 0 \tag{3.208}$$

A matriz densidade de par, populações electrónicas de par e energias são calculadas por expressões análogas às utilizadas para o estado não ligante fundamental:

$$\pi_p^{\infty}(X_1'X_2'|X_1X_2) = 2\,\pi_{p,v}^{\infty}(X_1'X_2'|X_1X_2) = 2\sum_{W=1}^{N}\sum_{Z=1}^{N} \pi_p^{\infty}(WW|ZZ)$$

$$= 2\left\{\sum_{W=1}^{N}\pi_p^{\infty}(WW|WW) + \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\pi_p^{\infty}(WW|ZZ)\right\}, \qquad (3.209)$$

$$\iint_{\substack{X_1'=X_1\\X_2'=X_2}} \pi_{p,v}^{\infty}(X_1'X_2'|X_1X_2)\,dV_1\,dV_2 = \sum_{W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\pi_p^{\infty}(WW|WW)\,dV_1\,dV_2$$

$$+\sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\pi_p^{\infty}(WW|ZZ)\,dV_1\,dV_2 = \frac{n(n-1)}{2} \qquad (3.210)$$

$$E_p^{\infty} = E_{p,cl}^{\infty} = \int_{X_1'=X_1}\hat{h}_1\,\rho_{p,v}^{\infty}(X_1'|X_1)\,dV_1 + \iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\,\pi_{p,v}^{\infty}(X_1'X_2'|X_1X_2)\,dV_1\,dV_2$$

$$= \sum_{W=1}^{N}\int_{X_1'=X_1}\left[-\frac{1}{2}\nabla_1^2 - \frac{Z_W}{\varepsilon\, r_{1W}}\right]\rho_p^{\infty}(W|W)\,dV_1 + \sum_{W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\pi_p^{\infty}(WW|WW)\,dV_1\,dV_2 \qquad (3.211)$$

Dentro da aproximação da super-molécula, a energia de estabilização é definida como,

$$\Delta E_p^{\infty} = E_p^{\infty} - E_0^{\infty}, \qquad (3.212)$$

e pode ser decomposta em termos de componentes atómicas:

$$\Delta E_p^{\infty} = \Delta E_{p,cl}^{\infty} = \sum_{W=1}^{N}\int_{X_1'=X_1}\left[-\frac{1}{2}\nabla_1^2 - \frac{Z_W}{\varepsilon\, r_{1W}}\right]\left\{\rho_p^{\infty}(W|W) - \rho_0^{\infty}(W|W)\right\}dV_1$$

$$+\sum_{W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\left\{\pi_p^{\infty}(WW|WW) - \pi_0^{\infty}(WW|WW)\right\}dV_1\,dV_2 \qquad (3.213)$$

Num estado covalente, o formalismo geral (equações 3.177 a 3.199) é inteiramente aplicável e a energia de estabilização é definida como:

$$\Delta E_k^{ov} = E_k^{ov} - E_0^{\infty} \tag{3.214}$$

Com base nas equações (3.196) a (3.199), (3.204) e (3.214), a energia de estabilização poderia ser decomposta nas seguintes componentes:

$$\Delta E_k^{ov} = \Delta E_{k,cl}^{ov} + E_{k,I}^{ov} + E_{k,II}^{ov} , \tag{3.215}$$

onde:

$$\begin{aligned}\Delta E_{k,cl}^{ov} = E_{k,cl}^{ov} - E_0^{\infty} = & \sum_{W=1}^{N}\sum_{Z=1}^{W-1}\frac{Z_W Z_Z}{\varepsilon r_{WZ}} + \sum_{W=1}^{N}\int_{X_1'=X_1}\hat{h}_1\left\{\rho_k^{ov\ Mull.}(W) - \rho_0^{\infty}(W|W)\right\} dV_1 \\ & + \sum_{W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\left\{\pi_k^{ov}(WW|WW) - \pi_0^{\infty}(WW|WW)\right\} dV_1\, dV_2 \\ & + \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\pi_k^{ov}(WW|ZZ)\, dV_1\, dV_2\end{aligned} \tag{3.216}$$

No entanto, no modelo de Ruedenberg, considera-se que um estado covalente se forma a partir de um estado não ligante promovido o que implica uma reformulação do formalismo usado. No caso limite da equação (3.208), podem estabelecer-se as seguintes relações entre os dois estados:

$$\rho_k^{ov\ Mull.}(W) = \rho_p^{\infty}(W|W) \tag{3.217}$$

$$\rho_{k,cl}^{ov}(X_1'|X_1) = \rho_p^{\infty}(X_1'|X_1) \tag{3.218}$$

$$\rho_k^{ov}(X_1'|X_1) = \rho_p^{\infty}(X_1'|X_1) + \rho_{k,I}^{ov}(X_1'|X_1) \tag{3.219}$$

$$\pi_{k,v}^{ov}(X_1'X_2'|X_1X_2) = \pi_p^{\infty}(X_1'X_2'|X_1X_2) + \pi_{k,sp}^{ov}(X_1'X_2'|X_1X_2) \tag{3.220}$$

Na equação (3.220), $\pi_{k,sp}^{ov}(X_1'X_2'|X_1X_2)$ é a chamada matriz de interpenetração e partilha electrónicas, cujas componentes atómicas são definidas como:

$$\pi_{k,sp}^{ov}(WW|WW) = \pi_{k,v}^{ov}(WW|WW) - \pi_p^{\infty}(WW|WW) \tag{3.221}$$

$$\pi_{k,sp}^{ov}(WW|ZZ) = \pi_{k,v}^{ov}(WW|ZZ) - \pi_p^{\infty}(WW|ZZ)\ ;\ W \neq Z \tag{3.222}$$

Neste contexto, a energia pode ser decomposta nas seguintes componentes,

$$E_k^{ov} = E_p^{\infty} + E_{k,clp}^{ov} + E_{k,sp}^{ov} + E_{k,I}^{ov} + E_{k,II}^{ov}\ , \tag{3.223}$$

onde $E_{k,clp}^{ov}$ e $E_{k,sp}^{ov}$ representam, respectivamente, as componentes quase clássica e de interpenetração e partilha, dadas por:

$$E_{k,clp}^{ov} = \sum_{W=1}^{N}\sum_{Z=1}^{W-1}\frac{Z_W Z_Z}{\varepsilon r_{WZ}} - \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\int_{X_1'=X_1}\frac{Z_Z}{\varepsilon r_{1Z}}\rho_p^{\infty}(W|W)\,dV_1$$

$$+ \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N}\iint_{\substack{X_1'=X_1\\X_2'=X_2}}\hat{h}_{12}\pi_p^{\infty}(WW|ZZ)\,dV_1\,dV_2 \tag{3.224}$$

e

$$E_{k,sp}^{ov} = \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12}\, \pi_{k,sp}^{ov}(X_1'X_2'|X_1X_2)\, dV_1\, dV_2 = \sum_{W=1}^{N} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12}\, \pi_{k,sp}^{ov}(WW|WW)\, dV_1\, dV_2$$

$$+ \sum_{W=1}^{N}\sum_{Z\neq W=1}^{N} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12}\pi_{k,sp}^{ov}(WW|ZZ)\, dV_1\, dV_2 \qquad (3.225)$$

A decomposição da energia de estabilização pode então ser reformulada:

$$\Delta E_k^{ov} = \Delta E_p^{\infty} + E_{k,clp}^{ov} + E_{k,sp}^{ov} + E_{k,I}^{ov} + E_{k,II}^{ov} \qquad (3.226)$$

No modelo anteriormente apresentado, a formação do estado covalente resulta de um mecanismo com dois passos:

No primeiro passo, os átomos transitam do estado não ligante fundamental para o estado não ligante promovido. Isto significa que os átomos permanecem a distâncias infinitas mas adquirem configurações electrónicas excitadas especialmente apropriadas à formação do estado covalente. No segundo passo, os átomos aproximam-se até às distâncias internucleares de equilíbrio e, preservando as respectivas populações electrónicas (relações 3.217 e 3.218), estabelecem as ligações covalentes entre si.

No entanto, um modelo deste tipo só é aplicável a agregados moleculares constituídos por átomos neutros o que constitui uma situação pouco comum. O caso mais vulgar é descrito pela equação (3.207), o que implica que as equações (3.217) a (3.226) têm de ser reformuladas de forma a incorporarem os efeitos associados à transferência de carga.

Em relação à matrizes densidade podem definir-se as seguintes relações modificadas,

$$\rho_{k,cl}^{ov}(X_1'|X_1) = \rho_p^{\infty}(X_1'|X_1) + \rho_{k,T}^{ov}(X_1'|X_1)\ , \qquad (3.227)$$

$$\rho_k^{ov\ Mull.}(W) = \rho_p^{\infty}(W|W) + \rho_{k,T}^{ov\ Mull.}(W)\ , \qquad (3.228)$$

$$\rho_{k,I}^{ov}(X_1'|X_1) = \rho_{k,IP}^{ov}(X_1'|X_1) = \rho_{p,IT}^{\infty}(X_1'|X_1)\ , \qquad (3.229)$$

e

$$\rho_k^{ov}(W/Z) = \rho_{k,IP}^{ov}(W/Z) + \rho_{p,IT}^{\infty}(W/Z) \tag{3.230}$$

onde $\rho_{k,IP}^{ov}(X_1'|X_1)$ e $\rho_{k,IP}^{ov}(W/Z)$ representam componentes calculadas para um estado covalente hipotético em que seja válida a condição (3.208) e $\rho_{k,T}^{ov}(X_1'|X_1)$, $\rho_{k,T}^{ov\,Mull.}(W)$, $\rho_{p,IT}^{\infty}(X_1'|X_1)$ e $\rho_{p,IT}^{\infty}(W/Z)$ representam correcções devidas à transferência de carga.

Em relação às matrizes densidade de par podem derivar-se as seguintes relações:

$$\pi_{k,v}^{ov}(X_1'X_2'|X_1X_2) = \pi_p^{\infty}(X_1'X_2'|X_1X_2) + \pi_{k,sp}^{ov}(X_1'X_2'|X_1X_2) + \pi_{k,T}^{ov}(X_1'X_2'|X_1X_2), \tag{3.231}$$

$$\pi_{k,v}^{ov}(WW|WW) = \pi_p^{\infty}(WW|WW) + \pi_{k,sp}^{ov}(WW|WW) + \pi_{k,T}^{ov}(WW|WW)\ , \tag{3.232}$$

$$\pi_{k,v}^{ov}(WW|ZZ) = \pi_p^{\infty}(WW|ZZ) + \pi_{k,sp}^{ov}(WW|ZZ) + \pi_{k,T}^{ov}(WW|ZZ)\ , \tag{3.233}$$

$$\pi_{k,I}^{ov}(X_1'X_2'|X_1X_2) = \pi_{k,Ip}^{ov}(X_1'X_2'|X_1X_2) + \pi_{k,IT}^{ov}(X_1'X_2'|X_1X_2)\ , \tag{3.234}$$

e

$$\pi_{k,I}^{ov}(WW|ZT) = \pi_{k,Ip}^{ov}(WW|ZT) + \pi_{k,IT}^{ov}(WW|ZT)\ , \tag{3.235}$$

onde $\pi_{k,sp}^{ov}(X_1'X_2'|X_1X_2)$, $\pi_{k,sp}^{ov}(WW|WW)$, $\pi_{k,sp}^{ov}(WW|ZZ)$, $\pi_{k,Ip}^{ov}(X_1'X_2'|X_1X_2)$ e $\pi_{k,Ip}^{ov}(WW|ZT)$ representam componentes calculadas para um estado covalente hipotético em que seja válida a condição (3.208) e $\pi_{k,T}^{ov}(X_1'X_2'|X_1X_2)$, $\pi_{k,T}^{ov}(WW|WW)$, $\pi_{k,T}^{ov}(WW|ZZ)$, $\pi_{k,IT}^{ov}(X_1'X_2'|X_1X_2)$ e $\pi_{k,IT}^{ov}(WW|ZT)$ representam correcções devidas à transferência de carga.

Neste contexto, a energia pode ser decomposta nas seguintes componentes:

$$E_k^{ov} = E_p^{\infty} + E_{k,clp}^{ov} + E_{k,sp}^{ov} + E_{k,T}^{ov} + E_{k,Ip}^{ov} + E_{k,IT}^{ov} + E_{k,II}^{ov} \tag{3.236}$$

Na equação (3.236), $E_{k,clp}^{ov}$, $E_{k,sp}^{ov}$, e $E_{k,Ip}^{ov}$ representam componentes calculadas para um estado covalente hipotético em que seja válida a condição (3.208) e $E_{k,T}^{ov}$ e $E_{k,IT}^{ov}$ representam correcções devidas à transferência de carga, calculadas como:

$$\begin{aligned} E_{k,T}^{ov} &= \int_{X_1'=X_1} \hat{h}_1 \, \rho_{k,T}^{ov}(X_1'|X_1) \, dV_1 + \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12} \, \pi_{k,T}^{ov}(X_1'X_2'|X_1X_2) \, dV_1 \, dV_2 \\ &= \sum_{W=1}^{N} \int_{X_1'=X_1} \hat{h}_1 \rho_{k,T}^{\Gamma\, Mull.}(W) \, dV_1 + \sum_{W=1}^{N} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12} \pi_{k,T}^{\Gamma}(WW|WW) \, dV_1 \, dV_2 \\ &+ \sum_{W=1}^{N} \sum_{Z\neq W=1}^{N} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12} \pi_{k,T}^{\Gamma}(WW|ZZ) \, dV_1 \, dV_2 \end{aligned} \tag{3.237}$$

e

$$\begin{aligned} E_{k,IT}^{ov} &= \int_{X_1'=X_1} \hat{h}_1 \, \rho_{k,IT}^{ov}(X_1'|X_1) \, dV_1 + \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12} \, \pi_{k,IT}^{ov}(X_1'X_2'|X_1X_2) \, dV_1 \, dV_2 \\ &= \sum_{W=1}^{N} \sum_{Z\neq W=1}^{N} \int_{X_1'=X_1} \hat{h}_1 \rho_{k,IT}^{ov}(W/Z) \, dV_1 + 2 \sum_{W=1}^{N} \sum_{Z=1}^{N} \sum_{T\neq Z=1}^{N} \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \hat{h}_{12} \pi_{k,IT}^{ov}(WW|ZT) \, dV_1 \, dV_2 \end{aligned} \tag{3.238}$$

A energia de estabilização pode então ser decomposta de acordo com a equação:

$$\Delta E_k^{ov} = \Delta E_p^{\infty} + E_{k,clp}^{ov} + E_{k,sp}^{ov} + E_{k,T}^{ov} + E_{k,IP}^{ov} + E_{k,IT}^{ov} + E_{k,II}^{ov} \quad (3.239)$$

As principais vantagens específicas do método de Ruedenberg são:

(i) A partição é definida para métodos *ab initio* monoconfiguracionais e multiconfiguracionais.

(ii) O formalismo é expresso explicitamente em termos da decomposição espacial das matrizes densidade e densidade de par. Isto permite uma interpretação física da interacção entre os átomos tendo em conta a natureza das ligações químicas que se estabelecem entre eles.

(iii) São definidos não só componentes diatómicas mas também componentes multicêntricas (triatómicas e tetratómicas).

(iv) A partição é aplicada à energia de estabilização.

Os principais inconvenientes deste método são:

(1) A partição é limitada a fragmentos monoatómicos.

(2) Os fragmentos são considerados entidades rígidas.

(3) O conceito de sobreposição electrónica, habitualmente utilizado na linguagem química, é substituído pelo conceito de interferência sem que isto traga vantagens evidentes para o modelo.

(4) As populações electrónicas são calculadas utilizando a aproximação de Mulliken o que introduz desnecessariamente mais um factor de arbitrariedade no formalismo.

3.5.4- Método de Clementi

O método de Clementi [100] foi introduzido para decompor a energia molecular em componentes atómicas dentro do formalismo RHF. Neste contexto, a energia molecular (equação 3.78) pode ser escrita como a soma de componentes monoatómicas e inter-atómicas (bicêntricas, tricêntricas e tetracêntricas),

$$E_k^{\Gamma} = \sum_{X=1}^{N} E_k^{\Gamma}(X) + \sum_{X=1}^{N}\sum_{Y=1}^{X-1} E_k^{\Gamma}(XY) + \sum_{X=1}^{N}\sum_{Y=1}^{X-1}\sum_{Z=1}^{Y-1} E_k^{\Gamma}(XYZ) + \sum_{X=1}^{N}\sum_{Y=1}^{X-1}\sum_{Z=1}^{Y-1}\sum_{W=1}^{Z-1} E_k^{\Gamma}(XYZW) \quad , \tag{3.240}$$

onde

$$E_k^{\Gamma}(X) = 2\sum_{i=k_1}^{k_{n/2}} h_{ii}(X) + \sum_{i=k_1}^{k_{n/2}}\sum_{j=k_1}^{k_{n/2}} G_{ij}(X) \tag{3.241}$$

$$E_k^{\Gamma}(XY) = 2\sum_{i=k_1}^{k_{n/2}} h_{ii}(XY) + \sum_{i=k_1}^{k_{n/2}}\sum_{j=k_1}^{k_{n/2}} G_{ij}(XY) \tag{3.242}$$

$$E_k^{\Gamma}(XYZ) = 2\sum_{i=k_1}^{k_{n/2}} h_{ii}(XYZ) + \sum_{i=k_1}^{k_{n/2}}\sum_{j=k_1}^{k_{n/2}} G_{ij}(XYZ) \tag{3.243}$$

$$E_k^{\Gamma}(XYZW) = \sum_{i=k_1}^{k_{n/2}}\sum_{j=k_1}^{k_{n/2}} G_{ij}(XYZW) \tag{3.244}$$

Nas equações (3.241) a (3.244), $h_{ii}(X)$, $h_{ii}(XY)$, $h_{ii}(XYZ)$ e $h_{ii}(XYZW)$ representam, respectivamente, as componentes monoatómicas, diatómicas, triatómicas e tetratómicas da energia monoelectrónica, enquanto $G_{ij}(X)$, $G_{ij}(XY)$, $G_{ij}(XYZ)$ e $G_{ij}(XYZW)$ têm o mesmo significado relativamente à energia bielectrónica.

Na formulação geral do método RHF, as funções orbitais (η_i) são funções absolutamente genéricas e, nestas circunstâncias, é difícil encontrar um critério rigoroso para a partição espacial das energias. A aproximação de Roothaan (equação 3.64) fornece uma boa metodologia para esta partição que é então expressa em termos das funções de base atómicas.

O formalismo de Clementi, dada a sua natureza geral, pode em princípio ser acoplado com outras metodologias de partição da densidade electrónica como as sugeridas por Bader *et*

al [108-110] (baseada na definição de fragmentos onde o teorema virial é localmente satisfeito) e Daudel *et al* [111] (baseada na definição de orbitais moleculares localizadas).

O método de Clementi apresenta as seguintes vantagens específicas:

(i) A partição espacial das energias monoelectrónicas é feita de uma forma mais rigorosa do que nos métodos de Bofill e Kollmar (ver discussão posterior na secção 3.5.7).

(ii) São definidos não só componentes diatómicas mas também componentes multicêntricas (triatómicas e tetratómicas).

(iii) A decomposição da energia pode ser feita, não só com base na aproximação de Roothaan, mas também com base noutros métodos de partição da densidade electrónica.

Os principais inconvenientes deste método são:

(1) A partição é limitada a fragmentos monoatómicos.

(2) Os fragmentos são considerados entidades rígidas.

(3) A decomposição só foi rigorosamente definida no formalismo RHF.

(4) O formalismo não foi explicitamente expresso em termos da decomposição espacial das matrizes densidade e densidade de par.

(5) A partição não é aplicada à energia de estabilização.

3.5.5- Método de Kollmar

O método de Kollmar [101] baseia-se na partição espacial da matriz de Fock e dos integrais monoelectrónicos, dentro de um formalismo RHF. De acordo com este método, a energia molecular (equação 3.81) é decomposta em componentes monoatómicas e diatómicas efectivas:

$$E_k^{\Gamma} = \sum_{X=1}^{N} E_k^{\Gamma\,ef.}(X) + \sum_{X=1}^{N}\sum_{Y=1}^{X-1} E_k^{\Gamma\,ef.}(XY) \qquad (3.245)$$

Na equação (3.245), as componentes monoatómicas e diatómicas são dadas respectivamente por,

$$E_k^{\Gamma\,ef.}(X) = \frac{1}{2}\sum_{p=n_X}^{m_X} P_{pq}^{\Gamma,k}\,(h_{pq} + F_{pq}^{\Gamma,k}) + \frac{1}{2}\,Z_X \sum_{Y\neq X=1}^{N} \frac{Z_Y}{\varepsilon\, r_{XY}}\,, \tag{3.246}$$

e

$$E_k^{\Gamma\,ef.}(XY) = \sum_{p=n_X}^{m_X}\sum_{q=n_Y}^{m_Y} P_{pq}^{\Gamma,k}\,(h_{pq} + F_{pq}^{\Gamma,k})\,, \tag{3.247}$$

onde n_X e m_X representam, respectivamente, a primeira e a última funções de base atómicas do átomo X.

O método de Kollmar apresenta analogias com diversos métodos de decomposição propostos dentro do formalismo semi-empírico [112-113] e apresenta como principal vantagem a sua simplicidade. De facto, a energia molecular é decomposta unicamente em componentes monoatómicas e diatómicas efectivas, o que poderia ser utilizado para construir potenciais a dois corpos muito utilizados em simulações moleculares [114]. No entanto, os fundamentos físicos deste método não são muito consistentes o que limita a sua aplicabilidade. Neste contexto, o método de Kollmar apresenta as seguintes principais desvantagens:

(1) O critério utilizado para a decomposição espacial dos termos monoelectrónicos (h_{pr}) é pouco rigoroso (ver discussão posterior na secção 3.5.7).

(2) A interpretação física associada à partição espacial da matriz de Fock não tem uma justificação fundamentada.

(3) A energia de repulsão nuclear é incluída nas componentes monoatómicas.

(4) A partição é limitada a fragmentos monoatómicos.

(5) Os fragmentos são considerados entidades rígidas.

(6) O método de decomposição só é válido dentro do formalismo RHF.

(7) A partição não é aplicada à energia de estabilização.

3.5.6- Método de Bofill

O método de Bofill [102] permite que a energia molecular seja decomposta em componentes atómicas com base na partição espacial da matriz densidade de par. De facto, conjugando as equações (3.37), (3.38), (3.49), (3.51) e (3.52) é possível obter uma expressão simplificada para a energia molecular expressa em termos da matriz densidade de par,

$$E_k^{\Gamma} = E_k^{\Gamma(0)} + \sum_{p=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{m}\sum_{s=1}^{m} D_{pqrs}^{\Gamma,k} \{pq|rs\} , \qquad (3.248)$$

onde $\{pq|rs\}$ são elementos da chamada matriz de Bopp [115], definidos como:

$$\{pq|rs\} = \frac{1}{2}\left[\frac{1}{n-1}(h_{pq} S_{rs} + h_{rs} S_{pq}) + (pq|rs)\right] \qquad (3.249)$$

As relações de conservação (3.48) e (3.49) são, em seguida, expressas em termos de componentes atómicas,

$$\sum_{X=1}^{N}\sum_{Y=1}^{N}\sum_{Z=1}^{N}\sum_{W=1}^{N} D^{\Gamma,k}(XY|ZW) = \frac{n(n-1)}{2} , \qquad (3.250)$$

e

$$2\sum_{Z=1}^{N}\sum_{W=1}^{N} D^{\Gamma,k}(XY|ZW) = (n-1) P^{\Gamma,k}(X|Y) , \qquad (3.251)$$

onde

$$D^{\Gamma,k}(XY|ZW) = \frac{1}{2}\sum_{p=n_X}^{m_X}\sum_{q=n_Y}^{m_Y}\sum_{r=n_Z}^{m_Z}\sum_{s=n_W}^{m_W} D^{\Gamma,k}_{pqrs}\, S_{pq}\, S_{rs} \tag{3.252}$$

$$P^{\Gamma,k}(X|Y) = \sum_{Z=1}^{N}\sum_{W=1}^{N} P^{\Gamma,k}_{pq}\, S_{pq} \tag{3.253}$$

A equação (3.250) pode ainda ser rearranjada como:

$$\sum_{X=1}^{N} D^{\Gamma,k}(XX|XX) + \sum_{X=1}^{N}\sum_{Y\neq X=1}^{N} D^{\Gamma,k}(XX|YY)$$

$$+ 2\sum_{X=1}^{N}\sum_{Y=1}^{N}\sum_{Z\neq Y=1}^{N} D^{\Gamma,k}(XX|YZ) + \sum_{X=1}^{N}\sum_{Y\neq X=1}^{N}\sum_{Z=1}^{N}\sum_{W\neq Z=1}^{N} D^{\Gamma,k}(XX|XX) = \frac{n(n-1)}{2} \tag{3.254}$$

Em analogia, a energia molecular também pode ser decomposta em componentes atómicas,

$$E^{\Gamma}_{k} = \sum_{X=1}^{N}\sum_{Y=1}^{N}\sum_{Z=1}^{N}\sum_{W=1}^{N} E^{\Gamma\,(1+2)}_{k}(XY|ZW) + \sum_{X=1}^{N}\sum_{Y=1}^{N} E^{\Gamma\,(0)}_{k}(X|Y)\ , \tag{3.255}$$

onde $E^{\Gamma\,(1+2)}_{k}(XY|ZW)$ é a energia electrónica associada ao quarteto ordenado (X,Y,Z,W),

$$E^{\Gamma\,(1+2)}_{k}(XY|ZW) = \sum_{p=n_X}^{m_X}\sum_{p=n_Y}^{m_Y}\sum_{r=n_Z}^{m_Z}\sum_{s=n_W}^{m_W} D^{\Gamma,k}_{pqrs}\ \{pq|rs\}\ , \tag{3.256}$$

e $E^{\Gamma\,(0)}_{k}(X|Y)$ é a energia de repulsão nuclear associada ao par ordenado (X,Y):

$$E_k^{\Gamma\,(0)}(X|Y) = \frac{Z_X\, Z_Y}{2\varepsilon\, r_{XY}} \tag{3.257}$$

A equação (3.255) pode ser reorganizada de uma forma semelhante à efectuada para a partição da matriz densidade de par na equação (3.254),

$$E_k^{\Gamma} = \sum_{X=1}^{N} E_k^{\Gamma}(XX|XX) + \sum_{X=1}^{N}\sum_{Y\neq X=1}^{N} E_k^{\Gamma}(XX|YY)$$
$$+ \sum_{X=1}^{N}\sum_{Y=1}^{N}\sum_{Z\neq Y=1}^{N} E_k^{\Gamma}(XX|YZ) + \sum_{X=1}^{N}\sum_{Y\neq X=1}^{N}\sum_{Z=1}^{N}\sum_{W\neq Z=1}^{N} E_k^{\Gamma}(XY|ZW) \tag{3.258}$$

onde

$$E_k^{\Gamma}(XY|ZW) = E_k^{\Gamma\,(1+2)}(XY|ZW) \tag{3.259}$$

$$E_k^{\Gamma}(XX|YZ) = E_k^{\Gamma\,(1+2)}(XX|YZ) \tag{3.260}$$

$$E_k^{\Gamma}(XX|XX) = E_k^{\Gamma\,(1+2)}(XX|XX) \tag{3.261}$$

e

$$E_k^{\Gamma}(XX|YY) = E_k^{\Gamma\,(1)}(XX|YY) + E_k^{\Gamma\,(R)}(XX|YY) \tag{3.262}$$

Na equação (3.262), as componentes monoelectrónica e de repulsão são dadas respectivamente por:

$$E_k^{\Gamma\,(1)}(XX|YY) = \frac{1}{2}\left[\frac{1}{n-1}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X}\sum_{r=n_Y}^{m_Y}\sum_{s=n_Y}^{m_Y}(h_{pq}S_{rs} + h_{rs}S_{pq})\, D_{pqrs}^{\Gamma,k}\right] \tag{3.263}$$

$$E_k^{\Gamma\,(R)}(XX|YY) = E_k^{\Gamma\,(2)}(XX|YY) + E_k^{\Gamma\,(0)}(X|Y) \tag{3.264}$$

Na equação (3.264), a componente bielectrónica repulsiva é definida como:

$$E_k^{\Gamma\,(2)}(XX|YY) = \frac{1}{2}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X}\sum_{r=n_Y}^{m_Y}\sum_{s=n_Y}^{m_Y} D_{pqrs}^{\Gamma,k}\,(pq|rs) \tag{3.265}$$

Num estado não ligante ($\Gamma = \infty$), as relações de conservação (3.254) e (3.251) podem ser simplificadas:

$$\sum_{X=1}^{N} D^{\infty,k}(XX|XX) + \sum_{X=1}^{N}\sum_{Y\neq X=1}^{N} D^{\infty,k}(XX|YY) = \frac{n(n-1)}{2} \tag{3.266}$$

$$2\sum_{Y=1}^{N} D^{\infty,k}(XX|YY) = (n-1)\,P^{\infty,k}\,(X|X) \tag{3.267}$$

Num estado deste tipo, a energia molecular (equação 3.258) só tem duas componentes não nulas:

$$E_k^{\infty} = \sum_{X=1}^{N} E_k^{\infty}(XX|XX) + \sum_{X=1}^{N}\sum_{Y\neq X=1}^{N} E_k^{\infty\,(1)}(XX|YY) \tag{3.268}$$

O método de Bofill apresenta as seguintes vantagens específicas:

(i) A partição é definida para métodos *ab initio* monoconfiguracionais e multiconfiguracionais.

(ii) O formalismo é expresso em termos da partição espacial da matriz densidade de par. Isto permite que a estrutura molecular e componentes da energia sejam discutidos tendo em conta apenas as populações electrónicas de par.

(iii) São definidos não só componentes diatómicas mas também componentes multicêntricas (triatómicas e tetratómicas).

Os principais inconvenientes deste método são:

(1) A partição é limitada a fragmentos monoatómicos.

(2) Os fragmentos são considerados entidades rígidas.

(3) O critério utilizado para a decomposição espacial dos termos monoelectrónicos (h_{pr}) é pouco rigoroso (ver discussão posterior na secção 3.5.7).

(4) O formalismo expresso em termos da matriz densidade de par leva a algumas ambiguidades na interpretação da partição espacial das componentes da energia. Por exemplo, $E_k^{\infty(1)}(XX|YY)$ na equação (3.268) é interpretada como uma componente de interacção da energia molecular num estado não ligante [102]. No entanto, desenvolvendo explicitamente a energia monoelectrónica total na equação (3.268) e recorrendo à relação de conservação (3.267),

$$
\begin{aligned}
E_k^{\infty(1)} &= \sum_{X=1}^{N}\sum_{Y=1}^{N} E_k^{\infty(1)}(XX|YY) \\
&= \frac{1}{2}\sum_{X=1}^{N}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X}\sum_{Y=1}^{N}\sum_{r=n_Y}^{m_Y}\sum_{s=n_Y}^{m_Y}\frac{1}{n-1}\left(h_{pq}S_{rs} + h_{rs}S_{pq}\right) D_{pqrs}^{\infty,k} \\
&= \frac{1}{2}\left\{\sum_{X=1}^{N}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X} h_{pq} \times \frac{1}{n-1}\sum_{Y=1}^{N}\sum_{r=n_Y}^{m_Y}\sum_{s=n_Y}^{m_Y} D_{pqrs}^{\infty,k} S_{rs}\right. \\
&\quad \left. + \sum_{Y=1}^{N}\sum_{r=n_Y}^{m_Y}\sum_{s=n_Y}^{m_Y} h_{rs} \times \frac{1}{n-1}\sum_{X=1}^{N}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X} D_{pqrs}^{\infty,k} S_{pq}\right\} \\
&= \frac{1}{2}\left\{\sum_{X=1}^{N}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X} P_{pq}^{\infty,k}\, h_{pq} + \sum_{Y=1}^{N}\sum_{r=n_Y}^{m_Y}\sum_{s=n_Y}^{m_Y} P_{rs}^{\infty,k}\, h_{rs}\right\} \\
&= \sum_{X=1}^{N}\sum_{p=n_X}^{m_X}\sum_{q=n_X}^{m_X} P_{pq}^{\infty,k}\, h_{pq} \ , \qquad (3.269)
\end{aligned}
$$

pode concluir-se esta componente é monocêntrica e não tem nenhuma componente de interacção.

(5) Na partição espacial das populações electrónicas de par (equação 3.254) e da energia (equação 3.255) são misturados termos associados a diferentes números de centros. Por exemplo, na terceira componente destas equações são misturados termos tricêntricos $\{D^{\Gamma,k}(XX|YZ)$ e $E_k^{\Gamma}(XX|YZ)$ com $X \neq Y \neq Z\}$ e termos bicêntricos $\{D^{\Gamma,k}(XX|XY)$ e $E_k^{\Gamma}(XX|XY)$ com $X \neq Y\}$. Na quarta componente destas equações são misturados termos tetracêntricos $\{ D^{\Gamma,k}(XY|ZW)$ e $E_k^{\Gamma}(XY|ZW)$ com $X \neq Y \neq Z \neq W \}$, termos tricêntricos $\{ D^{\Gamma,k}(XY|XZ)$ e $E_k^{\Gamma}(XY|XZ)$ com $X \neq Y \neq Z \}$ e termos bicêntricos $\{ D^{\Gamma,k}(XY|XY)$ e $E_k^{\Gamma}(XY|XY)$ com $X \neq Y\}$.

(6) A partição não é aplicada à energia de estabilização.

3.5.7- Método de Melo e Ramos

Neste trabalho é proposto um novo método de decomposição das interacções moleculares que se baseia no formalismo dos integrais indistinguíveis e na partição espacial do agregado molecular em t fragmentos definidos de uma forma completamente flexível.

As populações electrónicas (equação 3.47) e energias monoelectrónicas (equação 3.51) são calculadas somando as contribuições associadas a todos os pares ordenados (p,q). No entanto, alguns deles (pares indistinguíveis) dão a mesma contribuição para as populações electrónicas e energias monoelectrónicas [1]. Consequentemente, é possível seleccionar os pares cujas contribuições diferem por simetria e exprimir as equações (3.47) e (3.51) em função deles:

$$\sum_{p=1}^{m} P_{pp}^{\Gamma,k} + 2 \sum_{p=1}^{m} \sum_{q=1}^{p-1} P_{pq}^{\Gamma,k} = n \tag{3.270}$$

$$E^{(1)} = \sum_{p=1}^{m} P_{pp}^{\Gamma,k} h_{pp} + 2 \sum_{p=1}^{m} \sum_{q=1}^{p-1} P_{pq}^{\Gamma,k} h_{pq} \tag{3.271}$$

Neste trabalho, esta metodologia é alargada ao cálculo das populações electrónicas de par (equação 3.48) e energias bielectrónicas (equação 3.52). Estas quantidades são calculadas habitualmente através da soma das contribuições associadas a todos os quartetos ordenados (p,q,r,s). No entanto, é também possível definir o conceito de quartetos indistinguíveis [116] simplificando o formalismo associado a estes cálculos. Por exemplo, o quarteto (p,q,r,s com $p \neq q \neq r \neq s$) está associado a um conjunto de oito quartetos indistinguíveis,

$$(p,q,r,s)\ ;\ (p,q,s,r)\ ;\ (q,p,r,s)\ ;\ (q,p,s,r)\ ;\ (r,s,p,q)\ ;\ (r,s,q,p)\ ;\ (s,r,p,q)\ ;\ (s,r,q,p)\ , \quad (3.272)$$

cujas contribuições para as populações electrónicas de par e energias bielectrónicas são iguais.

De acordo com o presente formalismo, as equações (3.48) e (3.52) são expressas em função dos quartetos distinguíveis:

$$\frac{1}{2}\sum_{p=1}^{m}\sum_{q=1}^{p}\sum_{r=1}^{p}\sum_{s=1}^{lim} f_{pqrs}^{\Gamma,k}\, S_{pq} S_{rs} = \frac{n\,(n-1)}{2} \quad (3.273)$$

e

$$E_k^{\Gamma\,(2)} = \frac{1}{2}\sum_{p=1}^{m}\sum_{q=1}^{p}\sum_{r=1}^{p}\sum_{s=1}^{lim} f_{pqrs}^{\Gamma,k}\,(pq|rs) \quad (3.274)$$

Nas duas últimas equações, lim e $f_{pqrs}^{\Gamma,k}$ são definidos respectivamente como:

$$\text{lim} = q \text{ if } p = r \text{ ; lim} = r \text{ if } p \neq r \text{ ,} \tag{3.275}$$

$$f^{\Gamma,k}_{pqrs} = n_{pqrs} D^{\Gamma,k}_{pqrs} \text{ ,} \tag{3.276}$$

onde n_{pqrs} é o número de integrais indistinguíveis associados ao conjunto de índices (p,q,r,s). Os coeficientes n_{pqrs} para os diversos tipos de quartetos são apresentados na tabela 3.1.

Tabela 3.1. Coeficientes n_{pqrs} para diferentes tipos de quartetos.

Tipos de quartetos (p,q,r,s)	Exemplos	n_{pqrs}
$p = q \wedge r \neq s$; $pqrs \neq 0$	(p,p,r,s) $p \neq r \neq s$	4
	(p,p,p,s) $p \neq s$	
$p = q \wedge r = s$; $pqrs \neq 0$	(p,p,r,r) $p \neq r$	2
$p = q \wedge r = s$; $pqrs = 0$	(p,p,p,p)	1
$p \neq q \wedge r \neq s$; $pqrs \neq 0$	(p,q,r,s) $p \neq q \neq r \neq s$	8
$p \neq q \wedge r \neq s$; $pqrs = 0$	(p,q,p,q) $p \neq q$	4

$pqrs = 0 \Rightarrow (p=r \wedge q=s) \vee (p=s \wedge q=r)$
$pqrs \neq 0 \Rightarrow (p \neq r \vee q \neq s) \wedge (p \neq s \vee q \neq r)$

Em seguida, os fragmentos são definidos de uma forma apropriada ao estudo das interacções existentes no agregado molecular. Neste contexto, os fragmentos são entidades absolutamente gerais que podem ser átomos, grupos de átomos, iões, monómeros, etc.

O método de decomposição, válido dentro do formalismo *ab initio* multiconfiguracional ou monoconfiguracional, é então desenvolvido seguindo a seguinte metodologia:

a) As populações electrónicas e energias monoelectrónicas são expressas em termos de pares distinguíveis (equações 3.270 e 3.271).

b) As populações electrónicas de par e energias bielectrónicas são expressas em termos de pares distinguíveis (equações 3.273 e 3.274).

O procedimento usado em a) e b) reduz significativamente o número de integrais a calcular simplificando a implementação computacional do algoritmo de decomposição.

c) As populações electrónicas e electrónicas de par são decompostas em componentes associadas aos fragmentos.

d) A decomposição da energia monoelectrónica é obtida substituindo a partição das populações electrónicas na equação (3.271).

e) A decomposição da energia monoelectrónica é obtida substituindo a partição das populações electrónicas de par na equação (3.274).

f) A metodologia de decomposição é alargada, dentro da aproximação da super-molécula, à energia de estabilização.

g) Quando os fragmentos que constituem o agregado molecular compartilham electrões, existe sempre uma certa arbitrariedade associada a este tipo de métodos de decomposição [99]. Isto resulta da falta de um critério rigoroso na partição espacial das populações electrónicas de sobreposição. De acordo com a aproximação de Mulliken, utilizada no método de Ruedenberg, estas populações devem ser equalitariamente divididas pelos fragmentos envolvidos, o que não tem uma base física devidamente fundamentada. Neste trabalho, a aproximação de Mulliken não é utilizada e as populações electrónicas de sobreposição são encaradas como tendo uma natureza intrinsecamente inter-fragmento.

h) Na partição das energias monoelectrónicas, um critério frequentemente utilizado [99,101-102] consiste em associar os termos ($P_{pq}^{\Gamma,k}\ h_{pq}$) com as mesmas regiões espaciais dos elementos da matriz densidade ($P_{pq}^{\Gamma,k}$). No entanto, uma análise mais rigorosa do problema permite concluir que os integrais h_{pq} contêm uma componente de atracção entre todos os núcleos e a densidade de sobreposição $P_{pq}^{\Gamma,k}\ \chi_p^*(X_1')\ \chi_q(X_1)$ [103]. Consequentemente, as energias monoelectrónicas podem ser decompostas como,

$$P_{pq}^{\Gamma,k}\ h_{pq} = P_{pq}^{\Gamma,k}\ T_{pq} + \sum_{A=1}^{t} P_{pq}^{\Gamma,k}\ V_{pq}^{A}\ , \qquad (3.277)$$

onde,

$$T_{pq} = -\frac{1}{2} \left\langle \chi_p \middle| \nabla_1^2 \middle| \chi_q \right\rangle \tag{3.278}$$

e

$$V_{pq}^{A} = - \left\langle \chi_p \middle| \sum_{X=N_A}^{M_A} \frac{Z_X}{\varepsilon\, r_{1X}} \middle| \chi_q \right\rangle \tag{3.279}$$

Na equação (3.277), $P_{pq}^{\Gamma,k} V_{pq}^{A}$ representa a energia de atracção entre os núcleos do fragmento A e a densidade $P_{pq}^{\Gamma,k}\, \chi_p^*(X_1')\, \chi_q(X_1)$ e $P_{pq}^{\Gamma,k}\, T_{pq}$ representa a energia cinética da mesma densidade de sobreposição. Na equação (3.279), N_A e M_A representam o primeiro e o último núcleos do fragmento A, respectivamente.

As componentes cinética e potencial das energias monoelectrónicas podem ser associadas a diferentes localizações espaciais tendo em conta as funções de base atómicas e núcleos envolvidos. Esta informação é apresentada na tabela 3.2.

Tabela 3.2. Localização espacial das componente cinética e potencial das energias monoelectrónicas.

Localização espacial das funções de base χ_p e χ_q	Localização espacial da componente cinética ($P_{pq}T_{pq}$)	Localização espacial da componente potencial ($P_{pq}V_{pq}^{A}$)
$\chi_p \in$A e $\chi_q \in$A	Intra-fragmento A	Intra-fragmento A
$\chi_p \in$B e $\chi_q \in$B	Intra-fragmento B	Inter-fragmento A-B
$\chi_p \in$A e $\chi_q \in$B	Inter-fragmento A-B	Inter-fragmento A-B
$\chi_p \in$B e $\chi_q \in$C	Inter-fragmento B-C	Inter-fragmento A-B-C

i) Em alguns casos, a formação do agregado molecular ocorre através de reacções químicas envolvendo a transferência de átomos. Para descrever este tipo de ocorrências, o modelo é definido de uma forma flexível de forma a permitir a alteração da composição dos fragmentos durante a formação do agregado molecular.

No presente modelo, a formação do agregado molecular a partir dos seus fragmentos isolados pode envolver os seguintes tipos de estados: estado não ligante, estado não covalente ligante e estado covalente. As configurações electrónicas podem diferir nos três tipos de estados. Consequentemente, serão usadas as designações k_{∞}, k_{no} e k_{ov} para representar as configurações electrónicas do sistema nos estados do tipo não ligante, não covalente ligante e covalente, respectivamente.

Nos dois primeiros estados, não existe qualquer tipo de ambiguidade na decomposição espacial das populações electrónicas e electrónicas de par nem das correspondentes energias monoelectrónicas e bielectrónicas. No último estado, as aproximações g) e h) são usadas para lidar com a arbitrariedade inerente à decomposição espacial das quantidades associadas à sobreposição electrónica.

O formalismo detalhado da presente metodologia de decomposição é apresentado em seguida:

Num estado do tipo não ligante, os fragmentos encontram-se a distância infinita. Consequentemente, as populações electrónicas e electrónicas de par associadas à sobreposição inter-fragmento são nulas. Neste contexto, as populações electrónicas podem ser expressas em termos de componentes intra-fragmento:

$$\sum_{A=1}^{t} P^{\infty,k_{\infty}}(A) = n \qquad (3.280)$$

Na equação (3.280), $P^{\infty,k_{\infty}}(A)$ é a população electrónica do fragmento no estado do tipo não ligante, definida como,

$$P^{\infty,k_\infty}(A) = \sum_{p=n_A^\infty}^{m_A^\infty} P_{pp}^{\infty,k_\infty} + 2\sum_{p=n_A^\infty}^{m_A^\infty}\sum_{q=n_A^\infty}^{p-1} P_{pq}^{\infty,k_\infty} S_{pq} = q_A^\infty \qquad (3.281)$$

onde $\chi_{n_A^\infty}$ e $\chi_{m_A^\infty}$ representam a primeira e última funções de base atómicas do fragmento A e q_A^∞ é a sua carga electrónica no estado não ligante considerado. Neste estado, a população electrónica de um fragmento A, que representa o número de pares electrónicos em que ambos os elementos pertencem a este fragmento, é dada por:

$$D^{\infty,k_\infty}(A) = \frac{1}{2}\sum_{p=n_A^\infty}^{m_A^\infty}\sum_{q=n_A^\infty}^{p}\sum_{r=n_A^\infty}^{p}\sum_{s=n_A^\infty}^{lim} f_{pqrs}^{\infty,k_\infty} S_{pq}S_{rs} = \frac{q_A^\infty (q_A^\infty - 1)}{2} \qquad (3.282)$$

Tendo em conta a natureza deste estado, as únicas componentes inter-fragmento das populações electrónicas de par não nulas correspondem aos pares em que um elemento pertence a um fragmento A e o outro pertence a um fragmento diferente B:

$$D^{\infty,k_\infty}(AB) = \frac{1}{2}\sum_{p=n_A^\infty}^{m_A^\infty}\sum_{q=n_A^\infty}^{p-1}\sum_{r=n_B^\infty}^{m_B^\infty}\sum_{s=n_B^\infty}^{r-1} f_{pqrs}^{\infty,k_\infty} S_{pq}S_{rs} = q_A^\infty\, q_B^\infty \qquad (3.283)$$

Nas equações (3.282) e (3.283), os coeficientes têm o mesmo significado previamente estabelecido na equação (3.276) e na tabela 3.1 e estão associados ao estado do tipo não ligante.

As populações electrónicas de par podem então ser calculadas como uma soma de componentes intra- e inter-fragmento:

$$\sum_{A=1}^{t} D^{\infty,k_\infty}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1} D^{\infty,k_\infty}(AB) = \frac{n(n-1)}{2} \qquad (3.284)$$

Apesar das populações electrónicas de par inter-fragmento serem não nulas , o mesmo não acontece relativamente às energias bielectrónicas devido às distâncias entre os fragmentos serem infinitas. Consequentemente, a energia do sistema pode ser expressa em termos de componentes intra-fragmento,

$$E^{\infty}_{k_\infty} = \sum_{A=1}^{t} E^{\infty}_{k_\infty}(A) \ , \tag{3.285}$$

onde,

$$E^{\infty}_{k_\infty}(A) = \sum_{X=N_A^\infty}^{M_A^\infty} \sum_{Y=N_A^\infty}^{X-1} \frac{Z_X Z_Y}{\varepsilon r^{\infty}_{XY}} + \sum_{p=n_A^\infty}^{m_A^\infty} P^{\infty,k_\infty}_{pp} h^{A}_{pp} + 2 \sum_{p=n_A^\infty}^{m_A^\infty} \sum_{q=n_A^\infty}^{p-1} P^{\infty,k_\infty}_{pq} h^{A}_{pq}$$

$$+ \frac{1}{2} \sum_{p=n_A^\infty}^{m_A^\infty} \sum_{q=n_A^\infty}^{p} \sum_{r=n_A^\infty}^{p} \sum_{s=n_A^\infty}^{lim} f^{\infty,k_\infty}_{pqrs}(pq|rs) \tag{3.286}$$

Na equação (3.286), N^{∞}_{A} e M^{∞}_{A} são os primeiro e último núcleos do fragmento A no estado não ligante e h^{A}_{pq} é definido como:

$$h^{A}_{pq} = T_{pq} + V^{A}_{pq} \tag{3.287}$$

Num estado do tipo não covalente ligante, os fragmentos encontram-se a distâncias finitas mas as populações electrónicas de sobreposição são ainda nulas. Consequentemente, as populações electrónicas e electrónicas de par são conservadas:

$$\sum_{A=1}^{t} P^{no,k_{no}}(A) = n \tag{3.288}$$

$$P^{no,k_{no}}(A) = \sum_{p=n_A^{\infty}}^{m_A^{\infty}} P_{pp}^{no,k_{no}} + 2\sum_{p=n_A^{\infty}}^{m_A^{\infty}}\sum_{q=n_A^{\infty}}^{p-1} P_{pq}^{no,k_{no}} S_{pq} = q_A^{\infty} = P^{\infty,k_{\infty}}(A) \qquad (3.289)$$

$$\sum_{A=1}^{t} D^{no,k_{no}}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1} D^{no,k_{no}}(AB) = \frac{n(n-1)}{2} \qquad (3.290)$$

$$D^{no,k_{no}}(A) = \frac{1}{2}\sum_{p=n_A^{\infty}}^{m_A^{\infty}}\sum_{q=n_A^{\infty}}^{p}\sum_{r=n_A^{\infty}}^{p}\sum_{s=n_A^{\infty}}^{lim} f_{pqrs}^{no,k_{no}} S_{pq} S_{rs} = \frac{q_A^{\infty}(q_A^{\infty}-1)}{2} = D^{\infty,k_{\infty}}(A) \qquad (3.291)$$

$$D^{no,k_{no}}(AB) = \frac{1}{2}\sum_{p=n_A^{\infty}}^{m_A^{\infty}}\sum_{q=n_A^{\infty}}^{p-1}\sum_{r=n_B^{\infty}}^{m_B^{\infty}}\sum_{s=n_B^{\infty}}^{r-1} f_{pqrs}^{no,k_{no}} S_{pq} S_{rs} = q_A^{\infty} q_B^{\infty} = D^{\infty,k_{\infty}}(AB) \qquad (3.292)$$

Neste caso, as interacções coulômbicas inter-fragmento deixam de ser nulas. Isto origina efeitos de polarização que preservam as populações electrónicas e electrónicas de par dos fragmentos mas alteram a sua distribuição pelas diferentes funções de base. Consequentemente, os elementos de matriz ($P_{pq}^{no,k_{no}}$ e $f_{pqrs}^{no,k_{no}}$) diferem normalmente dos correspondentes elementos ($P_{pq}^{\infty,k_{\infty}}$ e $f_{pqrs}^{\infty,k_{\infty}}$). Também podem ocorrer relaxações da geometria molecular dentro de cada fragmento ($r_{XY}^{no} \neq r_{XY}^{\infty}$, $X,Y \in A$).

São definidos três novos termos inter-fragmento da energia que resultam da distância entre os fragmentos se tornar finita:

. Energias de atracção entre a densidade electrónica de um fragmento A e os núcleos de outro fragmento B.

. Energias de repulsão entre as densidades electrónicas de dois fragmentos A e B.

. Energias de repulsão entre os núcleos de dois fragmentos A e B.

A energia do sistema pode então ser calculada pela seguinte equação,

$$E_{k_{no}}^{no} = \sum_{A=1}^{t} E_{k_{no}}^{no}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1} E_{k_{no}}^{no}(AB)\ , \qquad (3.293)$$

onde,

$$E^{no}(A) = \sum_{X=N_A^{\infty}}^{M_A^{\infty}} \sum_{Y=N_A^{\infty}}^{X-1} \frac{Z_X Z_Y}{\varepsilon r_{XY}^{no}} + \sum_{p=n_A^{\infty}}^{m_A^{\infty}} P_{pp}^{no,k_{no}} h_{pp}^{A} + 2 \sum_{p=n_A^{\infty}}^{m_A^{\infty}} \sum_{q=n_A^{\infty}}^{p-1} P_{pq}^{no,k_{no}} h_{pq}^{A}$$

$$+ \frac{1}{2} \sum_{p=n_A^{\infty}}^{m_A^{\infty}} \sum_{q=n_A^{\infty}}^{p} \sum_{r=n_A^{\infty}}^{p} \sum_{s=n_A^{\infty}}^{lim} f_{pqrs}^{no,k_{no}} (pq|rs) \ , \qquad (3.294)$$

e

$$E^{no}(AB) = \sum_{X=N_A^{\infty}}^{M_A^{\infty}} \sum_{Y=N_B^{\infty}}^{M_B^{\infty}} \frac{Z_X Z_Y}{\varepsilon r_{XY}^{no}} + \sum_{p=n_A^{\infty}}^{m_A^{\infty}} P_{pp}^{no,k_{no}} V_{pp}^{B} + \sum_{p=n_B^{\infty}}^{m_B^{\infty}} P_{pp}^{no,k_{no}} V_{pp}^{A}$$

$$+ 2 \left(\sum_{p=n_A^{\infty}}^{m_A^{\infty}} \sum_{q=n_A^{\infty}}^{p-1} P_{pq}^{no,k_{no}} V_{pq}^{B} + \sum_{p=n_B^{\infty}}^{m_B^{\infty}} \sum_{q=n_B^{\infty}}^{p-1} P_{pq}^{no,k_{no}} V_{pq}^{A} \right)$$

$$+ \frac{1}{2} \sum_{p=n_A^{\infty}}^{m_A^{\infty}} \sum_{q=n_A^{\infty}}^{p-1} \sum_{r=n_B^{\infty}}^{m_B^{\infty}} \sum_{s=n_B^{\infty}}^{r-1} f_{pqrs}^{no,k_{no}} (pq|rs) \ . \qquad (3.295)$$

Dentro da aproximação da super-molécula, a energia de estabilização ($\Delta E_{k_{no}}^{no}$) pode ser definida como,

$$\Delta E_{k_{no}}^{no} = E_{k_{no}}^{no} - E_{k_{\infty}}^{\infty} \ , \qquad (3.296)$$

e pode ser decomposta de acordo com a equação,

$$\Delta E_{k_{no}}^{no} = \sum_{A=1}^{t} \Delta E_{k_{no}}^{no}(A) + \sum_{A=1}^{t} \sum_{B=1}^{A-1} E_{k_{no}}^{no}(AB) \ , \qquad (3.297)$$

onde $\Delta E^{no}_{k_{no}}(A) = E^{no}_{k_{no}}(A) - E^{\infty}_{k_{\infty}}(A)$. (3.298)

Na equação (3.298), $\Delta E^{no}_{k_{no}}(A)$ representa a variação de energia do fragmento A entre os estados do tipo não ligante e não covalente ligante, devida aos efeitos de alteração da configuração electrónica ($k_{\infty} \rightarrow k_{no}$), polarização e relaxação da geometria interna, que são induzidos pelas interacções coulômbicas com os outros fragmentos. Na mesma equação, $E^{no}_{k_{no}}(AB)$ representa a interacção coulômbica entre os fragmentos A e B num estado do tipo não covalente ligante.

Num estado do tipo covalente, as componentes inter-fragmento das populações electrónicas e electrónicas de par deixam de se nulas. Podem ainda ocorrer transferências de carga e de grupos de átomos entre os fragmentos.

Quando alguns átomos são transferidos entre o fragmento A e outros fragmentos, os seus núcleos diferem num estado do tipo covalente e nos outros tipos de estados.

Num estado covalente, o conceito de população electrónica de um fragmento é de alguma forma arbitrário. Usando a aproximação g), apenas a população electrónica intrínseca ($P^{ov,k_{ov}}(A)$) é associada ao fragmento A. Consequentemente, as populações electrónicas de um fragmento normalmente diferem num estado do tipo covalente e nos outros estados, devido a efeitos de sobreposição e transferência de carga electrónicas com outros fragmentos.

As populações electrónicas podem agora ser decompostas em componentes intra-fragmento ($P^{ov,k_{ov}}(A)$) e inter-fragmento ($P^{ov,k_{ov}}(AB)$):

$$\sum_{A=1}^{t} P^{ov,k_{ov}}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1} P^{ov,k_{ov}}(AB) = n \ , \tag{3.299}$$

$$P^{ov,k_{ov}}(A) = \sum_{p=n_A^{ov}}^{m_A^{ov}} P^{ov,k_{ov}}_{pp} + 2\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_A^{ov}}^{p-1} P^{ov,k_{ov}}_{pq} S_{pq} = q^{ov}(A) \ , \tag{3.300}$$

$$P^{ov,k_{ov}}(AB) = 2\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_B^{ov}}^{m_B^{ov}} P_{pq}^{ov,k_{ov}}\, S_{pq} = q^{ov}(AB)\ . \qquad (3.301$$

Nas equações anteriores, $\chi_{n_A^{ov}}$ e $\chi_{m_A^{ov}}$ são a primeira e última funções de base do fragmento A, $q^{ov}(A)$ é carga electrónica intrínseca do fragmento A e $q^{ov}(AB)$ é carga electrónica de sobreposição entre os fragmentos A e B, num estado do tipo covalente.

Como as populações electrónicas de sobreposição são não nulas, existem agora componentes das populações electrónicas de par onde um ou dois elementos estão associados à densidade de sobreposição inter-fragmento. Além disso, podem existir também componentes de ordem mais elevada associados a trímeros e tetrâmeros.

As populações electrónicas de par são então expandidas como uma soma de componentes intra-fragmento e inter-fragmento (bicêntricas, tricêntricas e tetracêntricas):

$$\sum_{A=1}^{t} D^{ov,k_{ov}}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1} D^{ov,k_{ov}}(AB) + \sum_{A=1}^{t}\sum_{B=1}^{A-1}\sum_{C=1}^{B-1} D^{ov,k_{ov}}(ABC)$$
$$+ \sum_{A=1}^{t}\sum_{B=1}^{A-1}\sum_{C=1}^{B-1}\sum_{D=1}^{C-1} D^{ov,k_{ov}}(ABCD) = \frac{n(n-1)}{2} \qquad (3.302)$$

Na equação (3.302), as componentes intra-fragmento são dadas por,

$$D^{ov,k_{ov}}(A) = \frac{1}{2}\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_A^{ov}}^{p}\sum_{r=n_A^{ov}}^{p}\sum_{s=n_A^{ov}}^{lim} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs}\ , \qquad (3.303)$$

as componentes inter-fragmento bicêntricas são calculadas como,

$$D^{ov,k_{ov}}(AB) = \frac{1}{2}\left[\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_A^{ov}}^{p-1}\sum_{r=n_B^{ov}}^{m_B^{ov}}\sum_{s=n_B^{ov}}^{r-1} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs} + \sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_A^{ov}}^{p-1}\sum_{r=n_A^{ov}}^{m_A^{ov}}\sum_{s=n_B^{ov}}^{m_B^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs}\right.$$

$$\left. + \sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_B^{ov}}^{m_B^{ov}}\sum_{r=n_B^{ov}}^{m_B^{ov}}\sum_{s=n_B^{ov}}^{r-1} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs} + \sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_B^{ov}}^{m_B^{ov}}\sum_{r=n_A^{ov}}^{m_A^{ov}}\sum_{s=n_B^{ov}}^{m_B^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs}\right], \quad (3.304)$$

as componentes inter-fragmento tricêntricas são definidas como,

$$D^{ov,k_{ov}}(ABC) = \frac{1}{2}\left[\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_A^{ov}}^{p-1}\sum_{r=n_B^{ov}}^{m_B^{ov}}\sum_{s=n_C^{ov}}^{m_C^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs} + \sum_{p=n_B^{ov}}^{m_B^{ov}}\sum_{q=n_B^{ov}}^{p-1}\sum_{r=n_A^{ov}}^{m_A^{ov}}\sum_{s=n_C^{ov}}^{m_C^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs}\right.$$

$$+ \sum_{p=n_C^{ov}}^{m_C^{ov}}\sum_{q=n_C^{ov}}^{p-1}\sum_{r=n_A^{ov}}^{m_A^{ov}}\sum_{s=n_B^{ov}}^{m_B^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs} + \sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_B^{ov}}^{m_B^{ov}}\sum_{r=n_B^{ov}}^{m_B^{ov}}\sum_{s=n_C^{ov}}^{m_C^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs}$$

$$\left. + \sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_C^{ov}}^{m_C^{ov}}\sum_{r=n_C^{ov}}^{m_C^{ov}}\sum_{s=n_B^{ov}}^{m_B^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs} + \sum_{p=n_B^{ov}}^{m_B^{ov}}\sum_{q=n_A^{ov}}^{m_A^{ov}}\sum_{r=n_A^{ov}}^{m_A^{ov}}\sum_{s=n_C^{ov}}^{m_C^{ov}} f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs}\right], \quad (3.305)$$

e as componentes inter-fragmento tetracêntricas são obtidas através da equação,

$$D^{ov,k_{ov}}(ABCD) = \frac{1}{2}\left[\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_B^{ov}}^{m_B^{ov}}\sum_{r=n_C^{ov}}^{m_C^{ov}}\sum_{s=n_D^{ov}}^{m_D^{ov}} \left(f_{pqrs}^{ov,k_{ov}}\, S_{pq}S_{rs} + f_{prqs}^{ov,k_{ov}}\, S_{pr}S_{qs} + f_{psqr}^{ov,k_{ov}} S_{ps}S_{qr}\right)\right] \quad (3.306)$$

Neste tipo de estados, são definidos um conjunto de novos termos da energia devido às populações electrónicas de sobreposição serem não nulas:

. Energias cinéticas monoelectrónicas associadas à densidade de sobreposição entre dois fragmentos A e B (incluídas em $E_{k_{ov}}^{ov}(AB)$).

. Energias de atracção monoelectrónicas entre a densidade de sobreposição de dois fragmentos A e B e os núcleos de ambos os fragmentos (incluídas em $E_{k_{ov}}^{ov}(AB)$).

. Energias de atracção monoelectrónicas entre a densidade de sobreposição de dois fragmentos A e B e os núcleos de um terceiro fragmento C (incluídas em $E_{k_{ov}}^{ov}(ABC)$).

. Energias bielectrónicas associadas com pares electrónicos onde um elemento pertence ao fragmento A e o outro pertence à densidade de sobreposição entre os fragmentos A e B (incluídas em $E_{k_{ov}}^{ov}(AB)$).

. Energias bielectrónicas associadas com pares electrónicos onde ambos os elementos pertencem à densidade de sobreposição entre os fragmentos A e B (incluídas em $E_{k_{ov}}^{ov}(AB)$).

. Energias bielectrónicas associadas com pares electrónicos onde um elemento pertence ao fragmento A e o outro pertence à densidade de sobreposição entre os fragmentos B e C (incluídas em $E_{k_{ov}}^{ov}(ABC)$).

. Energias bielectrónicas associadas com pares electrónicos onde um elemento pertence à densidade de sobreposição entre os fragmentos A e B e o outro pertence à densidade de sobreposição entre os fragmentos A e C (incluídas em $E_{k_{ov}}^{ov}(ABC)$).

. Energias bielectrónicas associadas com pares electrónicos onde um elemento pertence à densidade de sobreposição entre os fragmentos A e B e o outro pertence à densidade de sobreposição entre os fragmentos C e D (incluídas em $E_{k_{ov}}^{ov}(ABCD)$).

Num estado covalente, a energia pode ser decomposta de acordo com a equação:

$$E_{k_{ov}}^{ov} = \sum_{A=1}^{t} E_{k_{ov}}^{ov}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1} E_{k_{ov}}^{ov}(AB) + \sum_{A=1}^{t}\sum_{B=1}^{A-1}\sum_{C=1}^{B-1} E_{k_{ov}}^{ov}(ABC) + \sum_{A=1}^{t}\sum_{B=1}^{A-1}\sum_{C=1}^{B-1}\sum_{D=1}^{C-1} E_{k_{ov}}^{ov}(ABCD) \; . \qquad (3.307)$$

As componentes intra-fragmento são dadas por:

$$E^{ov}_{k_{ov}}(A) = \sum_{X=N_A^{ov}}^{M_A^{ov}} \sum_{Y=N_A^{ov}}^{X-1} \frac{Z_X Z_Y}{\varepsilon r^{ov}_{XY}} + \sum_{p=n_A^{ov}}^{m_A^{ov}} P^{ov,k_{ov}}_{pp} h^A_{pp} + 2 \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_A^{ov}}^{p-1} P^{ov,k_{ov}}_{pq} h^A_{pq}$$

$$+ \frac{1}{2} \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_A^{ov}}^{p} \sum_{r=n_A^{ov}}^{p} \sum_{s=n_A^{ov}}^{lim} f^{ov,k_{ov}}_{pqrs}(pq|rs) \ , \qquad (3.308)$$

As componentes inter-fragmento bicêntricas são calculadas como:

$$E^{ov}_{k_{ov}}(AB) = \sum_{X=N_A^{ov}}^{M_A^{ov}} \sum_{Y=N_B^{ov}}^{M_B^{ov}} \frac{Z_X Z_Y}{\varepsilon r^{ov}_{XY}} + \sum_{p=n_A^{ov}}^{m_A^{ov}} P^{ov,k_{ov}}_{pp} V^B_{pp} + \sum_{p=n_B^{ov}}^{m_B^{ov}} P^{ov,k_{ov}}_{pp} V^A_{pp}$$

$$+ 2 \left[\sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_A^{ov}}^{p-1} P^{ov,k_{ov}}_{pq} V^B_{pq} + \sum_{p=n_B^{ov}}^{m_B^{ov}} \sum_{q=n_B^{ov}}^{p-1} P^{ov,k_{ov}}_{pq} V^A_{pq} + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_B^{ov}}^{m_B^{ov}} P^{ov,k_{ov}}_{pq} (h^A_{pq} + h^B_{pq}) \right]$$

$$+ \frac{1}{2} \left[\sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_A^{ov}}^{p-1} \sum_{r=n_B^{ov}}^{m_B^{ov}} \sum_{s=n_B^{ov}}^{r-1} f^{ov,k_{ov}}_{pqrs}(pq|rs) + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_A^{ov}}^{p-1} \sum_{r=n_A^{ov}}^{m_A^{ov}} \sum_{s=n_B^{ov}}^{m_B^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) \right.$$

$$\left. + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_B^{ov}}^{m_B^{ov}} \sum_{r=n_B^{ov}}^{m_B^{ov}} \sum_{s=n_B^{ov}}^{r-1} f^{ov,k_{ov}}_{pqrs}(pq|rs) + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_B^{ov}}^{m_B^{ov}} \sum_{r=n_A^{ov}}^{m_A^{ov}} \sum_{s=n_B^{ov}}^{m_B^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) \right] \qquad (3.309)$$

As componentes inter-fragmento tricêntricas são definidas como:

$$E^{ov}_{k_{ov}}(ABC) = 2 \left[\sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_B^{ov}}^{m_B^{ov}} P^{ov,k_{ov}}_{pq} V^C_{pq} + \sum_{p=n_B^{ov}}^{m_B^{ov}} \sum_{q=n_C^{ov}}^{m_C^{ov}} P^{ov,k_{ov}}_{pq} V^A_{pq} + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_C^{ov}}^{m_C^{ov}} P^{ov,k_{ov}}_{pq} V^B_{pq} \right]$$

$$+ \frac{1}{2} \left[\sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_A^{ov}}^{p-1} \sum_{r=n_B^{ov}}^{m_B^{ov}} \sum_{s=n_C^{ov}}^{m_C^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) + \sum_{p=n_B^{ov}}^{m_B^{ov}} \sum_{q=n_B^{ov}}^{p-1} \sum_{r=n_A^{ov}}^{m_A^{ov}} \sum_{s=n_C^{ov}}^{m_C^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) \right.$$

$$+ \sum_{p=n_C^{ov}}^{m_C^{ov}} \sum_{q=n_C^{ov}}^{p-1} \sum_{r=n_A^{ov}}^{m_A^{ov}} \sum_{s=n_B^{ov}}^{m_B^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_B^{ov}}^{m_B^{ov}} \sum_{r=n_B^{ov}}^{m_B^{ov}} \sum_{s=n_C^{ov}}^{m_C^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs)$$

$$\left. + \sum_{p=n_A^{ov}}^{m_A^{ov}} \sum_{q=n_C^{ov}}^{m_C^{ov}} \sum_{r=n_C^{ov}}^{m_C^{ov}} \sum_{s=n_B^{ov}}^{m_B^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) + \sum_{p=n_B^{ov}}^{m_B^{ov}} \sum_{q=n_A^{ov}}^{m_A^{ov}} \sum_{r=n_A^{ov}}^{m_A^{ov}} \sum_{s=n_C^{ov}}^{m_C^{ov}} f^{ov,k_{ov}}_{pqrs}(pq|rs) \right] \qquad (3.310)$$

As componentes inter-fragmento tetracêntricas são obtidas através da equação:

$$E^{ov}_{k_{ov}}(ABCD)=\frac{1}{2}\left[\sum_{p=n_A^{ov}}^{m_A^{ov}}\sum_{q=n_B^{ov}}^{m_B^{ov}}\sum_{r=n_C^{ov}}^{m_C^{ov}}\sum_{s=n_D^{ov}}^{m_D^{ov}}\left\{ f^{ov,k_{ov}}_{pqrs}(pq|rs)+f^{ov,k_{ov}}_{prqs}(pr|qs)+f^{ov,k_{ov}}_{psqr}(ps|qr)\right\}\right] \quad (3.311)$$

Nas equações (3.308) a (3.311), n_A^{ov} e m_A^{ov} representam o primeiro e último núcleos do fragmento A, num estado do tipo covalente.

Dentro da aproximação da super-molécula, a energia de estabilização é definida como,

$$\Delta E^{ov}_{k_{ov}} = E^{ov}_{k_{ov}} - E^{\infty}_{k_{\infty}} \quad (3.312)$$

e pode ser decomposta de acordo com a equação,

$$\Delta E^{ov}_{k_{ov}} = \sum_{A=1}^{t}\Delta E^{ov}_{k_{ov}}(A) + \sum_{A=1}^{t}\sum_{B=1}^{A-1}E^{ov}_{k_{ov}}(AB) + \sum_{A=1}^{t}\sum_{B=1}^{A-1}\sum_{C=1}^{B-1}E^{ov}_{k_{ov}}(ABC) + \sum_{A=1}^{t}\sum_{B=1}^{A-1}\sum_{C=1}^{B-1}\sum_{D=1}^{C-1}E^{ov}_{k_{ov}}(ABCD)\ , \quad (3.313)$$

onde $\Delta E^{ov}_{k_{ov}}(A) = E^{ov}_{k_{ov}}(A) - E^{\infty}_{k_{\infty}}(A)$. (3.314)

Na equação (3.314), $\Delta E^{ov}_{k_{ov}}(A)$ representa a variação de energia do fragmento A entre os estados do tipo não ligante e covalente, devida aos efeitos de alteração da configuração electrónica ($k_{\infty} \rightarrow k_{ov}$), polarização, relaxação da geometria interna, transferência de carga electrónica e transferência de átomos, que são induzidos pelas interacções coulômbicas e sobreposição de carga electrónica com os outros fragmentos. Na mesma equação, $E^{ov}_{k_{ov}}(AB)$, $E^{ov}_{k_{ov}}(ABC)$ e $E^{ov}_{k_{ov}}(ABCD)$ representam, respectivamente, as energias de interacção entre fragmentos A e B, dentro do trímero (ABC) e dentro do tetrâmero (ABCD), num estado do tipo covalente.

As reacções envolvendo a transferência de grupos ligados covalentemente têm sido extensivamente estudadas quer por métodos experimentais [117-126] quer por métodos teóricos [127-155]. Num agregado molecular, onde ocorra uma reacção de transferência de um só grupo covalentemente ligado, podem ser consideradas duas regiões: uma *região reactiva* que inclui os dois fragmentos envolvidos no processo de transferência e uma *região não reactiva* que não está directamente envolvida no processo. No vácuo, apenas existe a região reactiva e a transferência de um grupo G entre dois fragmentos A e B pode ser descrita pela seguinte equação química:

$$\underset{\substack{\text{Estado inicial}\\ \text{não ligante } (\infty_1)}}{A - G + B} \xrightarrow{\text{Associação}} \underset{\substack{\text{Complexo ligante 1}\\ (BC_1)}}{[A - G : B]} \xrightarrow{\text{Transferência do grupo G}} \underset{\substack{\text{Complexo ligante 2}\\ (BC_2)}}{[A : G - B]} \xrightarrow{\text{Dissociação}} \underset{\substack{\text{Estado final}\\ \text{não ligante } (\infty_2)}}{A + G - B} \qquad (3.315)$$

Escolhendo o estado inicial não ligante como referência, a energia de estabilização pode ser decomposta como,

$$\Delta E^{\Gamma}_{k_\Gamma} = E^{\Gamma}_{k_\Gamma} - E^{\infty_1}_{k_{\infty_1}} \; (\Gamma = BC_1,\ BC_2 \text{ ou } \infty_2), \qquad (3.316)$$

e pode ser decomposta de acordo com a equação,

$$\Delta E^{\Gamma}_{k_\Gamma} = \Delta E^{\Gamma}_{k_\Gamma,\text{intra}} + \Delta E^{\Gamma}_{k_\Gamma,\text{inter}} \qquad (3.317)$$

Na equação (3.317), $\Delta E^{\Gamma}_{k_\Gamma,\text{intra}}$ é a componente intra-fragmento da energia de estabilização e representa a estabilidade intrínseca dos fragmentos do estado Γ em relação aos fragmentos do estado inicial,

$$\Delta E^{\Gamma}_{k_\Gamma,\text{intra}} = \Delta E^{\Gamma}_{k_\Gamma}(A) + \Delta E^{\Gamma}_{k_\Gamma}(B)\ , \qquad (3.318)$$

com,

$$\Delta E^{\Gamma}_{k_\Gamma}(A) = E^{\Gamma}_{k_\Gamma}(A) - E^{\infty_1}_{k_{\infty_1}}(A)\ , \qquad (3.319)$$

e

$$\Delta E^{\Gamma}_{k_\Gamma}(B) = E^{\Gamma}_{k_\Gamma}(B) - E^{\infty_1}_{k_{\infty_1}}(B) \ . \qquad (3.320)$$

Deve salientar-se que o grupo G deve ser incluído no fragmento A no estado inicial não ligante e no complexo ligante 1 e no fragmento B nos outros dois estados. $E^{\infty_1}_{k_{\infty_1}}(A)$ e $E^{\infty_1}_{k_{\infty_1}}(B)$ são calculadas pela equação (3.286). $E^{\Gamma}_{k_\Gamma}(A)$ e $E^{\Gamma}_{k_\Gamma}(B)$ são calculadas pelas equações (3.286), (3.294) e (3.308) se Γ é, respectivamente, um estado não ligante, um estado não covalente ligante ou um estado covalente. $\Delta E^{\Gamma}_{inter}$ é a componente inter-fragmento da energia de estabilização e representa a energia de interacção entre os fragmentos A e B no estado Γ,

$$\Delta E^{\Gamma}_{k_\Gamma,inter} = E^{\Gamma}_{k_\Gamma}(AB) \ , \qquad (3.321)$$

sendo calculada pela equação (3.295) se Γ é um estado não covalente ligante e pela equação (3.309) se Γ é um estado covalente. Esta componente é obviamente nula no estado final não ligante.

Os complexos ligantes têm um especial interesse no estudo deste tipo de sistemas. A sua energia de estabilização relativa ($\Delta\Delta E^{BC_1 \to BC_2}$) pode ser calculada como,

$$\Delta\Delta E^{BC_1 \to BC_2} = \Delta E^{BC_2}_{k_{BC_2}} - \Delta E^{BC_1}_{k_{BC_1}} = E^{BC_2}_{k_{BC_2}} - E^{BC_1}_{k_{BC_1}} \ , \qquad (3.322)$$

e pode ser decomposta de acordo com a equação:

$$\Delta\Delta E^{BC_1 \to BC_2} = \Delta\Delta E^{BC_1 \to BC_2}_{intra} + \Delta\Delta E^{BC_1 \to BC_2}_{inter} \qquad (3.323)$$

Na equação (3.323), $\Delta\Delta E^{BC_1 \to BC_2}_{intra}$ é a componente inter-fragmento $\Delta\Delta E^{BC_1 \to BC_2}$, definida como,

$$\Delta\Delta E^{BC_1 \to BC_2}_{intra} = (E^{BC_2}_{k_{BC_2}}(A) + E^{BC_2}_{k_{BC_2}}(B)) - (E^{BC_1}_{k_{BC_1}}(A) + E^{BC_1}_{k_{BC_1}}(B)) \ , \qquad (3.324)$$

e $\Delta\Delta E_{inter}^{BC_1 \rightarrow BC_2}$ é a componente inter-fragmento de $\Delta\Delta E^{BC_1 \rightarrow BC_2}$, dada por,

$$\Delta\Delta E_{inter}^{BC_1 \rightarrow BC_2} = E^{BC_2}(AB) - E^{BC_1}(AB) \qquad (3.325)$$

A componente inter-fragmento representa a diferença de energia de interacção entre os pares (A-G:B) e (A:G-B). Normalmente, esta componente é essencialmente electrostática e, consequentemente, favorece separação de cargas nos fragmentos.

A componente intra-fragmento representa a diferença de energia entre os fragmentos (A, G-B) e (A-G, B). Esta componente está associada à diferença de afinidade intrínseca para o grupo G entre os fragmentos A e B. Se $\Delta\Delta E_{intra}^{BC_1 \rightarrow BC_2} < 0$, o fragmento B apresenta uma maior afinidade intrínseca para o grupo G do que o fragmento A. Se $\Delta\Delta E_{intra}^{BC_1 \rightarrow BC_2} > 0$, verifica-se a situação inversa.

Na presença de um meio (proteína, água, etc.) devem ser considerados dois efeitos adicionais:

. *Efeito dieléctrico* - A magnitude das interacções coulômbicas depende do valor da constante dieléctrica (ε) do meio. Em meios hidrofóbicos ($\varepsilon \rightarrow 1$) e no vácuo ($\varepsilon=1$), a magnitude destas interacções é similar. Em meios hidrofílicos ($\varepsilon>>1$), estas interacções são muito mais fracas.

. *Efeito de fragmento* - Certas regiões do meio (por exemplo a primeira esfera de hidratação numa solução aquosa ou os aminoácidos vizinhos num ambiente proteico) podem ter interacção directa com a região reactiva. Neste caso, estas regiões devem ser incluídas num ou mais fragmentos não reactivos e, consequentemente, devem ser consideradas duas componentes adicionais na energia de estabilização: uma componente de interacção entre o meio e a região reactiva e uma componente associada à reorganização do meio. Neste contexto, os meios hidrofílicos estabilizam preferencialmente estruturas carregadas enquanto os meios hidrofóbicos estabilizam preferencialmente estruturas neutras.

A transferência de protão é uma das reacções principais deste tipo e desempenha um papel chave em muitos mecanismos biológicos tais como reacções enzimáticas e interacções droga-receptor [156-158]. Normalmente, uma reacção deste tipo pode ser expressa por uma equação química do tipo (3.315) [159],

$$\underset{\text{Estado inicial não ligante } (\infty_1)}{A-H+B} \xrightarrow{\text{Associação}} \underset{\text{Complexo ligante 1 } (BC_1)}{[A-H:B]} \xrightarrow{\text{Transferência de protão}} \underset{\text{Complexo ligante 2 } (BC_2)}{[A:H-B]} \xrightarrow{\text{Dissociation}} \underset{\text{Estado final não ligante } (\infty_2)}{A+H-B} \quad (3.326)$$

e está associado a uma superfície de energia potencial do tipo duplo poço [159-160] (ver figura 3.5). Neste contexto, a metodologia de decomposição proposta neste trabalho é inteiramente aplicável a reacções de transferência de protão.

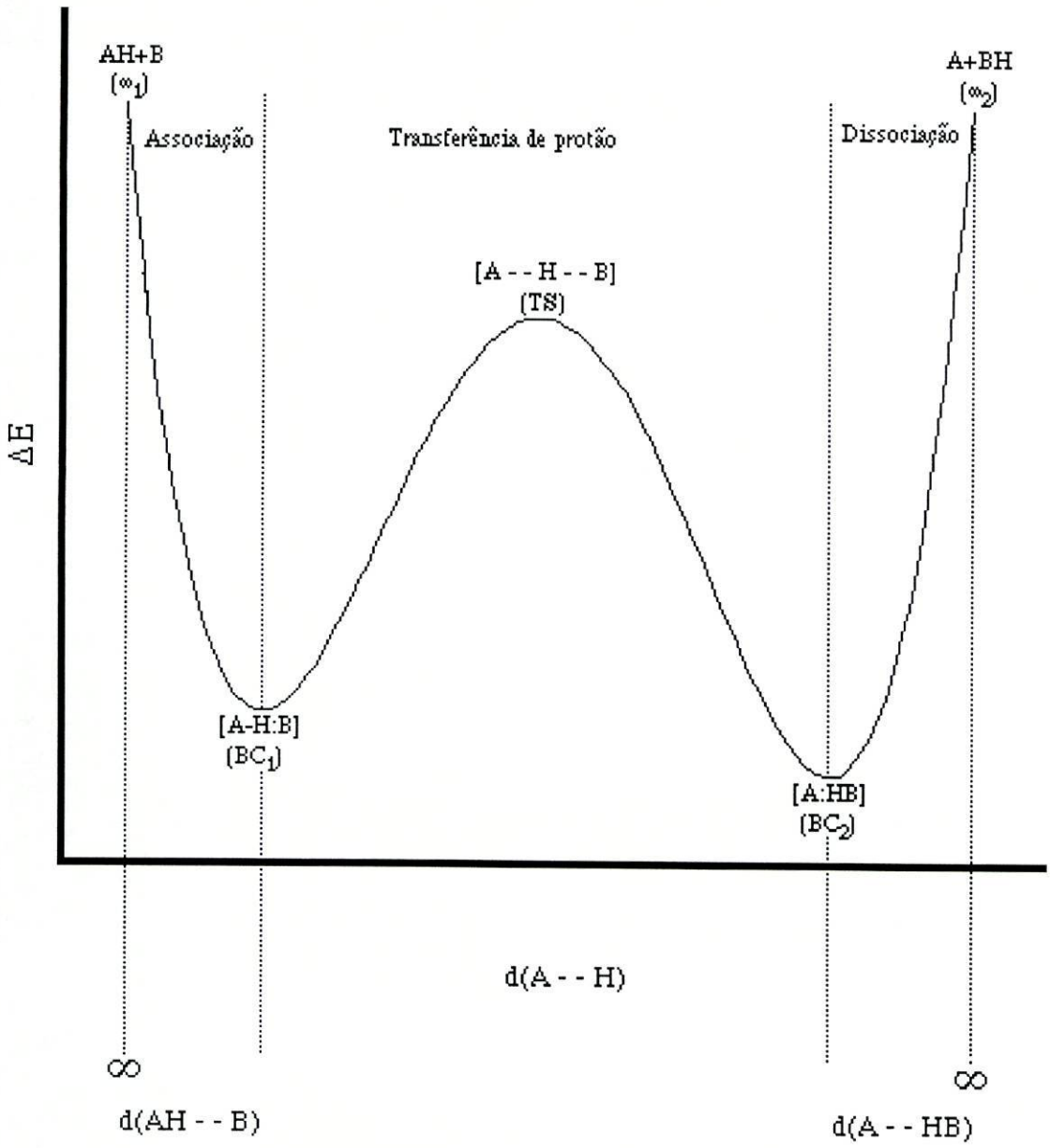


Figura 3.5. Superfície de energia potencial do tipo duplo poço para reacções de transferência de protão.

As propriedades das proteínas dependem da sua estrutura tridimensional, que por sua vez é condicionada pela carga das cadeias laterais dos aminoácidos. Reacções de transferência protónica envolvendo cadeias laterais ácidas e básicas podem provocar alterações conformacionais nas proteínas e influenciar as suas propriedades [161-162]. Neste contexto, os complexos (Arg-Carb) que se estabelecem entre as cadeias laterais da arginina e de aminoácidos do tipo carboxilato (glutamato e aspartato), referidas no capítulo 1, representam um dos mais importantes sistemas deste tipo [163-164].

Nas interacções Arg-Carb, podem ser considerados os complexos ligantes duplo ião e neutro que são habitualmente simulados usando os modelos moleculares apropriados metilguanidínio-acetato ($MGH^+:Ac^-$) e metilguanidina-ácido acético (MG:HAc), respectivamente (ver figura 3.6). Usando estes modelos moleculares, estas interacções podem ser expressas por uma equação química do tipo (3.326):

$$\underset{\substack{\text{Estado inicial}\\ \text{não ligante } (\infty_1)}}{MGH^+ + Ac^-} \xrightarrow{\text{Associação}} \underset{\substack{\text{Complexo duplo ião}\\ \text{(zwitt.)}}}{\left[MGH^+:Ac^-\right]} \xrightarrow{\text{Transferência de protão}} \underset{\substack{\text{Complexo neutro}\\ \text{(neut.)}}}{[MG:HAc]} \xrightarrow{\text{Dissociação}} \underset{\substack{\text{Estado final}\\ \text{não ligante } (\infty_2)}}{MG + HAc} \qquad (3.327)$$

Neste contexto, este sistema pode ser estudado com base numa superfície de energia potencial típica das reacções de transferência de protão apresentada na figura 3.5, onde o complexo ligante 1 (BC_1) corresponde ao duplo ião [$MGH^+:Ac^-$] e o complexo ligante 2 (BC_2) corresponde à estrutura neutra [MG:HAc].

Utilizando os modelos moleculares referidos, o método de decomposição proposto neste trabalho é também aplicável ao estudo das interacções Arg-Carb.

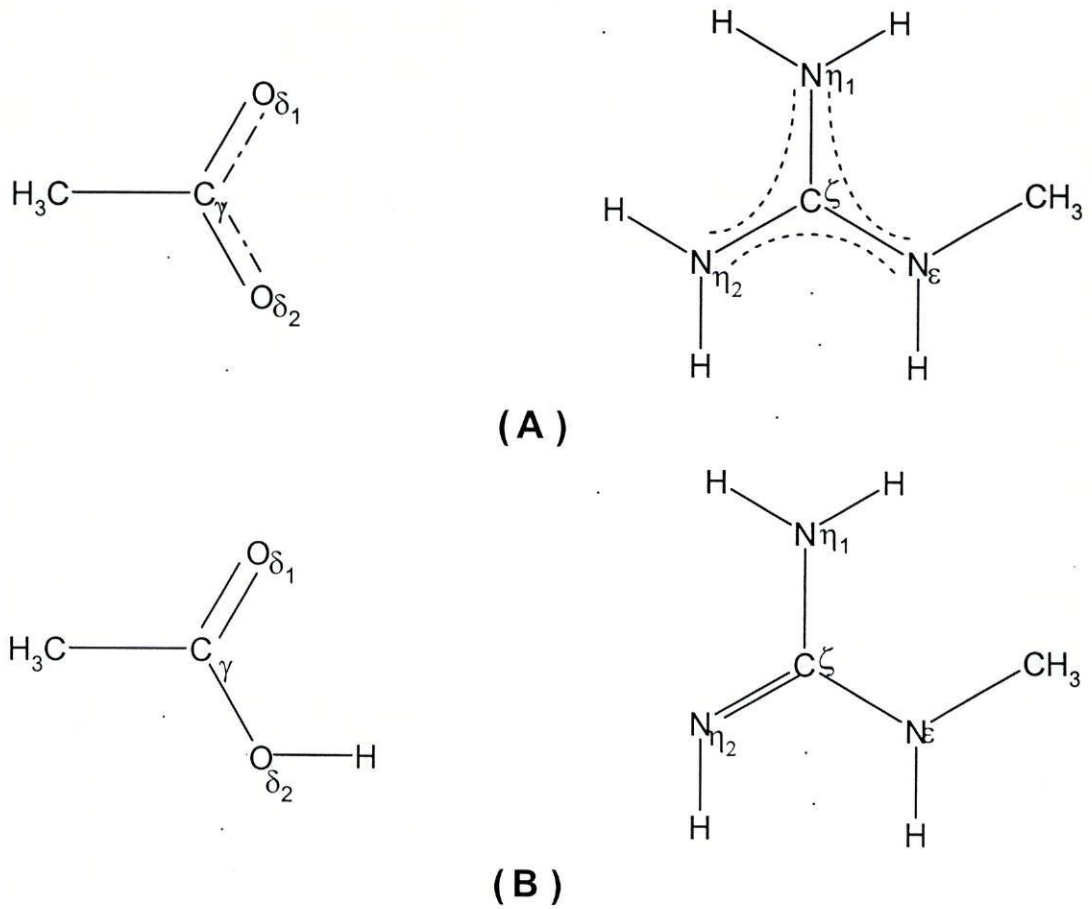


Figura 3.6. Representação esquemática dos modelos moleculares (A) metilguanidínio-acetato e (B) metilguanidina-ácido acético.

O esquema de partição aqui proposto baseia-se explicitamente na decomposição espacial das matrizes densidade e densidade de par pelo que apresenta analogias com os métodos propostos por Ruedenberg [99] e Bofill [102]. A principal vantagem da presente metodologia resulta das populações electrónicas e electrónicas de par serem calculadas dentro do formalismo dos integrais indistinguíveis o que reduz significativamente o número de termos a calcular.

Considerando um agregado molecular constituído por t fragmentos atómicos rígidos, num estado covalente, pode estabelecer-se a seguinte correspondência entre as populações electrónicas calculadas pelo métodos de Melo e Ramos e de Ruedenberg:

$$P^{ov,k_{ov}}(W) + \frac{1}{2}\sum_{Z\neq W=1}^{t} P^{ov,k_{ov}}(WZ) = \int_{X_1'=X_1} \rho_{k_{ov}}^{ov\ Mull.}(W)\, dV_1 \qquad (3.328)$$

O método de Bofill baseia-se unicamente na decomposição espacial das populações electrónicas de par pelo que, neste caso, não é possível estabelecer uma relação do tipo (3.328).

A enorme simplificação resultante do uso do formalismo dos integrais indistinguíveis é particularmente evidente se as populações electrónicas de par calculadas por este método forem comparadas com as calculadas pelos métodos de Ruedenberg e Bofill:

$$D^{ov,k_{ov}}(W) = D^{ov,k_{ov}}(WW|WW) = \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \pi_{k_{ov}}^{ov}(WW|WW)\, dV_1\, dV_2 \qquad (3.329)$$

$$\begin{aligned} D^{ov,k_{ov}}(WZ) &= D^{ov,k_{ov}}(WW|ZZ) + D^{ov,k_{ov}}(ZZ|WW) + 2\left\{D^{ov,k_{ov}}(WW|ZW) + D^{ov,k_{ov}}(WW|WZ)\right. \\ &\left. + D^{ov,k_{ov}}(ZZ|WZ) + D^{ov,k_{ov}}(ZZ|ZW)\right\} + D^{ov,k_{ov}}(ZW|ZW) + D^{ov,k_{ov}}(ZW|WZ) + D^{ov,k_{ov}}(WZ|WZ) \\ &+ D_{k_{ov}}^{ov,k_{ov}}(WZ|ZW) = \iint_{\substack{X_1'=X_1 \\ X_2'=X_2}} \left[\pi_{k_{ov}}^{ov}(WW|ZZ) + \pi_{k_{ov}}^{ov}(ZZ|WW) + 2\left\{\pi_{k_{ov}}^{ov}(WW|ZW)\right.\right. \\ &\left.\left. + \pi_{k_{ov}}^{ov}(WW|WZ) + \pi_{k_{ov}}^{ov}(ZZ|WZ) + \pi_{k_{ov}}^{ov}(ZZ|ZW)\right\} + \pi_{k_{ov}}^{ov}(ZW|ZW) + \pi_{k_{ov}}^{ov}(ZW|WZ) \right. \\ &\left. + \pi_{k_{ov}}^{ov}(WZ|WZ) + \pi_{k_{ov}}^{ov}(WZ|ZW)\right] dV_1\, dV_2 \qquad (3.330) \end{aligned}$$

$$D^{ov,k_{ov}}(WZT) = 2\left\{ D^{ov,k_{ov}}(WW|ZT) + D^{ov,k_{ov}}(WW|TZ) + D^{ov,k_{ov}}(ZZ|WT) + D^{ov,k_{ov}}(ZZ|TW) + D^{ov,k_{ov}}(TT|WZ) + D^{ov,k_{ov}}(TT|ZW) \right\} + D^{ov,k_{ov}}(WZ|ZT) + D^{ov,k_{ov}}(ZT|WZ) + D^{ov,k_{ov}}(WZ|TZ) + D^{ov,k_{ov}}(TZ|WZ) + D^{ov,k_{ov}}(ZW|ZT) + D^{ov,k_{ov}}(ZT|ZW) + D^{ov,k_{ov}}(ZW|TZ) + D^{ov,k_{ov}}(TZ|ZW) + D^{ov,k_{ov}}(WT|TZ) + D^{ov,k_{ov}}(TZ|WT) + D^{ov,k_{ov}}(WT|ZT) + D^{ov,k_{ov}}(ZT|WT) + D^{ov,k_{ov}}(TW|TZ) + D^{ov,k_{ov}}(TZ|TW) + D^{ov,k_{ov}}(TW|ZT) + D^{ov,k_{ov}}(ZT|TW) + D^{ov,k_{ov}}(ZW|WT) + D^{ov,k_{ov}}(WT|ZW) + D^{ov,k_{ov}}(ZW|TW) + D^{ov,k_{ov}}(TW|ZW) + D^{ov,k_{ov}}(WZ|WT) + D^{ov,k_{ov}}(WT|WZ) + D^{ov,k_{ov}}(WZ|TW) + D^{ov,k_{ov}}(TW|WZ) = \iint\limits_{\substack{X_1'=X_1 \\ X_2'=X_2}} \Big[2\left\{ \pi^{ov}_{k_{ov}}(WW|ZT) + \pi^{ov}_{k_{ov}}(WW|TZ) + \pi^{ov}_{k_{ov}}(ZZ|WT) + \pi^{ov}_{k_{ov}}(ZZ|TW) + \pi^{ov}_{k_{ov}}(TT|WZ) + \pi^{ov}_{k_{ov}}(TT|ZW) \right\} + \pi^{ov}_{k_{ov}}(WZ|ZT) + \pi^{ov}_{k_{ov}}(ZT|WZ) + \pi^{ov}_{k_{ov}}(WZ|TZ) + \pi^{ov}_{k_{ov}}(TZ|WZ) + \pi^{ov}_{k_{ov}}(ZW|ZT) + \pi^{ov}_{k_{ov}}(ZT|ZW) + \pi^{ov}_{k_{ov}}(ZW|TZ) + \pi^{ov}_{k_{ov}}(TZ|ZW) + \pi^{ov}_{k_{ov}}(WT|TZ) + \pi^{ov}_{k_{ov}}(TZ|WT) + \pi^{ov}_{k_{ov}}(WT|ZT) + \pi^{ov}_{k_{ov}}(ZT|WT) + \pi^{ov}_{k_{ov}}(TW|TZ) + \pi^{ov}_{k_{ov}}(TZ|TW) + \pi^{ov}_{k_{ov}}(TW|ZT) + \pi^{ov}_{k_{ov}}(ZT|TW) + \pi^{ov}_{k_{ov}}(ZW|WT) + \pi^{ov}_{k_{ov}}(WT|ZW) + \pi^{ov}_{k_{ov}}(ZW|TW) + \pi^{ov}_{k_{ov}}(TW|ZW) + \pi^{ov}_{k_{ov}}(WZ|WT) + \pi^{ov}_{k_{ov}}(WT|WZ) + \pi^{ov}_{k_{ov}}(WZ|TW) + \pi^{ov}_{k_{ov}}(TW|WZ) \Big] dV_1\, dV_2 \tag{3.331}$$

$$D^{ov,k_{ov}}(WZTV) = D^{ov,k_{ov}}(WZ|TV) + D^{ov,k_{ov}}(TV|WZ) + D^{ov,k_{ov}}(WZ|VT) + D^{ov,k_{ov}}(VT|WZ) + D^{ov,k_{ov}}(ZW|TV) + D^{ov,k_{ov}}(TV|ZW) + D^{ov,k_{ov}}(ZW|VT) + D^{ov,k_{ov}}(VT|ZW) + D^{ov,k_{ov}}(WT|ZV) + D^{ov,k_{ov}}(ZV|WT) + D^{ov,k_{ov}}(WT|VZ) + D^{ov,k_{ov}}(VZ|WT) + D^{ov,k_{ov}}(TW|ZV) + D^{ov,k_{ov}}(ZV|TW) + D^{ov,k_{ov}}(TW|VZ) + D^{ov,k_{ov}}(VZ|TW) + D^{ov,k_{ov}}(WV|ZT) + D^{ov,k_{ov}}(ZT|WV) + D^{ov,k_{ov}}(WV|TZ) + D^{ov,k_{ov}}(TZ|WV) + D^{ov,k_{ov}}(VW|ZT) + D^{ov,k_{ov}}(ZT|VW) + D^{ov,k_{ov}}(VW|TZ) + D^{ov,k_{ov}}(TZ|VW) = \iint\limits_{\substack{X_1'=X_1 \\ X_2'=X_2}} \Big[\pi^{ov}_{k_{ov}}(WZ|TV) + \pi^{ov}_{k_{ov}}(TV|WZ) + \pi^{ov}_{k_{ov}}(WZ|VT) + \pi^{ov}_{k_{ov}}(VT|WZ) + \pi^{ov}_{k_{ov}}(ZW|TV) + \pi^{ov}_{k_{ov}}(TV|ZW) + \pi^{ov}_{k_{ov}}(ZW|VT) + \pi^{ov}_{k_{ov}}(VT|ZW) + \pi^{ov}_{k_{ov}}(WT|ZV) + \pi^{ov}_{k_{ov}}(ZV|WT) + \pi^{ov}_{k_{ov}}(WT|VZ) + \pi^{ov}_{k_{ov}}(VZ|WT) + \pi^{ov}_{k_{ov}}(TW|ZV) + \pi^{ov}_{k_{ov}}(ZV|TW) + \pi^{ov}_{k_{ov}}(TW|VZ) + \pi^{ov}_{k_{ov}}(VZ|TW) + \pi^{ov}_{k_{ov}}(WV|ZT) + \pi^{ov}_{k_{ov}}(ZT|WV) + \pi^{ov}_{k_{ov}}(WV|TZ) + \pi^{ov}_{k_{ov}}(TZ|WV) + \pi^{ov}_{k_{ov}}(VW|ZT) + \pi^{ov}_{k_{ov}}(ZT|VW) + \pi^{ov}_{k_{ov}}(VW|TZ) + \pi^{ov}_{k_{ov}}(TZ|VW) \Big] dV_1\, dV_2 \tag{3.332}$$

Em relação à decomposição da energia molecular, a filosofia do presente método apresenta maiores afinidades com a adoptada pelo método de Clementi [100]. Consequentemente, é utilizado um critério mais rigoroso para lidar com a arbitrariedade

inerente à partição espacial da densidade de sobreposição e das energias monoelectrónicas, relativamente ao adoptado pelos métodos de Ruedenberg [99], Kollmar [101] e Bofill [102]. No entanto, o presente método é muito mais geral que o de Clementi e apresenta as seguintes vantagens adicionais:

(i) A partição é aplicada, não só à energia molecular, mas também à energia de estabilização do sistema.

(ii) Os fragmentos, em que o agregado molecular é subdividido, são entidades absolutamente gerais (átomos, grupos de átomos, iões, monómeros, etc.).

(iii) O modelo é definido de uma forma flexível tendo em conta um conjunto de efeitos importantes tais como: polarização, interacções coulômbicas, sobreposição, transferência de carga electrónica, relaxação da geometria dos fragmentos e transferência de grupos covalentemente ligados.

(iv) O formalismo é expresso explicitamente em termos da decomposição espacial das matrizes densidade e densidade de par. Isto permite uma interpretação física da interacção entre os fragmentos tendo em conta a natureza das ligações químicas que se estabelecem entre eles.

(v) As componentes da densidade, densidade de par, energia e energia de estabilização são calculadas usando unicamente os integrais indistinguíveis. Isto reduz significativamente o tempo de computação e o espaço em disco necessários para a implementação do algoritmo de decomposição.

(vi) A metodologia de decomposição é válida dentro do formalismo *ab initio* monoconfiguracional e multiconfiguracional.

(vii) No estudo de agregados moleculares onde ocorra uma reacção de transferência de um grupo covalentemente ligado, o presente algoritmo permite a decomposição da energia de estabilização em componentes intra- e inter-fragmento. A primeira componente está associada com a diferença de afinidade intrínseca, para o grupo transferido, entre os fragmentos envolvidos. A segunda componente está associada com a diferença de magnitude da interacção entre os fragmentos nos complexos ligantes inicial e final. Esta metodologia de interpretação permite uma melhor compreensão dos principais efeitos responsáveis pela estabilização preferencial dos estados estacionários relevantes das respectivas superfícies de energia potencial.

As principais desvantagens deste método são:

(1) O formalismo de partição não inclui o tratamento explícito das interacções com o meio, especialmente se ele for tratado por um modelo contínuo.

(2) A *correcção de counterpoise* não é introduzida explicitamente no modelo.

Referências

[1] R. Daudel, G. Leroy, D. Peeters e M. Sana, *Quantum Chemistry*, John Wiley & Sons, New York (1983).

[2] W. Kolos e L. Wolniewicz, *J. Chem. Phys.*, **41**, 3674 (1964).

[3] M. Born e J. R. Oppenheimer, *Ann. Phys.*, **87**, 457 (1927).

[4] J. von Newman, *Nachr. Acad. wiss. Gottingen, Math. Phisik. Kl. IIa. Mat. Physic. Chem. Abt.*, 245 (1927).

[5] P. A. M. Dirac, *Proc. Cambridge Phil. Soc.*, **25**, 62 (1929).

[6] P. A. M. Dirac, *Proc. Cambridge Phil. Soc.*, **26**, 376 (1930).

[7] P. A. M. Dirac, *Proc. Cambridge Phil. Soc.*, **27**, 240 (1931).

[8] K. Husimi, *Proc. Phys. Math. Soc. Japan*, **22**, 264 (1940).

[9] P. O. Löwdin, *J. Chem. Phys.*, **18**, 365 (1950).

[10] P. O. Löwdin, *Phys. Rev.*, **97**, 1474 (1955).

[11] R. McWeeny, *Proc. Roy. Soc (London)*, **A232**, 114 (1955).

[12] R. McWeeny, *Proc. Roy. Soc (London)*, **A235**, 496 (1956).

[13] R. McWeeny, *Rev. Mod. Phys.*, **32**, 335 (1960).

[14] D. R. Bates, K. Ledsham e A. L. Stewart, *Phil. Trans. Roy. Soc. London*, **A246**, 215 (1953).

[15] H. Wind, *J. Chem. Phys.*, **42**, 2371 (1965).

[16] C. C. J. Roothaan, *Rev. Mod. Phys.*, **23**, 69 (1951).

[17] R. J. Barthlett e J.F. Stanton, em *Reviews in Computational Chemistry*, Vol. 5, Ed. K. B. Lipkowitz e D. B. Boyd, VCH Publishers Inc., New York (1994).

[18] R. Shepard, *Adv. Chem. Phys.*, **69**, 63 (1987).

[19] I. Shavitt, em *Methods of Electronic Structure Theory*, Ed. H. F. Shaffer, Plenum, New York (1977).

[20] D. R. Hartree, *Proc. Camb. Philos. Soc.*, **24**, 89 (1928).

[21] V. Fock, *Z. Phys.*, **61**, 126 (1930).

[22] G. Náray-Szabó, P. R. Surján e J. G. Ángyán, *Applied Quantum Chemistry*, D. Reidel Publishing Company, Boston (1987).

[23] P. O. Löwdin, *Adv. Chem. Phys.*, **2**, 207 (1959).

[24] M. Møller e M. S. Plesset, *Phys. Rev.*, **46**, 618 (1934).

[25] R. J. Barthlett e D. M. Silver, *Int. J. Quantum Chem. Symp.*, **9**, 183 (1974).

[26] R. J. Barthlett e D. M. Silver, *Phys. Rev.*, **A10**, 1927 (1974).

[27] R. J. Barthlett, *J. Phys. Chem.*, **93**, 1697 (1989).

[28] L. Brillouin, *Act. Sci. Ind.*, **71**, (1933).

[29] C. M. Reeves e M. C. Harrison, *J. Chem. Phys.*, **39**, 11 (1963).

[30] E. Clementi e D. R. Davis, *J. Comput. Chem.*, **1**, 223 (1966)

[31] R. C. Raffenetti, *J. Chem. Phys.*, **58**, 4452 (1973).

[32] W. J. Hehre, L. Radom, P. v.R. Schleyer e J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, New York (1986).

[33] W. J. Hehre, R. F. Stewart e J. A. Pople, *J. Chem. Phys.*, **51**, 2657 (1969).

[34] T. H. Dunning e P. J. Hay, *Modern Theoretical Chemistry*, Plenum, New York (1976).

[35] R. Ditchfield, W. J. Hehre e J. A. Pople, *J. Chem. Phys.*, **54**, 724 (1971).

[36] W. J. Hehre, R. Ditchfield, e J. A. Pople, *J. Chem. Phys.*, **56**, 2257 (1972).

[37] R. Krishnan, J. S. Binkley, R. Seeger e J.A. Pople, *J. Chem. Phys.*, **72**, 650 (1980)

[38] R. Ahlrichs, *Chem. Phys. Lett.*, **15**, 609 (1972).

[39] P. C. Hariharen e J. A. Pople, *Theor. Chim. Acta (Berl.)*, **28**, 213 (1973).

[40] W. J. Hehre, R. Ditchfield e J. A. Pople, *J. Chem. Phys.*, **56**, 2257 (1972).

[41] T. Clark, J. Chandrasekhar, G. W. Spitznagel e P. v.R. Schleyer, *J. Comput. Chem.*, **4**, 294 (1983).

[42] M. J. Frisch, J. A. Pople e J. S. Binkley, *J. Chem. Phys.*, **80**, 3265 (1984).

[43] P. C. Hariharan e J. A. Pople, *Theoret. Chim. Acta (Berl.)*, **28**, 213 (1973).

[44] R. G. Parr, *J. Chem. Phys.*, **20**, 239 (1952).

[45] J. A. Pople e G. A. Segal, *J. Chem. Phys.*, **43**, S136 (1965).

[46] J. A. Pople e G. A. Segal, *J. Chem. Phys.*, **44**, 3289 (1966).

[47] J. A. Pople, D. L. Beveridge e A. P. Dobosh, *J. Chem. Phys.*, **47**, 2026 (1967).

[48] J. A. Pople, D. P. Santry e G. A. Segal, *J. Chem. Phys.*, **43**, S129 (1965).

[49] N. C. Baird, M. J. S. Dewar e R. Sustmann, *J. Chem. Phys.*, **50**, 1275 (1969).

[50] M. J. S. Dewar e E. Hassebach, *J. Am. Chem. Soc.*, **92**, 590 (1970).

[51] R. C. Bingham, M. J. S. Dewar e D. M. Lo, *J. Am. Chem. Soc.*, **97**, 1285 (1975).

[52] M. J. S. Dewar e G. Klopman, *J. Am. Chem. Soc.*, **89**, 3089 (1967).

[53] M. J. S. Dewar e W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977).

[54] M. J. S. Dewar, E. G. Zoebisch, E. F. Healey e S. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).

[55] J. J. P. Stewart, *J. Comp. Chem.*, **10**, 209 (1989).

[56] L. Oleari, L. DiSipio e G. DeMichells, *Mol. Phys.*, **10**, 97 (1966).

[57] A. van der Avoird, P.E.S. Wolmer, F. Mulder e R.M. Berns, *Top. Curr. Chem.*, **93**, 1 (1980).

[58] *Molecular Interactions*, Vol. 1-3, Ed. H. Ratajczak e W. J. Orville-Thomas, Wiley, Chichester (1982).

[59] *Hydrogen Bonds*, Vol. 120, Ed. P. Schuster, Springer, Berlin (1984).

[60] P. Hobza e R. Zahradník, *Intermolecular Complexes, the Role of Van der Waals Systems in Physical Chemistry and in the Biodisciplines*, Elsevier,Amsterdam, (1988).

[61] T.-H. Tang, W.-J. Hu, D.-Y. Yan e Y.-P. Cui, *J. Mol. Struct. (Theochem)*, **207**, 319 (1990).

[62] M. Kubicki e P. W. Codding, *Can. J. Chem.*, **71**, 68 (1993).

[63] M. Aida, *Bull. Chem. Soc. Japan*, **66**, 3423 (1993).

[64] P. Hobza, H. L. Selzle e E. W. Schlag, *J. Am. Chem. Soc.*, **116**, 3500 (1994).

[65] M. Bures e J. Berus, *Collect. Czech. Chem. Commun.*, **59**, 1251 (1994).

[66] A. Magalhães, B. Maigret, J. Hoflack, J. A. N. F. Gomes e H. A. Scheraga, *J. Protein Chem.*, **13**, 195 (1994).

[67] A. M. Sapse, S. Kabir e G. Snyder, *J. Mol. Struct. (Theochem)*, **339**, 227 (1995).

[68] B. Pirard, G. Baudoux e F. Durant, *Acta Cryst.*, **B51**, 103 (1995).

[69] T.-H. Tang e Y.-P. Cui, *Can. J. Chem.*, **74**, 1162 (1996).

[70] M.-J. Malliavin e C. Coudray, *J. Chem. Phys.*, **106**, 2323 (1997).

[71] M. Yu e M. Dolg, *Chem. Phys. Lett.*, **273**, 229 (1997).

[72] J. M. M. Cordeiro, *Int. J. Quantum Chem.*, **65**, 709 (1997).

[73] M. L. McKee e S. D. Worley, *J. Phys. Chem. A*, **101**, 5600 (1997).

[74] A. Karpfen e P. Schuster em *The Chemical Physics of Solvation. Part A: Theory of Solvation*, Ed. R. R. Dogonadze, E. Kalman, A. A. Kornyshev e J. Ulstrup, Tokyo (1985).

[75] S. L. Price em *Computer Simulations in Material Sciences : Interatomic Potentials; Simulations Techniques and Applications*, Ed. E. Meyer e V. Pontikis, Kluwer Academic Publ., London (1974).

[76] J. N. Murrell em *Orbital Theories of Molecules and Solids*, Ed. N. H. March, Oxford University Press, London (1974).

[77] P. Claverie em *Intermolecular Interactions from Diatomics to Biopolymers*, Ed. B. Pullman, John Wiley & Sons, Chichester (1978).

[78] J. N. Murrell, M. Randic, e D. R. Williams, *Proc. Roy. Soc., London, A*, **284**, 566 (1965).

[79] J. N. Murrell e G. Shaw, *J. Chem. Phys.*, **46**, 1768 (1967).

[80] R. Sustmann e G. Binsch, *Mol. Phys.*, **20**, 1 (1971).

[81] M. V. Basilevsky e M. M. Berenfeld, *Int. J. Quantum Chem.*, **6**, 23 (1972).

[82] J. G. C. M. van Duijneveldt e F. P. van Duijneveldt, *Chem. Phys. Lett.*, **17**, 425 (1972).

[83] D. M. Chipman, J. D. Bowman e J. O. Hirschfelder, *J. Chem. Phys.*, **59**, 2830 (1973).

[84] R. J. Barthlett e H. Weinstein, *Chem. Phys. Lett.*, **30**, 441 (1975).

[85] E. Kochanski e J. F. Gouyet, *Mol. Phys.*, **29**, 693 (1975).

[86] B. Jeziorsky e M. van Hemert, *Mol. Phys.*, **31**, 713 (1976).

[87] B. Jeziorsky, K. Szalewick e G. Chalesinski, *Int. J. Quantum Chem.*, **14**, 271 (1978).

[88] J. P. Daudey, P. Claverie e J. P. Malrieu, *Int. J. Quantum Chem.*, **8**, 1 (1974).

[89] I. C. Hayes e A. Stones, *Mol. Phys.*, **53**, 83 (1981).

[90] J. Tomasi em *Molecular Interactions*, Vol. 3, Ed. H. Ratajczak e W. J. Orville-Thomas, John Wiley & Sons, Chichester (1982).

[91] P. S. Epstein, *Phys. Rev.*, **28**, 695 (1926).

[92] R. K. Nesbet, *Proc. Roy. Soc.*, **A230**, 312 (1955).

[93] K. B. Kestner, *J. Chem. Phys.*, **48**, 252 (1968).

[94] S. Boys e F. Bernardi, *Mol. Phys.*, **19**, 553 (1970).

[95] F. Gianinetti, M. Raimondi e F. Tornaghi, *Int. J. Quantum Chem.*, **60**, 57 (1986).

[96] J. R. Collins e G. A. Galup., *Chem. Phys. Lett.*, **123**, 56 (1986).

[97] K. Morokuma, *J. Chem. Phys.*, **55**, 1236 (1971).

[98] K. Kitaura e K. Morokuma, *Int. J. Quantum Chem.*, **10**, 325 (1976).

[99] K. Ruedenberg, *Rev. Mod. Phys.*, **34**, 326 (1962).

[100] E. Clementi, *J. Chem. Phys.*, **46**, 3842 (1967).

[101] H. Kollmar, *Theor. Chim. Acta (Berl.)*, **50**, 235 (1978).

[102] J. M. Bofill, A. Povill e J. Rubio, *Chem. Phys. Lett.*, **222**, 51 (1994).

[103] A. Melo e M. J. Ramos, *Int. J. Quantum Chem.*, (1998).

[104] R. Cammi e J. Tomasi, *Theor. Chim. Acta (Berl.)*, **69**, 11 (1986).

[105] S. Nagase, T. Fueno, S. Yamabe e K. Kitaura, *Theor. Chim. Acta (Berl.)*, **49**, 309 (1978).

[106] R. Cammi, H. J. Hofmann e J. Tomasi, *Theor. Chim. Acta (Berl.)*, **76**, 297 (1989).

[107] R. S. Mulliken, *J. Chem. Phys.*, **23**, 1833 (1955).

[108] R. F. W. Bader e P. M. Peddall, *J. Chem. Phys.*, **56**, 3320 (1972).

[109] R. F. W. Bader, P. M. Peddall e J. Jr. Peslak, *J. Chem. Phys.*, **58**, 557 (1973).

[110] R. F. W. Bader, A. J. Duke e R. R. Messer, *J. Am. Chem. Soc.*, **95**, 7715 (1973).

[111] R. Daudel, R. F. W. Bader, M. E. Stephens e D. S. Borrett, *J. Can. Chem.*, **52**, 1310 (1974).

[112] M. J. S. Dewar e D. H. Lo, *J. Am. Chem. Soc.*, **93**, 7201 (1971).

[113] H. Fisher e H. Kollmar, *Theor. Chim. Acta (Berl.)*, **16**, 163 (1970).

[114] M. P. Allen e D. J. Tildesley, *Computer Simulations of Liquids*, Oxford University Press, Oxford (1990).

[115] F. Bopp, *Z. Physik*, **156**, 348 (1959).

[116] D. B.Cook, *Ab initio Valence Calculations in Chemistry*, Butterworths, London (1974).

[117] D. A. Brown, W. K. Glass, R. Mageswaran e B. Girmay, *Magn. Reson. Chem.*, **26**, 970 (1988).

[118] M. R. Rabinowitz, J. W. Sutherland, P. M. Patterson e R. B. Klemm, *J. Phys. Chem.*, **95**, 674 (1991).

[119] K. Héberger, M. Walbiner e H. Fisher, Angew. *Chem., Int. Ed. Engl.*, **31**, 635 (1992).

[120] N. Kondo, H. Fueno, H. Fujimoto, M. Makino, H. Nakaoka, I. Aaoki e S. Uemura, *J. Org. Chem.*, **59**, 5254 (1994).

[121] L. Shimoni, J. P. Glusker e C.W. Bock, *J. Phys. Chem.*, **100**, 2957 (1996).

[122] D. Fulle, H. F. Hamann, H. Hipler e J. Trol, *J. Chem. Phys.*, **105**, 983 (1996).

[123] S. A. Kandel, T. P. Rakitzis, T. Levon e R. N. Zare, *J. Chem. Phys.*, **105**, 7550 (1996).

[124] J. B. Griffin e P. B. Armentrout, *J. Chem. Phys.*, **107**, 5345 (1997).

[125] T. Richardz, J. Gale e G. Klose, *Chem. Phys. Lett.*, **271**, 79 (1997).

[126] M. B. More, D. Ray e P. B. Armentrout, *J. Phys. Chem. A*, **101**, 7007 (1997).

[127] R. Arnaud, *New. J. Chem.*, **15**, 615 (1991).

[128] S. Schröder, N. Buckley, N. J. Oppenheimer e P.A. Kollman, *J. Am. Chem. Soc.*, **114**, 8232 (1992).

[129] R. Arnaud, H. Postlethwaite e V. Barone, *J. Phys. Chem.*, **98**, 5913 (1994).

[130] I. Lee, C. K. Kim, D. S. Chung e B.- S. Lee, *J. Org. Chem.*, **59**,4490 (1994).

[131] B. S. Jursic e Z. Zdravkovski, *J. Org. Chem.*, **59**,3015 (1994).

[132] M. A. McAllister e T.T. Tidwell, *J. Org. Chem.*, **59**,4506 (1994).

[133] D. Buttar e D. M. Hirst, *J. Chem. Soc. Faraday Trans.*, **90**, 1811 (1994).

[134] K. D. Dobbs e D. A. Dixon, *J. Phys. Chem.*, **98**, 5290 (1994).

[135] J. Andrés, V. Moliner, J. Krechl e E. Silla, *J. Phys. Chem.*, **98**, 3664 (1994).

[136] M. Solà, J. L. Andrés, M. Duran, A. Lledós e J. Bertran, *Theor. Chim. Acta (Berl.)*, **91**, 333 (1995).

[137] M.F. Ruiz-López e D. Rinaldi, *J. Chem. Phys.*, **103**, 9249 (1995).

[138] C. Muguruma, N. Koga, K. Kitaura e K. Morokuma, *J. Chem. Phys.*, **103**, 9274 (1995).

[139] S. Sakai, *J. Phys. Chem.*, **99**, 5883 (1995).

[140] B. S. Jurcic, *Chem. Phys. Lett.*, **244**, 263 (1995).

[141] A. Bouyacoub, Y. Jean e F. Volatron, *J. Mol. Struct. (Theochem)*, **371**, 51 (1996).

[142] X. Assfeld, J. Garapon, D. Rinaldi, M. F. Ruiz-López e J.L. Rivail, *J. Mol. Struct. (Theochem)*, **371**, 107 (1996).

[143] I. Tuñón, F.R. Tortonda, J.L. Pascual-Ahuir, and E. Silla, *J. Mol. Struct. (Theochem)*, **371**, 117 (1996).

[144] R. Sumathi, B. Engels e S. D. Peyerimhoff, *J. Chem. Phys.*, **105**, 8117 (1996).

[145] Q. Cui, D. G. Musaev, M. Svenson e K. Morokuma, *J. Phys. Chem.*, **100**, 10936 (1996).

[146] W. T. Lee e R.I. Masel, *J. Phys. Chem.*, **100**, 10945 (1996).

[147] T. Peng, D.H. Zang, J.Z.H. Zang e R. Schinke, *Chem. Phys. Lett.*, **248**, 37 (1996).

[148] P. E. Sinclair e C. R. A. Catlow, *J. Chem Soc., Faraday Trans.*, **92**, 2099 (1996).

[149] W.H. Thompson e W.H. Miller, *J. Chem. Phys.*, **106**, 142 (1997).

[150] T. Takayanagi e G. C. Schatz, *J. Chem. Phys.*, **106**, 3227 (1997).

[151] J. C. Corchado e J. Espinosa-Garcia, *J. Chem. Phys.*, **106**, 4013 (1997).

[152] D. Sengupta e M. T. Mguyen, *J. Chem. Phys.*, **106**, 9703 (1997).

[153] R. Ramachandra, J. Semekowitsch e R. E. Wyatt, *Chem. Phys. Lett.*, **270**, 387 (1997).

[154] J. Espinosa-Garcia e J. C. Corchado, *J. Phys. Chem. A*, **101**, 7336 (1997).

[155] E. Jemmis, K. T. Giju e J. Leszczynski, *J. Phys. Chem. A*, **101**, 7389 (1997).

[156] J. P. Dijkman, R. Osman e M. Weinstein, *Int. J. Quantum Chem., Quantum Biol. Symp.*, **14**, 211 (1987).

[157] D. Hadzi, *J. Mol. Struct. (Theochem)*, **177**, 1 (1988).

[158] J. P. Dijkman, R. Osman e M. Weinstein, *Int. J. Quantum Chem.*, **35**, 241 (1989).

[159] S. Scheiner, *Acc. Chem. Res.*, **18**, 174 (1985).

[160] L. Pardo, A. P. Mazurek e R. Osman, *Int. J. Quantum Chem.*, **37**, 701 (1990).

[161] C. J: Cassaday, S. R. Carr, K. Zhang e A. Chung-Philiphs, *J. Org. Chem.*, **60**, 1704 (1995).

[162] C. Scharnagl, J. Hettenkofer e S. F. Fisher, *Int. J. Quantum Chem., Quantum Biol. Symp.*, **21**, 33 (1994).

[163] D.J. Barlow e J. M. Thornton, *J. Mol. Biol.*, **168**, 867 (1983).

[164] J. Singh, J. M. Thornton, M. Snarey e S. F. Campbell, *FEBS Lett.*, **224**, 161 (1987).

Capítulo 4: Resultados

Se eu vim ao mundo, foi
Só para desflorar florestas virgens,
E desenhar meus próprios pés na areia inexplorada.
O mais que eu faço não vale nada.

José Régio, extracto de *O Cântico Negro, Poemas de Deus e do Diabo*, em *Antologia Poética*, Círculo dos Leitores, Lisboa (1993).

4.1 - Introdução

O nosso estudo teórico de reacções de associação protease serina:inibidor desenvolveu-se segundo duas vertentes principais.

Numa primeira vertente, foi efectuado um estudo detalhado das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI, previamente apresentadas no capítulo 1, utilizando ciclos termodinâmicos apropriados e os métodos da dinâmica molecular, perturbação termodinâmica e integração termodinâmica descritos no capítulo 2. Os resultados obtidos são apresentados na secção 4.2. Nestas simulações, foram produzidos ficheiros de trajectórias de grandes dimensões cujo armazenamento em disco provocou grandes problemas. Por isso, os compressores conversativos clássicos (compress, gzip, pack e compact) bem como o algoritmo BS-VLC, descrito na secção 2.5, foram aplicados na compressão destes ficheiros. Resultados típicos obtidos nestas compressões são apresentados na secção 4.3.

Numa segunda vertente, foi efectuado um estudo quântico das interacções arginina-carboxilato (Arg-Carb), que tal como foi salientado na secção 1.4, desempenham um papel relevante no mecanismo de inibição do factor β-XII a. Neste estudo, foram utilizados métodos *semi-empíricos* (AM1) e *ab initio* (6-31G** e MP2/6-31G**) e os modelos moleculares apropriados metilguanidínio-acetato e metilguanidina-ácido acético. Com o objectivo de estabelecer um modelo interpretativo adequado, o método de decomposição de Melo e Ramos, elaborado no âmbito deste trabalho e descrito na secção 3.5, foi também aplicado no estudo destas interacções. Os resultados obtidos são apresentados na secção 4.4.

4.2 - Estudo das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI

As interacções ligantes que estabilizam o complexo tripsina:PTI foram previamente apresentadas na figura (1.10) da secção 1.3. Estas interacções foram classificadas em interacções que envolvem somente a cadeia principal (apresentadas na figura 1.11) e interacções envolvendo cadeias laterais (apresentadas na figura 1.12). Esta última classe pode

ainda ser subdividida em interacções envolvendo somente cadeias laterais e interacções envolvendo uma cadeia lateral e a cadeia principal.

Como foi salientado na secção 1.3, não é possível determinar directamente a contribuição de cada uma destas interacções para a estabilidade do complexo tripsina:PTI. Nestas circunstâncias, optou-se por avaliar indirectamente estas contribuições através da mutação de um conjunto de resíduos específicos na região de contacto do complexo tripsina:PTI, originando uma tripsina (MTripsina) ou PTI (MPTI) modificados.

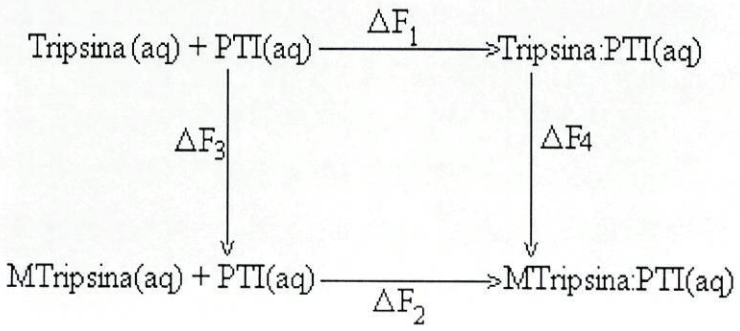
As mutações foram efectuadas de forma a substituir um resíduo, envolvido numa determinada interacção, por outro resíduo que destruísse a respectiva especificidade. As mutações efectuadas podem ser classificadas em duas classes: mutações físicas e mutações não físicas.

As mutações físicas envolvem a substituição do resíduo original por outro resíduo, o mais semelhante possível, do ponto de vista topológico, mas com características químicas opostas. Pretende-se assim anular a interacção em estudo sem provocar alterações estruturais significativas no sistema. Os resultados teóricos obtidos para estas mutações podem ser directamente comparados com resultados experimentais.

No entanto, não é possível encontrar mutações físicas adequadas ao estudo de todas as interacções ligantes do complexo tripsina:PTI. Por exemplo, as interacções envolvendo somente a cadeia principal não são afectadas por mutações deste tipo. Nestes casos, optou-se por efectuar mutações não físicas, onde o átomo responsável pela interacção é substituído por um pseudo-átomo preservando as suas constantes de força internas mas anulando os seus parâmetros electrostáticos e de van der Waals. Os resultados teóricos obtidos para estas mutações, embora também permitam avaliar a contribuição das respectivas interacções para a estabilidade do complexo tripsina:PTI, não podem ser directamente comparados com resultados experimentais. Por outro lado, no caso da interacção His^{40}(tripsina)-Arg^{17}(PTI),não existe nenhum aminoácido topologicamente similar à arginina que seja apropriado para a mutação pretendida. Neste caso, optou-se pela mutação da arginina, que apresenta um ião guanidínio com grupo terminal da cadeia lateral, pela arginina neutra em que o ião guanidínio é substituído por um grupo neutro guanidina. Com esta mutação obtém-se o efeito pretendido, destruição da especificidade da interacção His^{40}(tripsina)-Arg^{17}(PTI), mas também não são possíveis comparações directas com resultados experimentais.

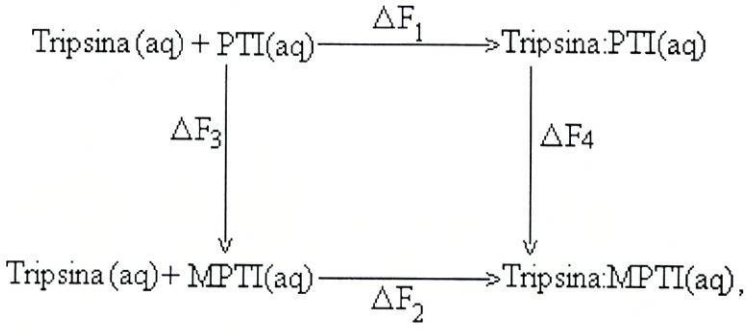
A interacção Gln^{192}(tripsina)-Cys^{14}(PTI) foi estudada através de três mutações (duas físicas e uma não física), com o objectivo de comparar as diferentes metodologias utilizadas no presente trabalho.

As interacções ligantes, responsáveis pela estabilidade do complexo tripsina:PTI, bem como as mutações utilizadas no seu estudo são apresentadas nas tabelas 4.1 e 4.2. A contribuição de cada uma destas interacções para a estabilidade do complexo foi avaliada recorrendo a ciclos termodinâmicos similares aos apresentados nas equações (2.78) e (2.79),



(4.1)

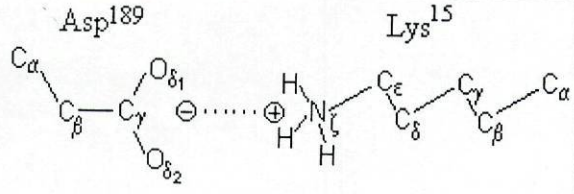
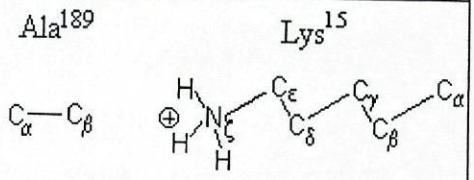
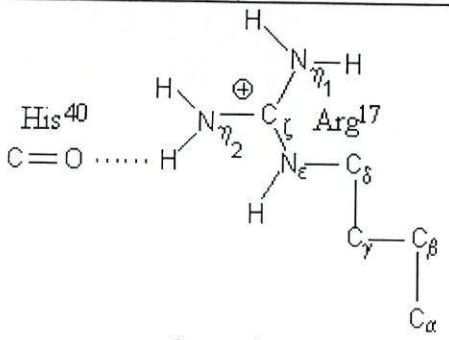
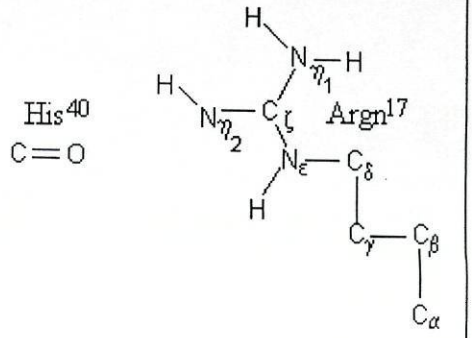
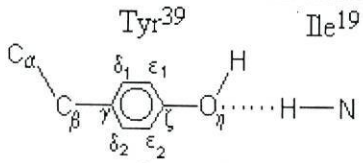
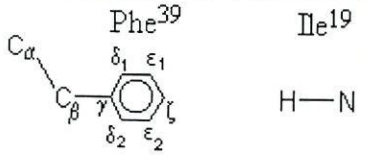
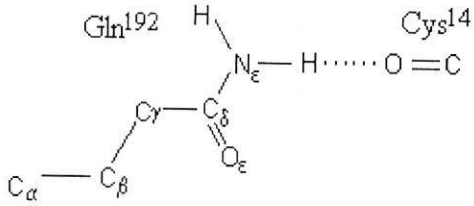
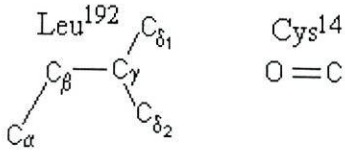
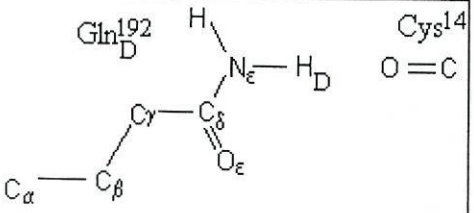
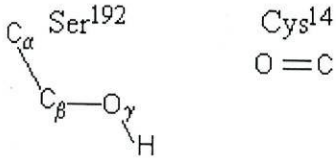
ou



(4.2)

assim como à aproximação do parâmetro de acoplamento e aos métodos da perturbação termodinâmica, da integração termodinâmica e da dinâmica molecular. Neste contexto, a variação da energia de Helmholtz ($\Delta\Delta A$), associada a cada mutação, pode ser determinada através da simulação dos processos não físicos 3 e 4 (nos ciclos 4.1 ou 4.2) e recorrendo à equação (2.82). Metodologias análogas têm sido também bastante utilizadas no estudo teórico de outros sistemas com interesse biológico [1-30].

Tabela 4.1. Interacções ligantes, envolvendo as cadeias laterais, na região de contacto tripsina:PTI e mutações utilizadas para avaliar a respectiva contribuição para a estabilidade do complexo.

Interacção original[a]	Mutação	Interacção final[a]
 (*sc*)	$Asp^{189} \rightarrow Ala$ **(Fis.)**	
 (*sc-mc*)	$Arg^{17} \rightarrow Argn^{b}$ **(N/ Fis.)**	
 (*sc-mc*)	$Tyr^{39} \rightarrow Phe$ **(Fis.)**	
 (*sc-mc*)	$Gln^{192} \rightarrow Leu$ **(Fis.)**	
	$Gln^{192} \rightarrow Gln_D{}^{c}$ **(N/ Fis.)**	
	$Gln^{192} \rightarrow Ser$ **(Fis.)**	

sc- interacções envolvendo cadeias laterais ; *sc-mc*- interacções envolvendo cadeias laterais e a cadeia principal.
Fis.- Mutação física ; **N/ Fis.**- Mutação não física.

[a] O primeiro resíduo pertence à tripsina e o segundo resíduo ao PTI.

[b] Arginina neutra.

[c] O átomo responsável pela interacção é mutado por um pseudo-átomo mantendo as suas constantes de força ligantes mas anulando os seus parâmetros electrostáticos de van der Waals.

Tabela 4.2. Interacções ligantes, envolvendo somente a cadeia principal, na região de contacto tripsina:PTI e mutações utilizadas para avaliar a respectiva contribuição para a estabilidade do complexo.

Interacção original[a]	Mutação	Interacção final[a]
Ser^{214} Lys^{15} C=O-----H—N (*mc*)	$Lys^{15} \rightarrow Lys_{D1}$[b] **(N/ Fis.)**	Ser^{214} Lys^{15}_{D} C=O H_{D}—N
Phe^{41} Arg^{17} C=O-----H—N (*mc*)	$Arg^{17} \rightarrow Arg_{D}$[b] **(N/ Fis.)**	Phe^{41} Arg^{17}_{D} C=O H_{D}—N
Gly^{193} N—H Lys^{15} Ser^{195} O=C N—H (*mc*)	$Lys^{15} \rightarrow Lys_{D2}$[b] **(N/ Fis.)**	Gly^{193} N—H Lys^{15}_{D} Ser^{195} O_{D}=C N—H
Gly^{216} Pro^{13} N—H-----O=C (*mc*)	$Pro^{13} \rightarrow Pro_{D}$[b] **(N/ Fis.)**	Gly^{216} Pro^{13}_{D} N—H O_{D}=C

mc- interacções envolvendo somente a cadeia principal.
N/ Fis.- Mutação não física.
[a] O primeiro resíduo pertence à tripsina e o segundo resíduo ao PTI.
[b] O átomo responsável pela interacção é mutado por um pseudo-átomo mantendo as suas constantes de força ligantes mas anulando os seus parâmetros electrostáticos de van der Waals.

As estruturas cristalográficas da tripsina, PTI e complexo tripsina:PTI incluem algumas águas de cristalização. No entanto, para obter uma solvatação mais realista destas estruturas, foram introduzidas nos modelos utilizados esferas de água TIP3P [31] de 18 Å de raio centradas na região mutada. Os sistemas modelo utilizados na simulação de cada mutação são apresentados na tabela 4.3.

No estudo de cada mutação, as estruturas modelo iniciais foram optimizadas, até que os desvios quadráticos médios dos gradientes da energia fossem inferiores a 0,0001 kcal/mol, usando 1000 passos do método *Steepest Descents* e os restantes passos do método *do gradiente conjugado*.

Tabela 4.3. Sistemas modelo utilizados na simulação das mutações de aminoácidos específicos região de contacto tripsina:PTI.

Interacção[a]	Mutação	Centro da esfera de 18 Å de águas TIP3P	Modelo da tripsina solvatada	Modelo do PTI solvatado	Modelo do complexo tripsina:PTI solvatado
Asp^{189}-Lys^{15} *(sc)*	$Asp^{189}\rightarrow Ala$ **(Fis.)**	$C_\beta(Asp^{189})$	223 aminoácidos + 393 águas TIP3P	Não utilizado	281 aminoácid + 417 águas TIP3
His^{40}-Arg^{17} *(sc-mc)*	$Arg^{17}\rightarrow Argn$[b] **(N/ Fis.)**	$C_\zeta(Arg^{17})$	Não utilizado	58 aminoácidos + 671 águas TIP3P	281 aminoácid + 434 águas TIP3
Tyr^{39}-Ile^{19} *(sc-mc)*	$Tyr^{39}\rightarrow Phe$ **(Fis.)**	$C_\zeta(Tyr^{39})$	223 aminoácidos + 601 águas TIP3P	Não utilizado	281 aminoácid + 465 águas TIP3
Gln^{192}-Cys^{14} *(sc-mc)*	$Gln^{192}\rightarrow Leu$ **(Fis.)**	$C_\gamma(Gln^{192})$	223 aminoácidos + 458 águas TIP3P	Não utilizado	281 aminoácid + 354 águas TIP3
	$Gln^{192}\rightarrow Gln_D$[c] **(N/ Fis.)**	$C_\gamma(Gln^{192})$	223 aminoácidos + 458 águas TIP3P	Não utilizado	281 aminoácid + 354 águas TIP3
	$Gln^{192}\rightarrow Ser$ **(Fis.)**	$C_\gamma(Gln^{192})$	223 aminoácidos + 458 águas TIP3P	Não utilizado	281 aminoácid + 354 águas TIP3
Ser^{214}-Lys^{15} *(mc)*	$Lys^{15}\rightarrow Lys_{D1}$[c] **(N/ Fis.)**	$N(Lys^{15})$	Não utilizado	58 aminoácidos + 678 águas TIP3P	281 aminoácid + 281 águas TIP3
Phe^{41}-Arg^{17} *(mc)*	$Arg^{17}\rightarrow Arg_D$[c] **(N/ Fis.)**	$N(Arg^{17})$	Não utilizado	58 aminoácidos + 645 águas TIP3P	281 aminoácid + 321 águas TIP3
Ser^{195}-Lys^{15} Gly^{193}-Lys^{15} *(mc)*	$Lys^{15}\rightarrow Lys_{D2}$[c] **(N/ Fis.)**	$C(Lys^{15})$	Não utilizado	58 aminoácidos + 683 águas TIP3P	281 aminoácid + 265 águas TIP3
Gly^{216}-Pro^{13} *(mc)*	$Pro^{13}\rightarrow Pro_D$[c] **(N/ Fis.)**	$C(Pro^{13})$	Não utilizado	58 aminoácidos + 673 águas TIP3P	281 aminoácid + 367 águas TIP3

sc- interacções envolvendo cadeias laterais ; *mc*- interacções envolvendo somente a cadeia principal;
sc-mc- interacções envolvendo cadeias laterais e a cadeia principal.
Fis.- Mutação física ; **N/ Fis.**- Mutação não física.
[a] O primeiro resíduo pertence à tripsina e o segundo resíduo ao PTI.
[b] Arginina neutra.
[c] O átomo responsável pela interacção é mutado por um pseudo-átomo mantendo as suas constantes de força ligan mas anulando os seus parâmetros electrostáticos de van der Waals.

Nas simulações dos processos 3 e 4 na mutação $Tyr^{39} \rightarrow Phe$ [32], foram utilizadas 29 janelas (0,0025, de 0,01 a 0,05 com incrementos de 0,01, de 0,1 a 0,9 com incrementos de 0,05, de 0,95 a 0,99 com incrementos de 0,05 e 0,9975). Nas simulações dos mesmos processos nas restantes mutações [33], foram utilizadas 27 janelas (de 0,01 a 0,05 com incrementos de 0,01, de 0,1 a 0,9 com incrementos de 0,05 e de 0,95 a 0,99 com incrementos de 0,05). Para cada janela, o sistema modelo foi equilibrado a 300 K usando dinâmica molecular com um intervalo de tempo de integração de 0,001 ps (os tempos de equilibração variaram entre 60 ps e 1,2 ns) e o algoritmo SHAKE foi aplicado a todas as ligações envolvendo átomos de hidrogénio. A estrutura equilibrada numa janela foi utilizada para iniciar a fase de equilibração da janela seguinte. O modelo da *região activa* foi utilizado em todas as simulações. Nestas circunstâncias, todas as moléculas de água e aminoácidos contidos na região activa, esfera de 15 Å centrada num átomo da região mutada (ver tabela 4.3), eram livres de se moverem nas simulações MD. Foram impostas forças harmónicas a todas as moléculas de água contidas na região fronteira (calote esférica de raio 15 Å e espessura 3 Å centrada no mesmo átomo da região mutada); isto foi implementado aplicando uma constante de força de 0,6 kcal mol^{-1} $Å^{-2}$ aos deslocamentos dos átomos de oxigénio das moléculas de água relativamente às suas posições iniciais. Os restantes resíduos foram incluídos na região inerte, pelo que os respectivos átomos foram fixados nas suas posições iniciais mas foram utilizados na determinação da energia potencial e das forças. Um corte esférico de 9 Å foi utilizado na determinação das interacções não ligantes, cuja actualização foi efectuada com uma frequência de 10 passos (0,01 ps). A temperatura foi mantida constante usando o método de Berendsen com uma constante de acoplamento (τ_T) de 0,1 ps. As trajectórias foram guardadas para análise com uma frequência de 20 passos (0,02 ps).

Nos cálculos de variações da energia de Helmholtz, foram utilizados os métodos da perturbação termodinâmica (TP) e da integração termodinâmica (TI). A metodologia *double-wide sampling* foi utilizada nos cálculos TP. Nos cálculos TI, as variações da energia de Helmholtz conformacional, ΔF_3 e ΔF_4, foram obtidas através de integração trapezoidal. Nestes cálculos, foi também utilizado o método de extrapolação de Brooks III (baseado na equação 2.86) a partir dos estados híbridos de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.Para cada janela λ_i, a convergência das propriedades calculadas, gradiente da energia potencial ($\partial V_\lambda / \partial \lambda$) e variações da energia de Helmholtz conformacional ($\Delta F_{\lambda_i \rightarrow \lambda_{i-1}}$ e $\Delta F_{\lambda_i \rightarrow \lambda_{i+1}}$), foi monitorizada em

função do tempo de simulação. Resultados típicos, obtidos no estudo do processo 3 (Tripsina→MTripsina) da mutação Tyr^{39}→Phe [32], são representados nas figuras (4.1) a (4.3).

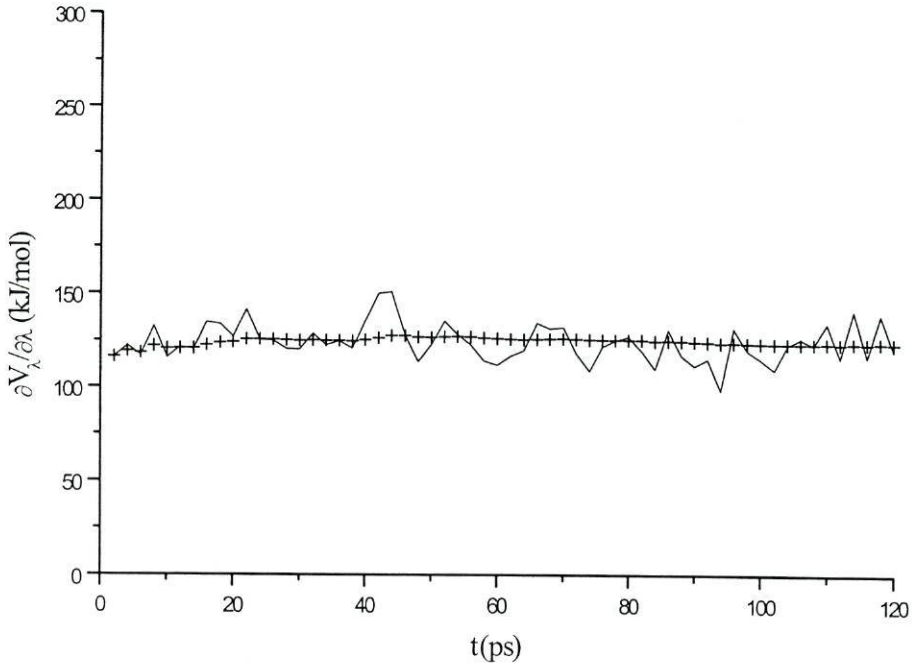


Figura 4.1. Convergência de $\partial V_\lambda / \partial \lambda$ para o estado híbrido λ=0,1 na simulação do processo 3 (Tripsina→MTripsina) da mutação Asp^{189}→Ala.

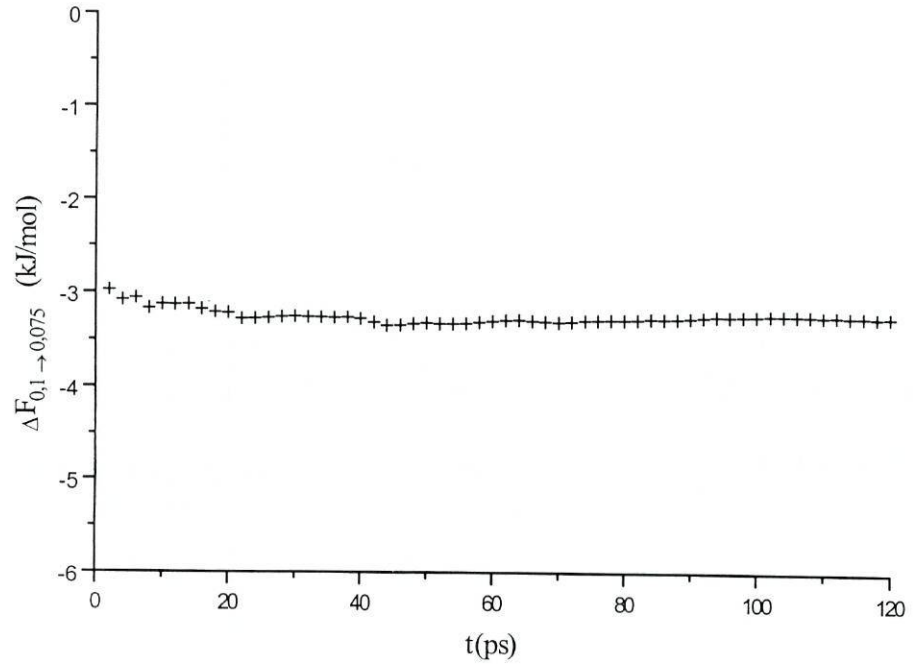


Figura 4.2. Convergência de $\Delta F_{0,1\rightarrow 0,075}$ para o estado híbrido λ=0,1 na simulação do processo 3 (Tripsina→MTripsina) da mutação Asp^{189}→Ala.

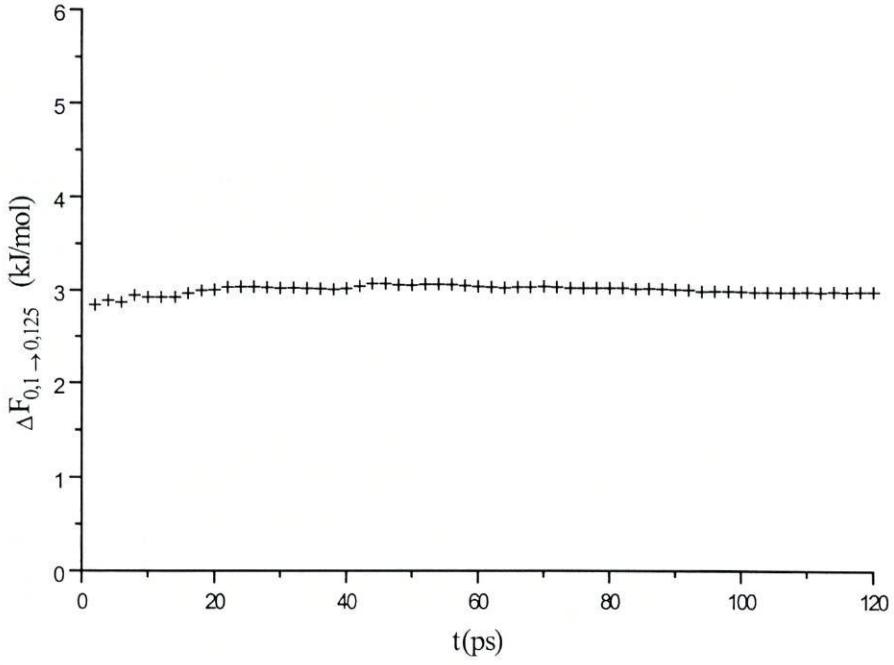


Figura 4.3. Convergência de $\Delta F_{0,1\rightarrow 0,125}$ para o estado híbrido λ=0,1 na simulação do processo 3 (Tripsina→MTripsina) da mutação Asp^{189}→Ala.

Uma vez atingido o equilíbrio estatístico, iniciava-se a fase de produção que era continuada até se obter a convergência das propriedades em estudo. Em cada janela, foi utilizado um mínimo de 120 ps de simulação MD. Todos os cálculos foram efectuados com o programa CHARMm©22 [34-36] e utilizando o respectivo potencial *united atoms*. Os tempos de simulação, tempos de CPU e computadores usados nestes cálculos são apresentados na tabela 4.4. Estes tempos são consideravelmente maiores que os utilizados na maioria de cálculos similares publicados na literatura [1-30].

Como foi salientado na secção 2.3.2, a convergência das médias de *ensemble* apresenta problemas específicos nas zonas onde as propriedades associadas mudam de sinal e nos *pontos extremos*. Nestas zonas, os tempos de simulação foram aumentados para assegurar uma boa convergência. Nos *pontos extremos*, foi também diminuído o espaçamento entre as janelas tendo em conta as grandes variações que as propriedades termodinâmicas aí apresentam.

Tabela 4.4. Dados referentes às simulações computacionais das mutações estudadas na região d contacto tripsina:PTI.

Interacção[a]	Mutação	Tempos de simulação (ns)			Tempos de CPU (dias)			Computado
		Processo 3	Processo 4	Total	Processo 3	Processo 4	Total	
Asp^{189}-Lys^{15} *(sc)*	Asp^{189}→Ala **(Fis.)**	4,560	5,760	10,320	43	56	99	HP/700
His^{40}-Arg^{17} *(sc-mc)*	Arg^{17}→Argn[b] **(N/ Fis.)**	4,560	4,080	8,640	14	13	27	Alpha/500
Tyr^{39}-Ile^{19} *(sc-mc)*	Tyr^{39}→Phe **(Fis.)**	6,240	6,400	12,640	60	62	122	HP/700
Gln^{192}-Cys^{14} *(sc-mc)*	Gln^{192}→Leu **(Fis.)**	5,520	5,880	11,400	53	56	109	HP/700
	Gln^{192}→Gln_D[c] **(N/ Fis.)**	5,760	4,440	10,200	55	43	98	HP/700
	Gln^{192}→Ser **(Fis.)**	5,400	4,920	10,320	52	47	99	HP/700
Ser^{214}-Lys^{15} *(mc)*	Lys^{15}→Lys_{D1}[c] **(N/ Fis.)**	3,240	3,360	6,600	31	32	63	HP/700
Phe^{41}-Arg^{17} *(mc)*	Arg^{17}→Arg_D[c] **(N/ Fis.)**	3,240	3,240	6,480	31	31	62	HP/700
Ser^{195}-Lys^{15} Gly^{193}-Lys^{15} *(mc)*	Lys^{15}→Lys_{D2}[c] **(N/ Fis.)**	4,920	3,480	8,400	16	11	27	Alpha/500
Gly^{216}-Pro^{13} *(mc)*	Pro^{13}→Pro_D[c] **(N/ Fis.)**	4,080	3,840	7,920	13	12	25	Alpha/500
Total		47,520	45,400	92,920	368	363	731	HP/700 e Alpha/500

sc- interacções envolvendo cadeias laterais ; *mc*- interacções envolvendo somente a cadeia principal;
sc-mc- interacções envolvendo cadeias laterais e a cadeia principal.
Fis.- Mutação física ; **N/ Fis.**- Mutação não física.
[a] O primeiro resíduo pertence à tripsina e o segundo resíduo ao PTI.
[b] Arginina neutra.
[c] O átomo responsável pela interacção é mutado por um pseudo-átomo mantendo as suas constantes de força ligante mas anulando os seus parâmetros electrostáticos de van der Waals.

O método dos blocos, descrito na secção 2.3.1, foi utilizado para calcular os erros estatísticos associados às médias de *ensemble* em estudo. Com este objectivo foi determinada a função de correlação $\rho(\nu)$ (ver equação 2.46) associada a cada uma destas quantidades. Utilizando as equações (2.47) e (2.48), calculou-se, em seguida, para cada caso, o respectivo comprimento de correlação (n_K^{min}). Exemplos típicos desta metodologia, obtidos na simulação do processo 3 (Tripsina→MTripsina) da mutação Tyr^{39}→Phe [32], são apresentados nas figuras (4.4) a (4.6). Os comprimentos de correlação assim obtidos apresentaram valores sempre inferiores a 100 conformações (2 ps). Neste contexto, utilizámos este valor para o comprimento de bloco nos cálculos dos erros estatísticos associados às diferentes médias de *ensemble*, o que constitui uma garantia de que as respectivas propriedades se encontravam descorrelacionadas e que a equação (2.45) era aplicável.

Os erros totais associados às variações da energia de Helmholtz conformacional (ΔF_3 e ΔF_4) foram calculados utilizando as equações (2.62) e (2.68) nas simulações TP e as equações (2.75) e (2.76) nas simulações TI.

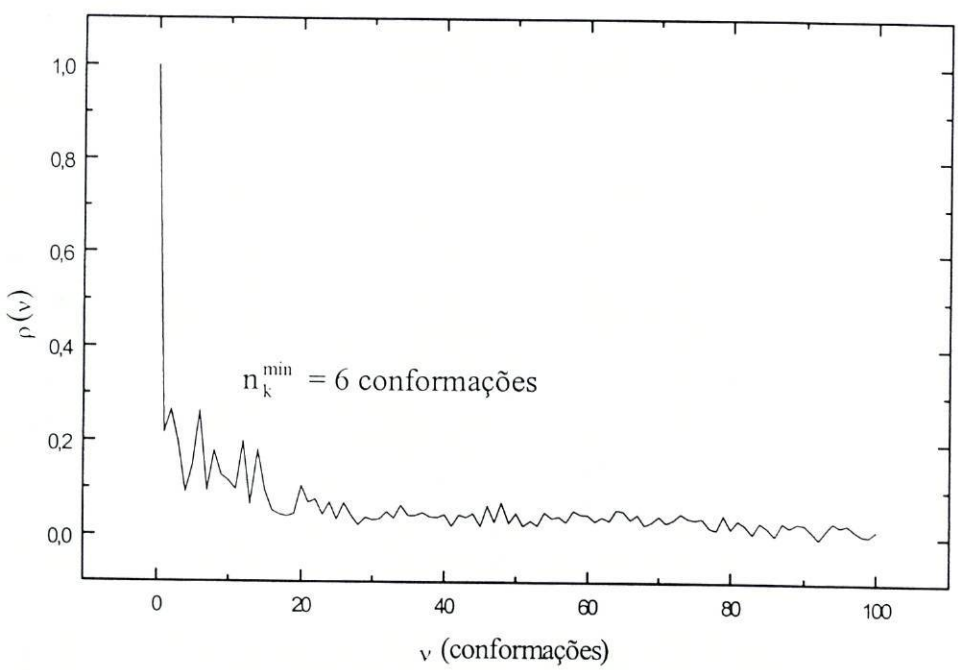


Figura 4.4. Função de correlação ($\rho(\nu)$) de $\partial V_\lambda/\partial\lambda$ para o estado híbrido λ=0,1 na simulação do processo 3 (Tripsina→MTripsina) da mutação Tyr^{39}→Phe.

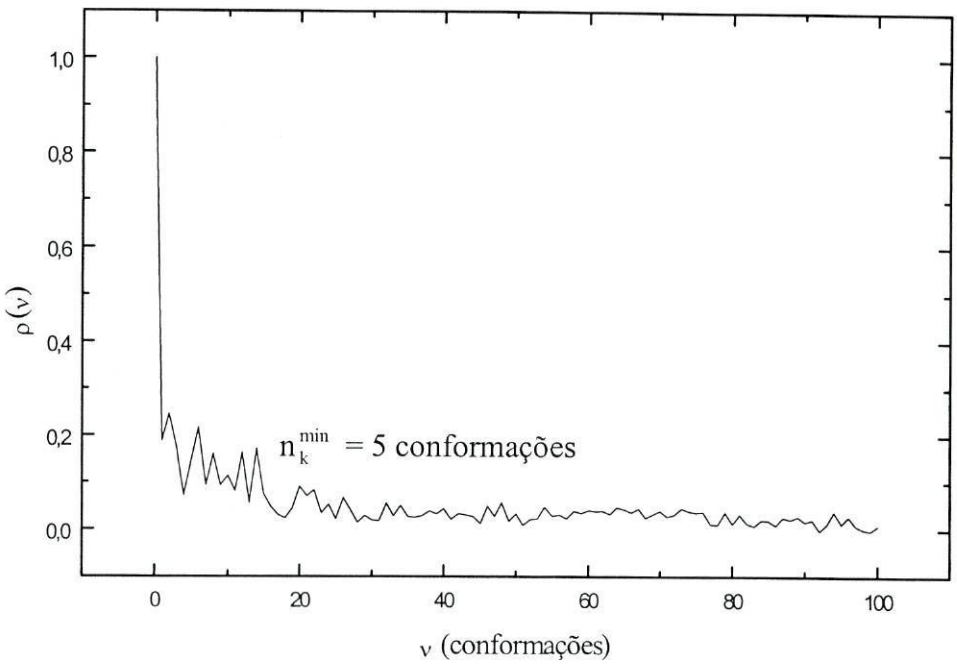


Figura 4.5. Função de correlação ($\rho(\nu)$) de $\Delta F_{0,1\rightarrow 0,075}$ para o estado híbrido λ=0,1 na simulação do processo 3 (Tripsina→MTripsina) da mutação Tyr^{39}→Phe.

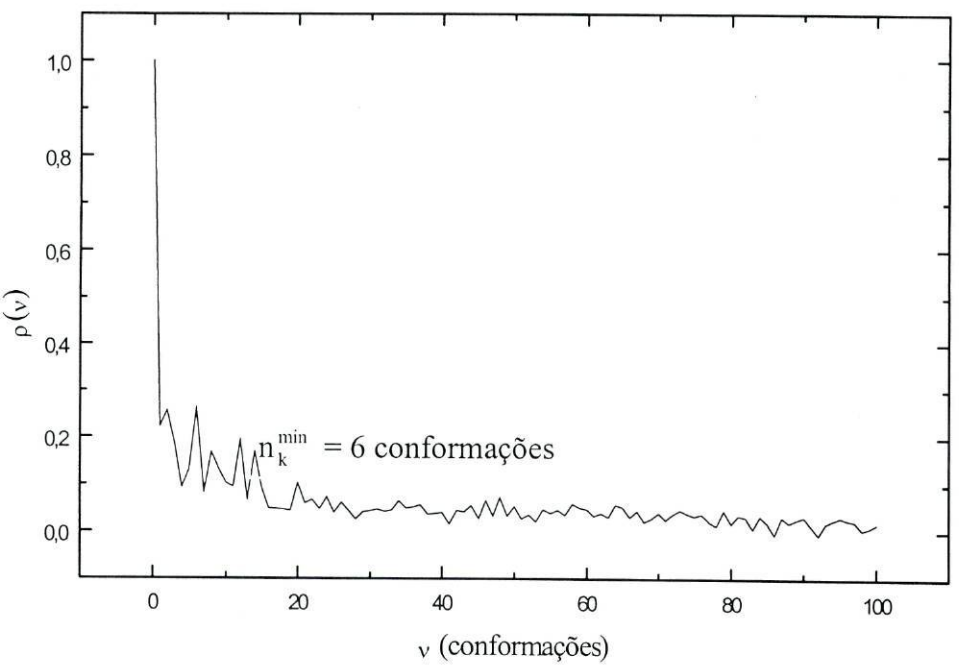


Figura 4.6. Função de correlação ($\rho(\nu)$) de $\Delta F_{0,1\rightarrow 0,125}$ para o estado híbrido λ=0,1 na simulação do processo 3 (Tripsina→MTripsina) da mutação Tyr^{39}→Phe.

Os resultados finais obtidos nestas simulações são apresentados na tabela 4.5 e no apêndice I.

.**Tabela 4.5.** Variações da energia de Helmholtz (kJ/mol) calculadas através de simulações moleculares.

Interacção[a]	Mutação	Variações da energia de Helmholtz	TP[b]	TI[c]	TI/TP[d]	TI(ext.)[e]
Asp^{189}-Lys^{15} (*sc*,**vs**)	Asp^{189}→Ala **(Fis.)**	ΔF_4	479,9±0,8	482,8±0,8	478,6±0,8	501±7
		ΔF_3	357,7±0,8	364,4±0,4	356,5±0,8	374±5
		ΔΔA≈ΔΔF	122±1	118,4±0,9	122±1	127±9
His^{40}-Arg^{17} (*sc-mc*,**s**)	Arg^{17}→$Argn^{f}$ **(N/ Fis.)**	ΔF_4	152,0±0,3	151,9±0,3	151,7±0,3	150±2
		ΔF_3	107,5±0,4	108,2±0,3	107,2±0,4	107±2
		ΔΔA≈ΔΔF	44,5±0,5	43,7±0,4	44,5±0,5	43±3
Tyr^{39}-Ile^{19} (*sc-mc*,**m**)	Tyr^{39}→Phe **(Fis.)**	ΔF_4	41,4±0,8	47,7±0,4	39±1	60±2
		ΔF_3	24±1	29,7±0,4	21±1	36±5
		ΔΔA≈ΔΔF	17±1	18,0±0,6	18±1	24±5
Gln^{192}-Cys^{14} (*sc-mc*,**w**)	Gln^{192}→Leu **(Fis.)**	ΔF_4	165±1	165,3±0,8	164±1	$1,6x10^2$±$1x10^2$
		ΔF_3	215±1	217,6±0,4	214±1	223±7
		ΔΔA≈ΔΔF	-50±1	-52,3±0,9	-50±1	$-6x10^2$±$1x10^2$
	Gln^{192}→Gln_D[g] **(N/ Fis.)**	ΔF_4	-1,63±0,08	-1,59±0,08	-1,63±0,08	-2,97±0,08
		ΔF_3	22,80±0,08	22,80±0,08	22,80±0,08	21,6±0,2
		ΔΔA≈ΔΔF	-24,4±0,1	-24,4±0,1	-24,4±0,1	-24,6±0,2
	Gln^{192}→Ser **(Fis.)**	ΔF_4	68±2	83,3±0,4	67±2	93±4
		ΔF_3	81±2	101,3±0,4	80±2	112±3
		ΔΔA≈ΔΔF	-13±3	-18,0±0,6	-13±3	-19±5
Ser^{214}-Lys^{15} (*mc*,**w**)	Lys^{15}→Lys_{D1}[g] **(N/ Fis.)**	ΔF_4	58,37±0,04	57,15±0,04	58,37±0,04	57,07±0,08
		ΔF_3	59,71±0,08	58,49±0,08	59,71±0,08	59,2±0,2
		ΔΔA≈ΔΔF	-1,34±0,09	-1,34±0,09	-1,34±0,09	-2,1±0,2
Phe^{41}-Arg^{17} (*mc*,**w**)	Arg^{17}→Arg_D[g] **(N/ Fis.)**	ΔF_4	71,76±0,04	70,33±0,04	71,76±0,04	70,0±0,2
		ΔF_3	75,10±0,08	75,06±0,08	75,10±0,08	74,8±0,3
		ΔΔA≈ΔΔF	-3,34±0,09	-4,73±0,09	-3,34±0,09	-4,8±0,4
Ser^{195}-Lys^{15} Gly^{193}-Lys^{15} (*mc*,**w**)	Lys^{15}→Lys_{D2}[g] **(N/ Fis.)**	ΔF_4	40,7±0,2	40,0±0,2	40,7±0,2	39,5±0,4
		ΔF_3	49,4±0,2	48,5±0,2	49,5±0,2	43,6±0,7
		ΔΔA≈ΔΔF	-8,7±0,3	-8,5±0,3	-8,8±0,3	-4,1±0,8
Gly^{216}-Pro^{13} (*mc*,**w**)	Pro^{13}→Pro_D[g] **(N/ Fis.)**	ΔF_4	83,2±0,1	81,6±0,1	83,1±0,1	82,1±0,5
		ΔF_3	103,6±0,2	102,4±0,2	103,6±0,2	102,6±0,8
		ΔΔA≈ΔΔF	-20,4±0,2	-20,8±0,3	-20,5±0,3	-20,5±0,9

sc- interacções envolvendo cadeias laterais ; *mc*- interacções envolvendo somente a cadeia principal;
sc-mc- interacções envolvendo cadeias laterais e a cadeia principal.
Fis.- Mutação física ; **N/ Fis.**- Mutação não física.
vs- muito forte; **s**- forte; **m**: intermédia ; **w**- fraca.
[a] O primeiro resíduo pertence à tripsina e o segundo resíduo ao PTI.
[b] Perturbação termodinâmica.
[c] Integração termodinâmica.
[d] Perturbação termodinâmica da primeira até à última janela; perturbação termodinâmica para os pontos extremos.
[e] Integração termodinâmica usando o método de extrapolação de Brooks III, a partir dos estados híbridos de referência λ=0.1, 0.3, 0.5, 0.7 e 0.9.
[f] Arginina neutra.
[g] O átomo responsável pela interacção é mutado por um pseudo-átomo mantendo as suas constantes de força ligantes mas anulando os seus parâmetros electrostáticos de van der Waals.

As variações da energia de Helmholtz de ligação (ΔΔA), associadas a cada mutação, calculadas pelos métodos da perturbação termodinâmica (TP) e da integração termodinâmica (TI) apresentam uma boa concordância. No entanto, as sua componentes (ΔF_3 e ΔF_4) apresentam maiores discrepâncias. A principal razão destas diferenças deriva de nos cálculos, usando o formalismo TI, as variações da energia de Helmholtz associadas às transições para os *pontos extremos* não serem calculadas. Estas variações normalmente não se cancelam em cada um dos processos não físicos (3 e 4), especialmente quando as referidas transições envolvem a aniquilação de um número diferente de átomos. No entanto, o seu efeito no ciclo termodinâmico global é habitualmente desprezável. Neste contexto, obteve-se uma concordância melhor entre os dois métodos introduzindo uma correcção perturbacional, associada às transições para os *pontos extremos*, aos resultados calculados pelo método da integração termodinâmica.

Dentro do formalismo TI, vários autores [6,9,14] encontraram uma variação aproximadamente linear de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ. No entanto, a zona vizinha dos pontos extremos (λ<0,1 ou λ>0,9) não foi amostrada directamente nestes cálculos, tendo as correspondentes médias de *ensemble* $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ sido obtidos através do método de extrapolação de Brooks III. No presente estudo, um comportamento linear para a função considerada pode ser observado para todas as mutações que envolvem apenas a alteração de parâmetros não ligantes dos átomos mutados. Nas restantes mutações, o comportamento linear é conservado nos *pontos centrais* e nos *pontos extremos* onde não ocorre aniquilação de átomos. No entanto, a variação de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ com λ apresenta grandes variações e afasta-se significativamente da linearidade sempre que a transição para um *ponto extremo* envolve a aniquilação de átomos. Estas variações são justificadas por a energia potencial dos átomos aniquilados tender para zero nesta zona, pelo que estes se desviam significativamente relativamente às suas posições de equilíbrio. Isto origina colisões repulsivas e um aumento bastante grande das componentes harmónicas da energia potencial ligante.

Numa mutação envolvendo a criação e o desaparecimento de átomos, como por exemplo a mutação $Asp^{189} \rightarrow Ala$ (ver figura 4.7), os átomos mutados são aniquilados quando $\lambda \rightarrow 0$ e apresentam grandes desvios em relação às suas posições de equilíbrio nesta zona. O comportamento dos átomos mutantes é análogo quando $\lambda \rightarrow 1$.

Numa mutação envolvendo apenas o desaparecimento de átomos, como por exemplo a mutação $Arg^{17} \rightarrow Argn$ (ver figura 4.8), os átomos mutados são aniquilados quando $\lambda \rightarrow 0$ e apresentam grandes desvios em relação às suas posições de equilíbrio nesta zona.

Numa mutação envolvendo apenas a alteração de parâmetros não ligantes, como por

exemplo a mutação $Lys^{15} \rightarrow Lys_{D1}$ (ver figura 4.9), não são alteradas as constantes de força internas dos átomos mutados. Nestas circunstâncias, estes átomos nunca se desviam significativamente em relação às suas posições de equilíbrio.

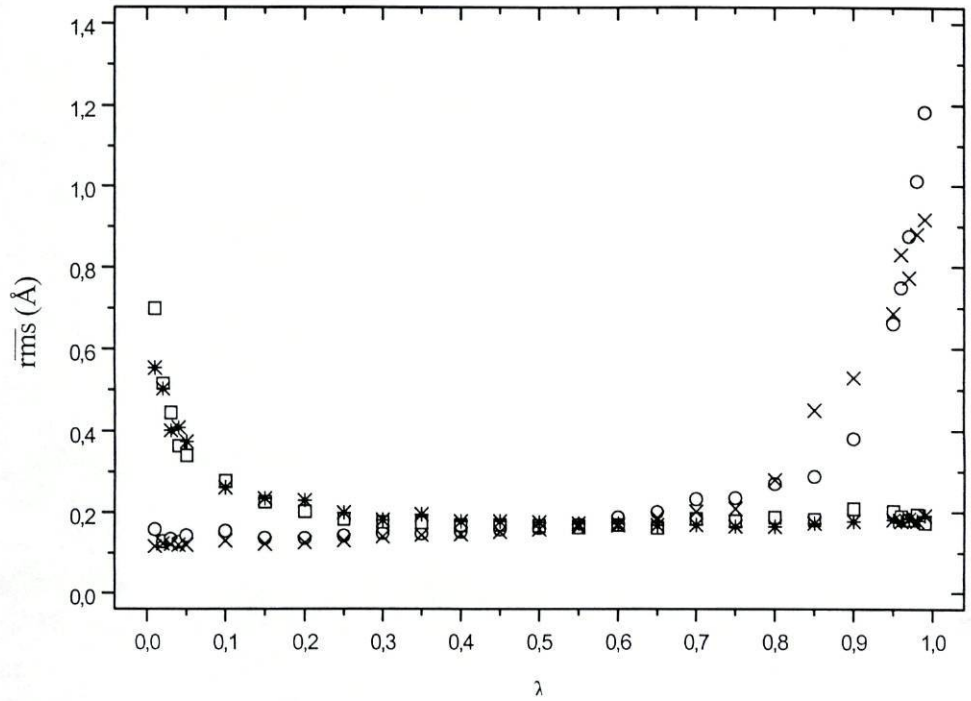


Figura 4.7. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) relativamente à estrutura média em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular dos átomos mutados durante o processo 3 (PTI→MPTI) (x); dos átomos mutados durante o processo 4 (Tripsina:PTI→Tripsina:MPTI) (o); dos átomos mutantes durante o processo 3 (PTI→MPTI) (*) e dos átomos mutados durante o processo 4 (Tripsina:PTI→Tripsina:MPTI) (o), na mutação $Asp^{189} \rightarrow Ala$.

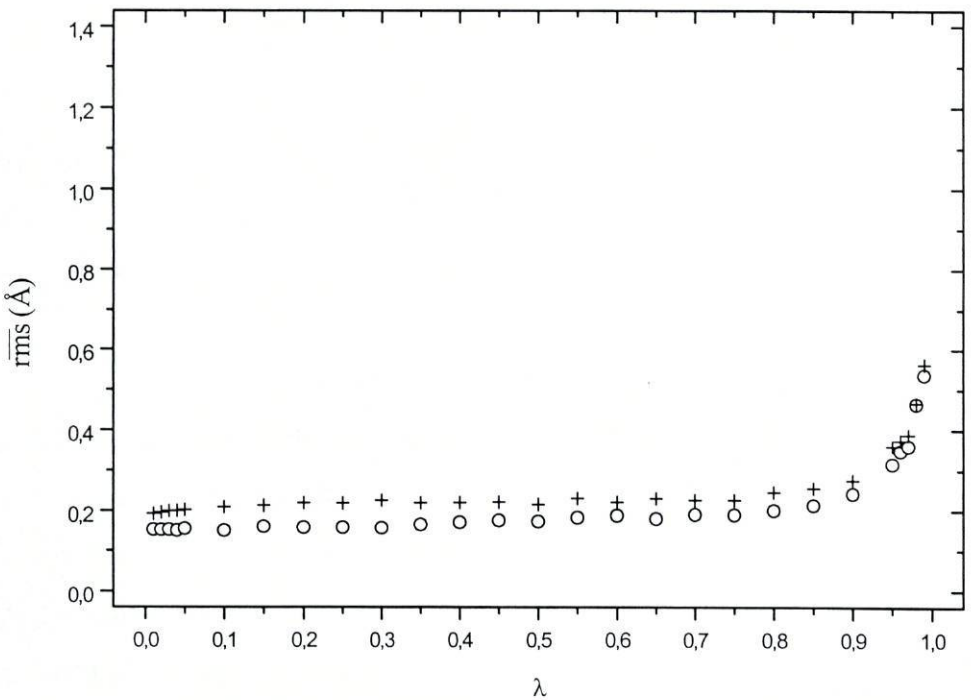


Figura 4.8. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) dos átomos mutados relativamente à estrutura média em função de λ, calculados na mutação $Arg^{17} \rightarrow Argn$ com base nas trajectórias obtidas por dinâmica molecular para o processo 3 (PTI→MPTI) (+) e para o processo 4 (Tripsina:PTI→Tripsina:MPTI) (o).

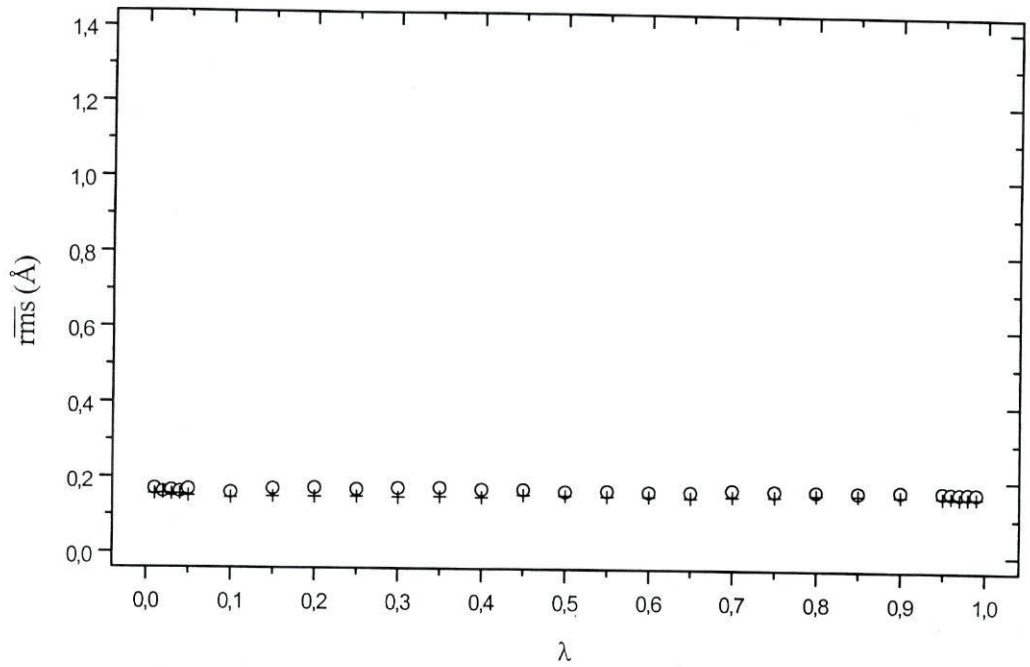


Figura 4.9. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) dos átomos mutados relativamente à estrutura média em função de λ, calculados na mutação $Lys^{15} \rightarrow Lys_{D1}$ com base nas trajectórias obtidas por dinâmica molecular para o processo 3 (PTI→MPTI) (+) e para o processo 4 (Tripsina:PTI→Tripsina:MPTI) (o).

Dentro do formalismo TI, o método de extrapolação de Brooks III tem sido bastante utilizado como uma forma de reduzir significativamente o tempo de simulação computacional [8,8,14].

Neste trabalho, os resultados obtidos com e sem extrapolação dentro do formalismo TI apresentam uma concordância razoável. Nos pontos extremos em que ocorre a aniquilação de átomos existem diferenças significativas entre os resultados obtidos pelas duas metodologias. No entanto, estas diferenças cancelam-se razoavelmente no ciclo termodinâmico global. Outra desvantagem detectada resulta deste método de extrapolação conduzir a erros estatísticos bastante maiores que as metodologias convencionais.

Para cada mutação, as componentes da energia de Helmholtz associadas aos processos não físicos (ΔF_3 e ΔF_4) são normalmente positivas. Como seria de esperar, a mutação de um resíduo específico por um não específico desestabiliza simultaneamente as espécies proteicas isoladas (tripsina ou PTI solvatados) e complexo tripsina:PTI solvatado. Este efeito de desestabilização é tanto maior quanto maiores forem as diferenças relativamente às propriedades químicas e à topologia entre o resíduo mutado e o resíduo mutante. De facto, as maiores desestabilizações ocorrem para mutações do tipo iónico→não polar (como por exemplo $Asp^{189} \rightarrow$Ala e $Arg^{17} \rightarrow$Argn) e para mutações envolvendo alterações topológicas significativas (como por exemplo $Gln^{192} \rightarrow$Leu). As mutações do tipo polar →polar ($Gln^{192} \rightarrow$Ser) e as restantes mutações, que são do tipo polar →não polar e não envolvem alterações topológicas significativas, provocam um efeito de desestabilização bastante menos significativo.

Um resultado inesperado ocorre para o processo 4 da mutação $Gln^{192} \rightarrow Gln_D$. De facto, a variação da energia de Helmholtz conformacional associada a este processo apresenta um valor ligeiramente negativo ($\Delta F_4 \approx -1,6$ kJ/mol). Isto significa que a mutação efectuada estabiliza ligeiramente o complexo tripsina:PTI. Com o objectivo de interpretar este resultado, determinaram-se as estruturas médias da interacção Gln^{192}-Cys^{14} no complexo tripsina:PTI para os estados inicial e final desta mutação (ver figuras 4.10 e 4.11).

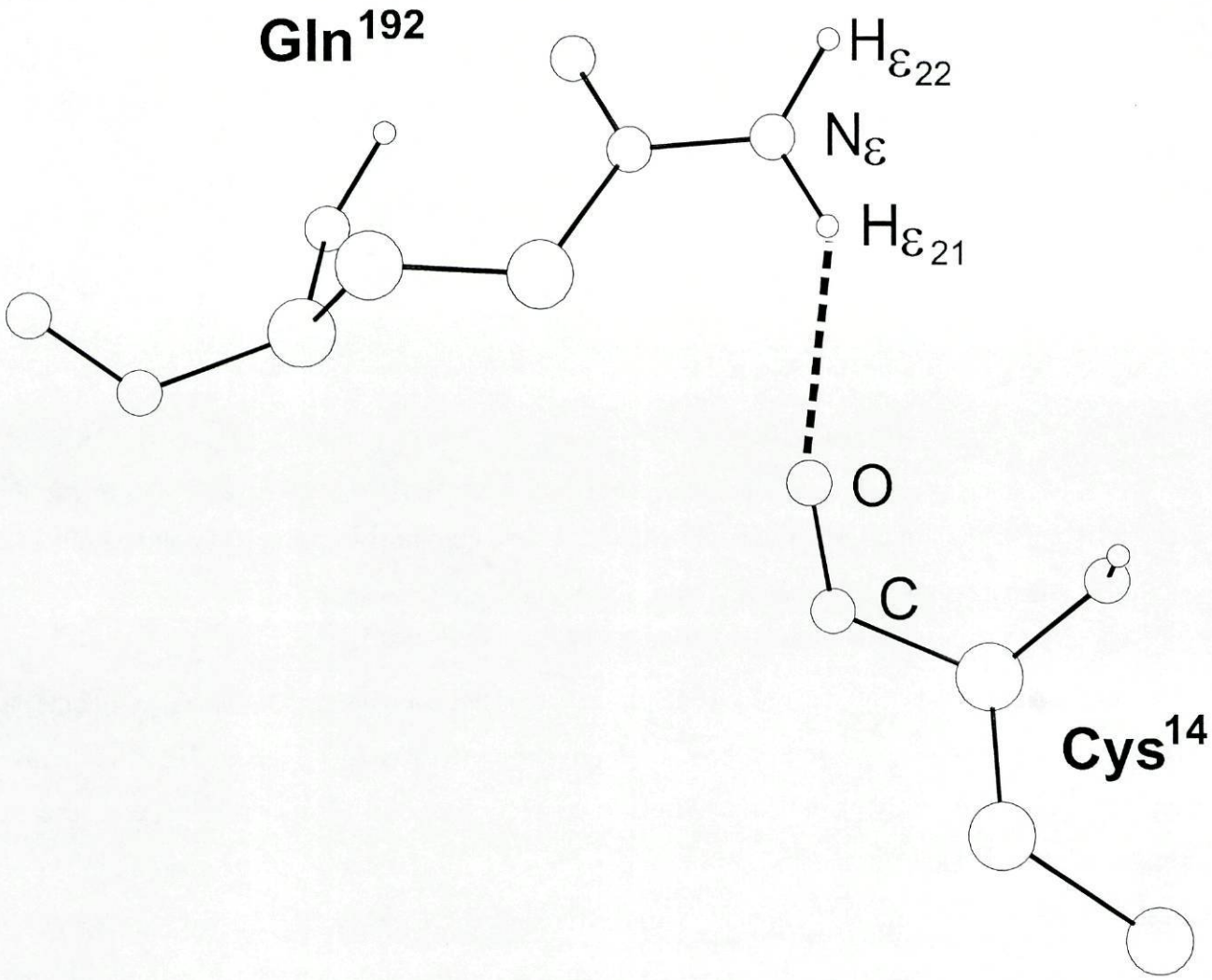


Figura 4.10. Estrutura média da interacção Gln^{192}-Cys^{14} no complexo tripsina:PTI obtida através de simulação MD para o estado inicial da mutação $Gln^{192} \rightarrow Gln_D$.

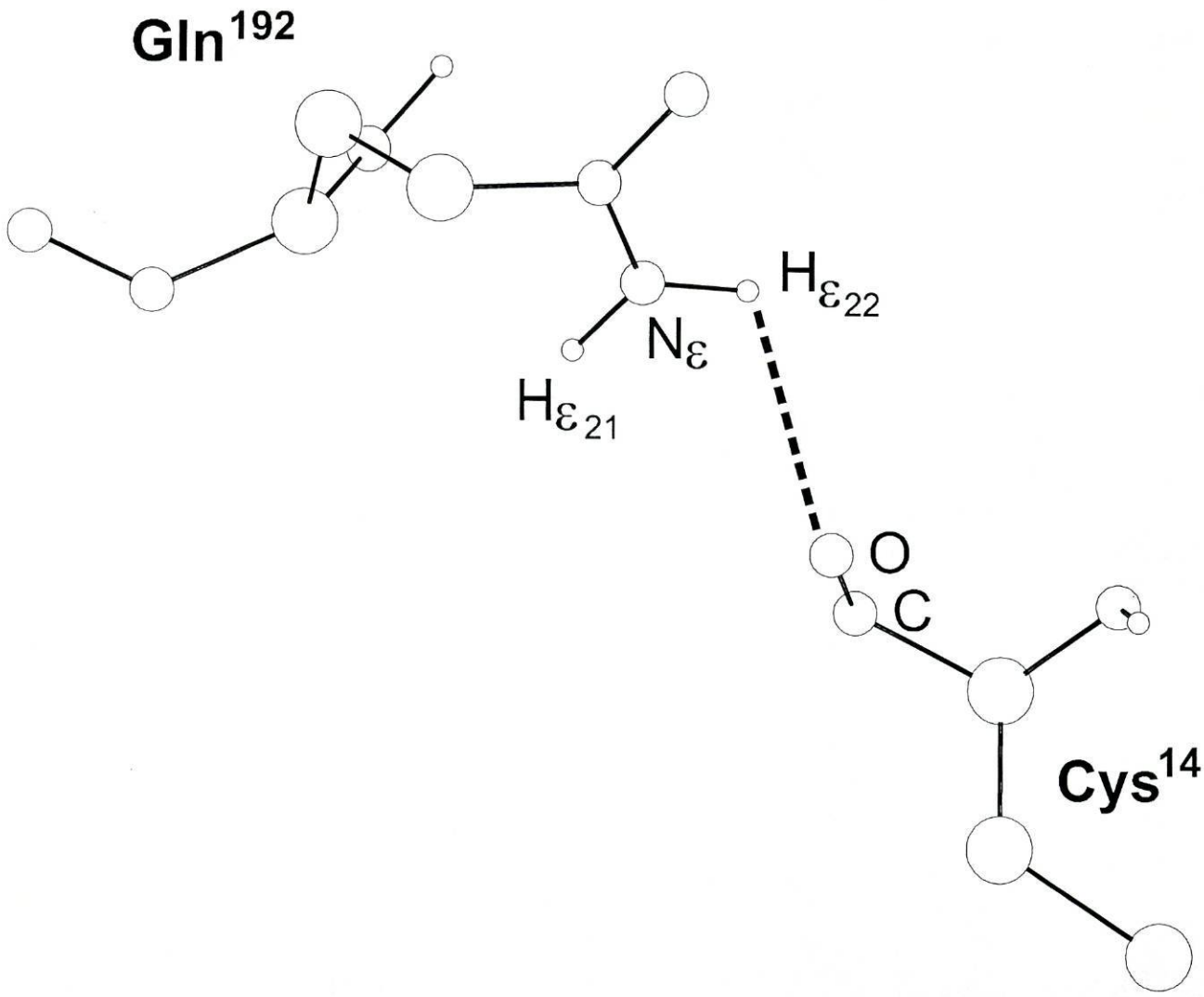


Figura 4.11. Estrutura média da interacção Gln^{192}-Cys^{14} no complexo tripsina:PTI obtida através de simulação MD para o estado final da mutação $Gln^{192} \rightarrow Gln_D$.

Analisando estas figuras pode concluir-se que, enquanto no estado inicial a interacção Gln^{192}-Cys^{14} está associada a uma ligação de hidrogénio envolvendo os átomos $H_{\varepsilon 21}$ do resíduo Gln^{192} e O do resíduo Cys^{14}, no estado final esta interacção está associada a uma ligação de hidrogénio envolvendo os átomos $H_{\varepsilon 22}$ do resíduo Gln^{192} e O do resíduo Cys^{14}. Isto significa que ao mesmo tempo que o átomo $H_{\varepsilon 21}$ é mutado, através da anulação dos seus parâmetros não ligantes, a sua interacção com o átomo O torna-se mais fraca e este estabelece uma ligação de hidrogénio alternativa com o átomo $H_{\varepsilon 22}$. As mesmas conclusões podem ser tiradas analisando os gráficos das distâncias $H_{\varepsilon 21}$-O e $H_{\varepsilon 22}$-O em função do tempo de simulação para os dois estados considerados (ver figuras 4.12 e 4.13). Neste contexto, compreende-se a variação cumulativa da energia de Helmholtz conformacional associada a este processo. ΔF_4 aumenta até λ=0,5 e a partir deste ponto começa a diminuir, sendo o seu valor global negativo (ver figura I.10).

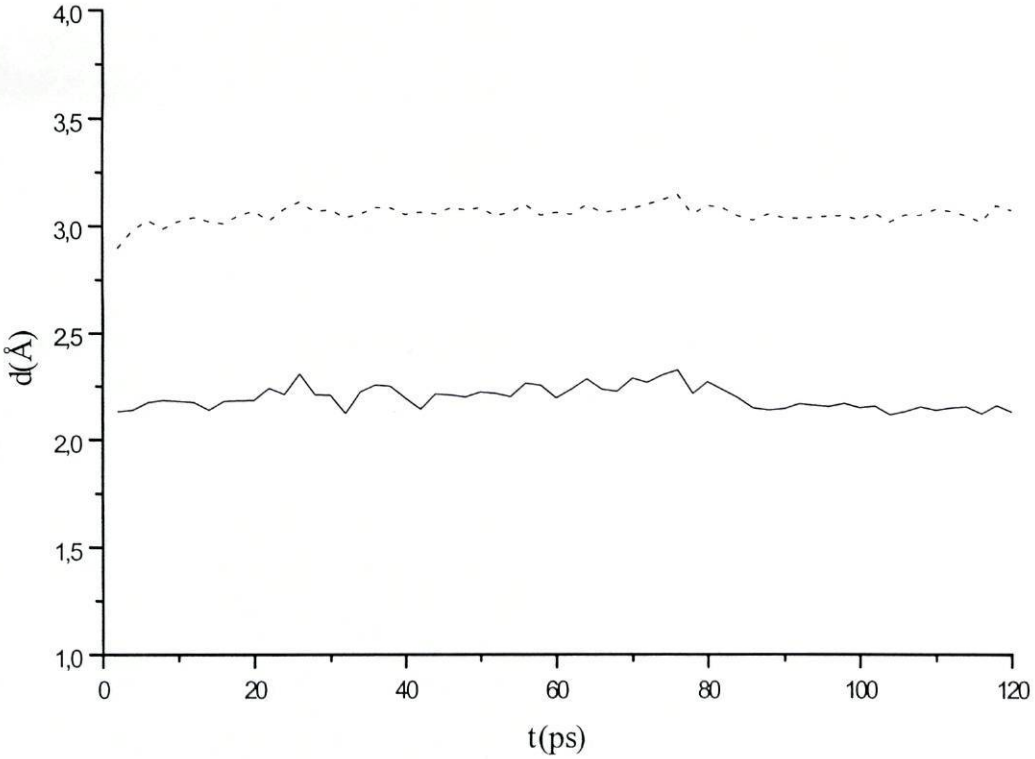


Figura 4.12. Distâncias $H_{\varepsilon 21}(Gln^{192})$-$O(Cys^{14})$ (—) e $H_{\varepsilon 22}(Gln^{192})$-$O(Cys^{14})$ (--)em função do tempo de simulação, para o estado inicial do processo 4 na mutação $Gln^{192} \rightarrow Gln_D$.

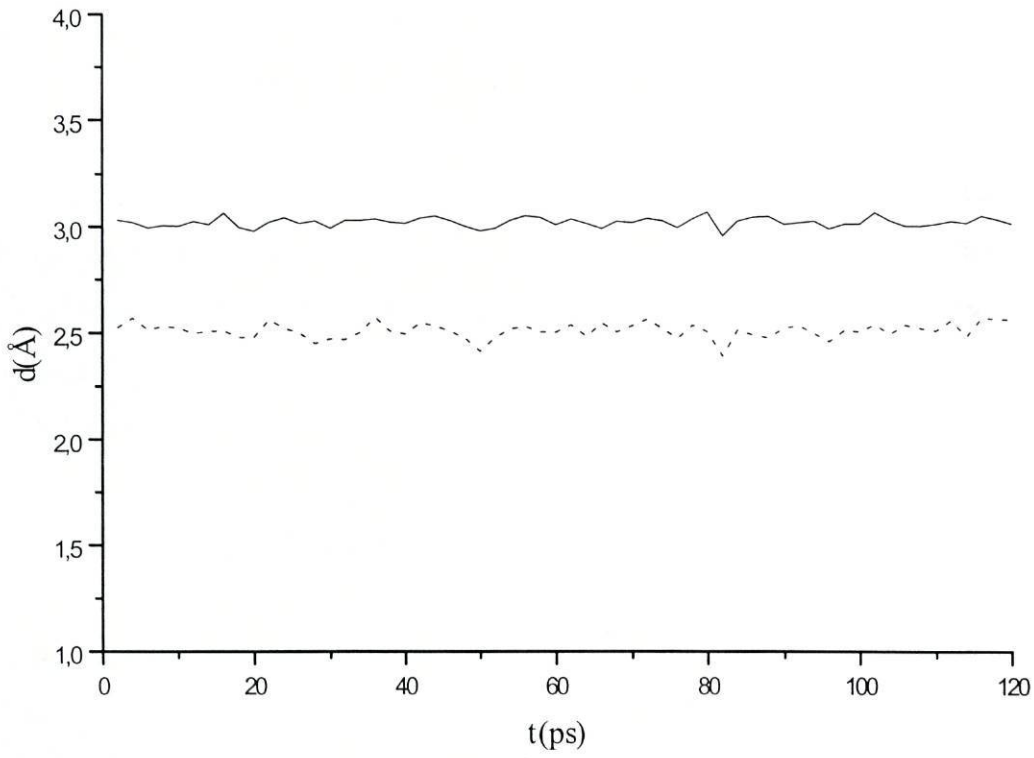


Figura 4.13. Distâncias $H_{\varepsilon 21}(Gln^{192})$-$O(Cys^{14})$(—) e $H_{\varepsilon 22}(Gln^{192})$-$O(Cys^{14})$ (--)em função do tempo de simulação, para o estado final do processo 4 na mutação $Gln^{192} \rightarrow Gln_D$.

Os valores das variações da energia de Helmholtz de ligação ($\Delta\Delta A$) dependem da natureza das interacções associadas, do meio em que ocorrem e das mutações utilizadas no seu estudo.

No processo de associação, um número elevado de moléculas de água são removidas das espécies proteicas tripsina e PTI para formar o complexo tripsina:PTI. A estabilização global originada por um dado par de resíduos (A,B) depende do balanço entre as ligações de hidrogénio que se estabelecem entre cada resíduo e o solvente no estado inicial (tripsina(aq) + PTI(aq)) e a interacção específica A-B no complexo tripsina:PTI(aq). O meio em que esta interacção ocorre também desempenha um papel importante. Num meio hidrofílico, os resíduos A e B podem estabelecer um grande número de interacções com o solvente e com outros resíduos polares ou carregados que permitem estabilizar o referido complexo.

Neste contexto, as interacções que se estabelecem entre cadeias laterais carregadas devem ser as mais fortes, seguindo-se as interacções entre uma cadeia lateral carregada e a cadeia principal, as interacções entre uma cadeia lateral polar e a cadeia principal e as interacções envolvendo somente a cadeia principal. No que se refere ao efeito do meio, a interacção Asp^{189}-Lys^{15} ocorre na cavidade ligante S_1 da tripsina que tem características fortemente hidrofóbicas. Quase todas as restantes interacções se estabelecem em regiões adjacentes que, embora menos hidrofóbicas, são ainda muito pouco acessíveis ao solvente. A única excepção é a interacção Tyr^{39}-Ile^{19} que ocorre numa região superficial do complexo tripsina:PTI de características bastante hidrofílicas.

A interacção Asp^{189}-Lys^{15}, que se estabelece entre cadeias laterais carregadas, é muito forte ($\Delta\Delta A(Asp^{189} \rightarrow Ala) \approx 120 kJ/mol$) apesar de ocorrer num meio hidrofóbico. A interacção His^{40}-Arg^{17}, por se estabelecer entre uma cadeia lateral carregada e a cadeia principal num meio hidrofóbico, é bastante menos forte mas ainda bastante atractiva ($\Delta\Delta A(Arg^{17} \rightarrow Argn) \approx 40 kJ/mol$).

Por outro lado, as interacções Tyr^{39}-Ile^{19} e Gln^{192}-Cys^{14}, envolvendo cadeias laterais polares e a cadeia principal, são bastante mais fracas. No entanto, a primeira interacção ocorre numa região muito mais hidrofílica que a segunda. Consequentemente, enquanto a mutação $Tyr^{39} \rightarrow Phe$ está associado a um valor de $\Delta\Delta A$ pequeno mas positivo, as mutações utilizadas no estudo da interacção Gln^{192}-Cys^{14} estão associadas a valores de $\Delta\Delta A$ negativos.

No estudo da interacção Gln^{192}-Cys^{14}, foram utilizadas três mutações ($Gln^{192} \rightarrow Leu$, $Gln^{192} \rightarrow Gln_D$ e $Gln^{192} \rightarrow Ser$). Os resultados obtidos pela mutação não física $Gln^{192} \rightarrow Gln_D$ são

fortemente influenciados pelas transformações anteriormente referidas que ocorrem no complexo tripsina:PTI durante o processo de mutação (ver figuras 4.10 a 4.13). Neste contexto, teria sido preferível mutar simultaneamente os átomos $H_{\varepsilon 21}$ e $H_{\varepsilon 22}$ ou o átomo O do resíduo Cys^{14} para obter o efeito pretendido (destruir a especificidade da interacção Gln^{192}-Cys^{14}). Os resultados obtidos nas duas restantes mutações físicas (Gln^{192}→Leu e Gln^{192}→Ser) são típicos relativamente a todas as interacções fracas estudadas na superfície de contacto do complexo tripsina:PTI. De facto, ambas as mutações desestabilizam quer a tripsina quer o complexo tripsina:PTI. No entanto, estas mutações ocorrem num meio hidrofílico na primeira espécie e num meio hidrofóbico na segunda. Consequentemente, a tripsina é sempre mais desestabilizada que o complexo tripsina:PTI, o que explica a estabilização global encontrada em ambas as mutações ($\Delta\Delta A<0$). Além disso, este efeito é mais acentuado na mutação Gln^{192}→Leu (polar→não polar; $\Delta\Delta A\approx$-50 kJ/mol) do que na mutação Gln^{192}→Ser (polar→polar; $\Delta\Delta A\approx$-13 kJ/mol).

Por fim as interacções polares envolvendo somente a cadeia principal (Ser^{214}-Lys^{15}; Phe^{41}-Arg^{17}; Ser^{195}-Lys^{15} , Gly^{193}-Lys^{15} e Gly^{216}-Pro^{13}) são fracas e ocorrem em regiões hidrofóbicas. Consequentemente, as mutações utilizadas no seu estudo estão sempre associadas a valores de $\Delta\Delta A$ negativos pelas mesmas razões anteriormente discutidas para a interacção Gln^{192}-Cys^{14}.

Para todas as mutações, foram ainda calculados os desvios quadráticos médios totais ($\overline{\text{rms}}$) das estruturas proteicas em função do parâmetro de acoplamento λ para os processos 3 e 4 nos ciclos (4.1) ou (4.2). Nestes cálculos, a estrutura de raios X e a estrutura média, obtida na simulação MD, foram usadas como referência. As flutuações foram calculadas separadamente para três conjuntos de átomos dos sistemas proteicos contidos na região activa: todos os átomos pesados, átomos pesados da cadeia principal e átomos pesados das cadeias laterais. Com base nos resultados obtidos (ver apêndice II) podem tirar-se as seguintes conclusões:

(1) Os $\overline{\text{rms}}$ calculados relativamente às estruturas de raios X são muito maiores que os calculados relativamente às estruturas médias obtidas nas simulações MD. Isto deve-se à natureza fundamentalmente distinta destes dois tipos de estruturas de referência. De facto, é perfeitamente natural que as flutuações de um sistema proteico solvatado sejam maiores

relativamente à sua estrutura cristalográfica (estrutura de raios X) do que relativamente à sua estrutura média em fase aquosa.

(2) As flutuações dos átomos pesados das cadeias laterais são sempre superiores às dos átomos pesados das cadeias laterais. Isto deve-se à localização preferencial dos primeiros átomos no exterior e dos segundos no interior das proteínas. As flutuações dos átomos pesados, em geral, apresentam valores obviamente intermédios relativamente às dos dois conjuntos anteriormente referidos.

(3) As estruturas dos sistemas proteicos solvatados em estudo não são muito afectadas pelas mutações efectuadas. De facto, as variações de $\overline{rms}$ com λ nunca são muito significativas. Este resultado é bastante importante para validar a metodologia utilizada no presente trabalho, pois os métodos de cálculo de variações da energia de Helmholtz aqui utilizados só são aplicáveis quando os estados inicial e final do sistema perturbado não diferem significativamente do ponto de vista estrutural. [3].

4.3- Compressão de ficheiros de trajectórias obtidos por dinâmica molecular.

O armazenamento dos ficheiros de trajectórias, produzidos através de dinâmica molecular no estudo das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI, representou um problema de difícil resolução. De facto, foram produzidos neste estudo 544 ficheiros deste tipo totalizando cerca de 65 *Gbytes* de informação.

Neste contexto, a utilização de algoritmos de compressão desempenhou um papel muito importante de forma a permitir armazenar toda esta informação para posterior análise. Por esta via, foi possível reduzir a informação a guardar para cerca de 25 *Gbytes* o que corresponde a uma eficiência média de compressão de cerca de 62%.

Nas compressões efectuadas, foram comparados os algoritmos conservativos clássicos e o algoritmo BS-VLC descrito na secção 2.5. Nesta secção, apresentam-se resultados típicos usando os ficheiros de trajectórias obtidos nas simulações dos processos 3 e 4 da mutação $Tyr^{39} \rightarrow Phe$, para o estado híbrido $\lambda=0,35$. Os modelos utilizados para a tripsina e complexo tripsina:PTI solvatados estão descritos na secção 4.2. As trajectórias obtidas tinham as seguintes dimensões:

$\mathbf{r}_u(1815,6000) \Rightarrow$ para a tripsina solvatada

$\mathbf{r}_u(1666,6000) \Rightarrow$ para o complexo tripsina:PTI solvatado

Nas compressões BS-VLC, foram utilizados 60 blocos. Foram efectuadas compressões preliminares, usando três coordenadas de referência alternativas (coordenadas atómicas, coordenadas atómicas dentro de um bloco e coordenadas atómicas descomprimidas) e um factor de escala 10^5. Os desvios quadráticos médios (rms), entre as trajectórias originais e as trajectórias obtidas após compressão e descompressão efectuadas sequencialmente usando o algoritmo BS-VLC e um factor de escala de 10^5, foram determinados em função do tempo de simulação.

Os resultados obtidos são apresentados na figura (4.14) e permitem confirmar que coordenadas atómicas, coordenadas atómicas dentro de um bloco e coordenadas atómicas descomprimidas originam erros cumulativos, cumulativos truncados e não cumulativos, respectivamente. Nestas circunstâncias, e tal como foi salientado na secção 2.5, as coordenadas atómicas descomprimidas constituem a melhor escolha para coordenadas de referência e foram usadas em todas as compressões BS-VLC efectuadas posteriormente.

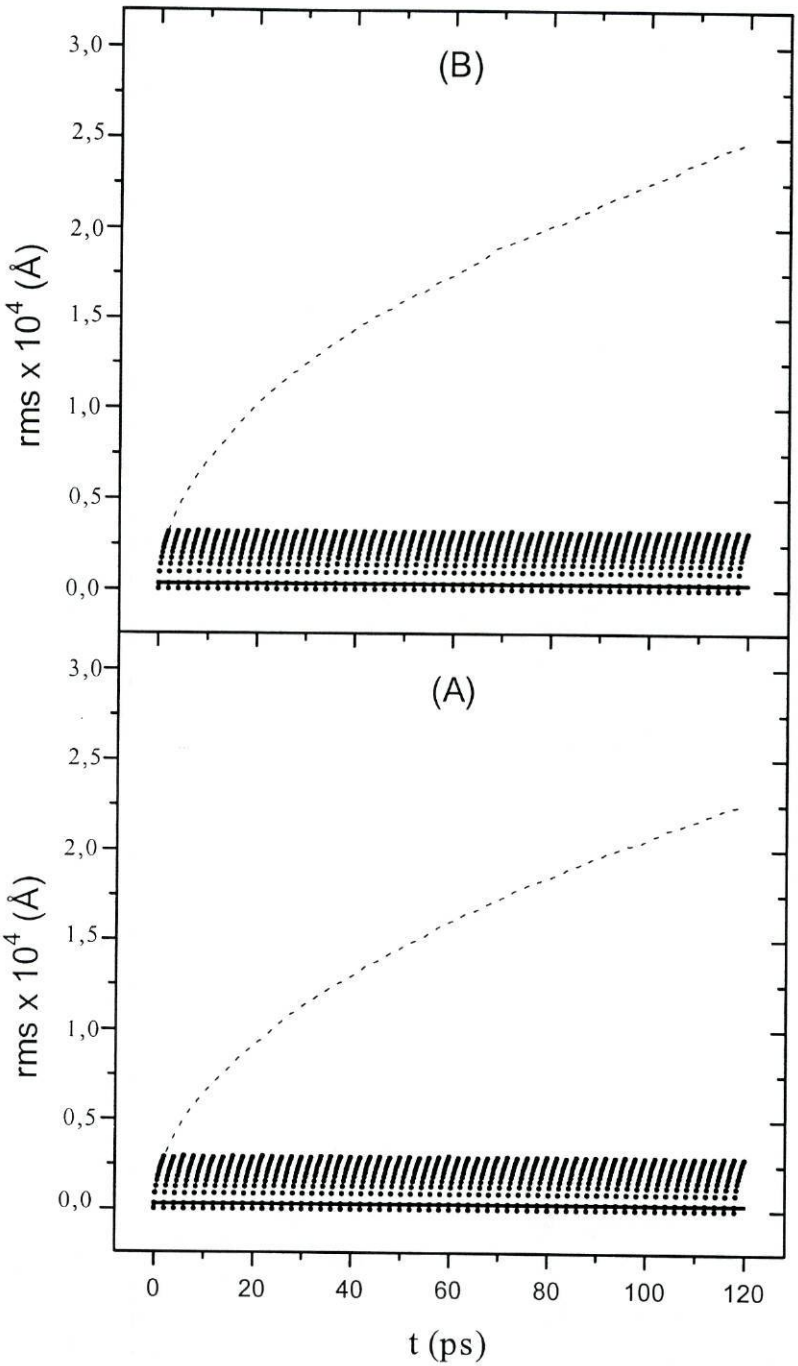


Figura 4.14. Desvios quadráticos médios (rms), entre as trajectórias originais obtidas através de dinâmica molecular para (A) a tripsina solvatada e (B) o complexo tripsina:PTI solvatado e as trajectórias obtidas após compressão e descompressão efectuadas sequencialmente usando o algoritmo BS-VLC e um factor de escala de 10^5, em função do tempo de simulação (t). Foram usadas três coordenadas de referência alternativas, coordenadas atómicas (- -), coordenadas atómicas dentro de um bloco (. .) e coordenadas atómicas descomprimidas (▬).

Para avaliar a eficiência do algoritmos BS-VLC, foram efectuadas compressões para os ficheiros da trajectórias da tripsina e do complexo tripsina:PTI usando 28 factores de escala diferentes. Os desvios quadráticos médios totais foram calculados para todos os casos. Os resultados obtidos são apresentados na figura (4.15) e na tabela 4.6. As eficiências de compressão obtidas com os diversos algoritmos são apresentadas na figura (4.16).

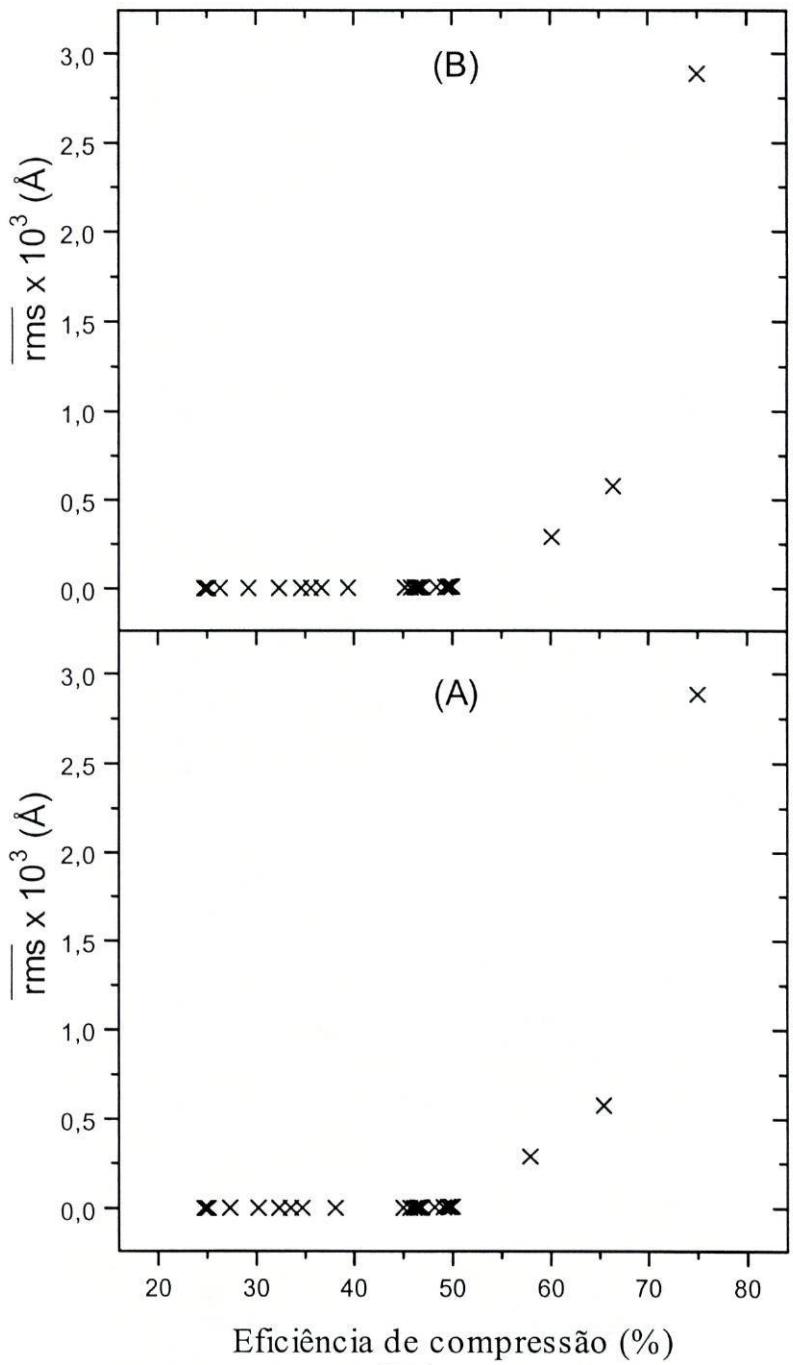


Figura 4.15. Desvios quadráticos totais ($\overline{rms}$), entre as trajectórias originais obtidas através de dinâmica molecular para (A) a tripsina solvatada e (B) o complexo tripsina:PTI solvatado e as trajectórias obtidas após compressão e descompressão efectuadas sequencialmente usando o algoritmo BS-VLC e 28 factores de escala diferentes, em função da eficiência de compressão.

Tabela 4.6. Eficiências de compressão e desvios quadráticos totais ($\overline{rms}$) entre as trajectórias originais, obtidas através de dinâmica molecular para a tripsina solvatada e o complexo tripsina:PTI solvatado, e as trajectórias obtidas após compressão e descompressão efectuadas sequencialmente usando o algoritmo BS-VLC e um factor de escala (Scale).

Tripsina			Tripsina:PTI		
Scale	Eficiência de compressão (%)	$\overline{rms}$ (Å)	Scale	Eficiência de compressão (%)	$\overline{rms}$ (Å)
$1{,}000\times 10^{7}$	24,7	0,000	$1{,}000\times 10^{7}$	24,7	0,000
$9{,}400\times 10^{6}$	24,8	$6{,}000\times 10^{-9}$	$9{,}400\times 10^{6}$	24,9	$6{,}000\times 10^{-9}$
$8{,}850\times 10^{6}$	24,9	$7{,}000\times 10^{-9}$	$8{,}300\times 10^{6}$	25,0	$9{,}000\times 10^{-9}$
$8{,}300\times 10^{6}$	25,0	$7{,}100\times 10^{-9}$	$7{,}200\times 10^{5}$	25,1	$1{,}400\times 10^{-8}$
$7{,}200\times 10^{5}$	25,1	$1{,}000\times 10^{-8}$	$4{,}108\times 10^{5}$	26,3	$3{,}664\times 10^{-7}$
$3{,}664\times 10^{5}$	27,3	$8{,}140\times 10^{-7}$	$3{,}664\times 10^{5}$	29,2	$6{,}090\times 10^{-7}$
$3{,}220\times 10^{5}$	30,2	$9{,}290\times 10^{-7}$	$3{,}220\times 10^{5}$	32,3	$7{,}790\times 10^{-7}$
$2{,}776\times 10^{5}$	32,3	$1{,}055\times 10^{-6}$	$2{,}776\times 10^{5}$	34,6	$9{,}050\times 10^{-7}$
$2{,}332\times 10^{5}$	33,5	$1{,}321\times 10^{-6}$	$2{,}332\times 10^{5}$	35,6	$1{,}055\times 10^{-6}$
$1{,}888\times 10^{5}$	34,7	$1{,}653\times 10^{-6}$	$1{,}888\times 10^{5}$	36,7	$1{,}285\times 10^{-6}$
$1{,}444\times 10^{5}$	38,1	$2{,}182\times 10^{-6}$	$1{,}444\times 10^{5}$	39,4	$1{,}685\times 10^{-6}$
$1{,}000\times 10^{5}$	45,0	$2{,}923\times 10^{-6}$	$1{,}000\times 10^{5}$	45,2	$3{,}197\times 10^{-6}$
$9{,}400\times 10^{4}$	45,7	$3{,}097\times 10^{-6}$	$9{,}400\times 10^{4}$	45,8	$3{,}421\times 10^{-6}$
$8{,}850\times 10^{4}$	46,1	$3{,}268\times 10^{-6}$	$8{,}850\times 10^{4}$	46,2	$3{,}651\times 10^{-6}$
$8{,}300\times 10^{4}$	46,4	$3{,}507\times 10^{-6}$	$8{,}300\times 10^{4}$	46,4	$3{,}884\times 10^{-6}$
$7{,}750\times 10^{4}$	46,6	$3{,}767\times 10^{-6}$	$7{,}750\times 10^{4}$	46,6	$4{,}130\times 10^{-6}$
$7{,}200\times 10^{4}$	46,7	$4{,}077\times 10^{-6}$	$7{,}200\times 10^{4}$	46,7	$4{,}399\times 10^{-6}$
$6{,}650\times 10^{4}$	46,8	$4{,}430\times 10^{-6}$	$6{,}650\times 10^{4}$	46,8	$4{,}699\times 10^{-6}$
$6{,}100\times 10^{4}$	46,9	$4{,}787\times 10^{-6}$	$6{,}100\times 10^{4}$	46,9	$5{,}021\times 10^{-6}$
$5{,}550\times 10^{4}$	47,0	$5{,}221\times 10^{-6}$	$5{,}550\times 10^{4}$	47,0	$5{,}399\times 10^{-6}$
$5{,}000\times 10^{4}$	48,2	$5{,}794\times 10^{-6}$	$5{,}000\times 10^{4}$	48,4	$5{,}849\times 10^{-6}$
$4{,}552\times 10^{4}$	49,1	$6{,}392\times 10^{-6}$	$4{,}552\times 10^{4}$	49,3	$6{,}404\times 10^{-6}$
$4{,}108\times 10^{4}$	49,5	$7{,}057\times 10^{-6}$	$4{,}108\times 10^{4}$	49,6	$7{,}052\times 10^{-6}$
$3{,}664\times 10^{4}$	49,8	$7{,}889\times 10^{-6}$	$3{,}664\times 10^{4}$	49,8	$8{,}042\times 10^{-6}$
$3{,}220\times 10^{4}$	50,0	$9{,}005\times 10^{-6}$	$3{,}220\times 10^{4}$	50,0	$9{,}134\times 10^{-6}$
$1{,}000\times 10^{3}$	57,9	$2{,}886\times 10^{-4}$	$1{,}000\times 10^{3}$	60,1	$2{,}887\times 10^{-4}$
$5{,}000\times 10^{2}$	65,4	$5{,}773\times 10^{-4}$	$5{,}000\times 10^{2}$	66,4	$5{,}772\times 10^{-4}$
$1{,}000\times 10^{2}$	75,0	$2{,}886\times 10^{-3}$	$1{,}000\times 10^{2}$	75,0	$2{,}887\times 10^{-3}$

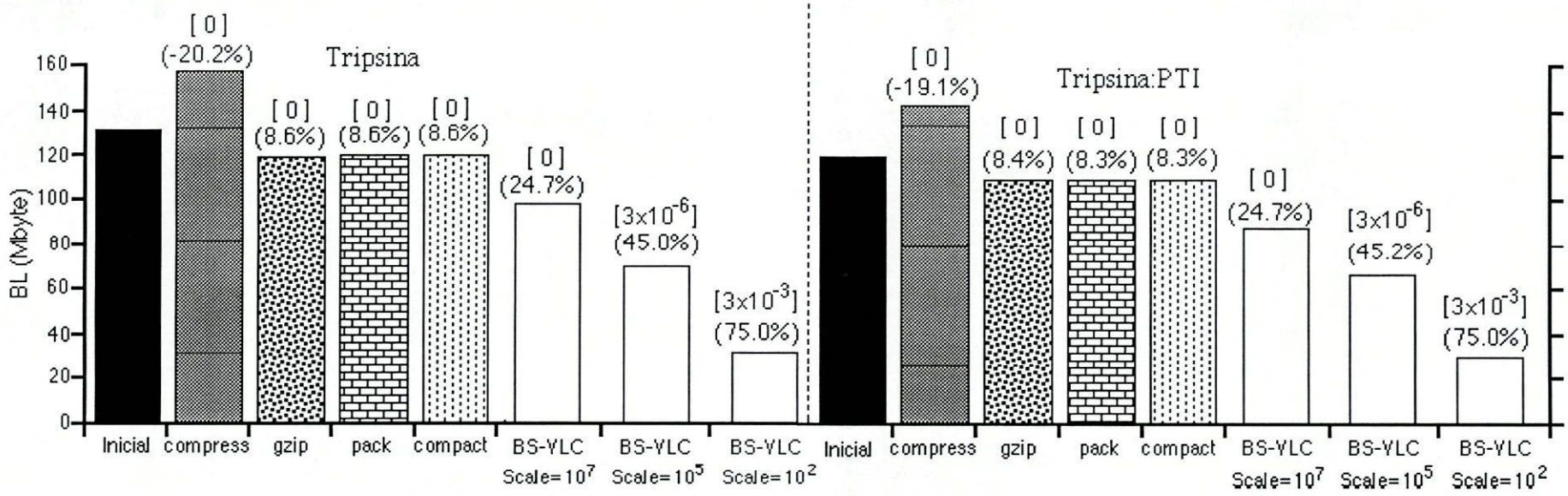


Figura 4.16. Dimensão (BL) dos ficheiros de trajectórias originais, obtidos através de dinâmica molecular para a tripsina solvatada e o complexo tripsina:PTI solvatado, e dos ficheiros comprimidos usando diversos algoritmos. As eficiências de compressão e desvios quadráticos médios totais, entre os ficheiros originais obtidas e os ficheiros obtidos após compressões e descompressões efectuadas sequencialmente, são também apresentados dentro de parêntesis curvos e rectos, respectivamente.

Os resultados obtidos permitem concluir que o algoritmo BS-VLC apresenta um comportamento quase conservativo ($\overline{rms}=0$) quando se usa um factor de escala próximo de 10^7. Nestas circunstâncias, o algoritmo BS-VLC quase triplica a eficiência de compressão do melhor algoritmo conservativo clássico (LZ77 usado no gzip). Além disso, podem-se obter eficiências de compressão bastante maiores (≈50%) com o algoritmo BS-VLC preservando um grau de precisão bastante elevado ($\overline{rms}$ entre 10^{-5} e 10^{-6}). Para obter eficiências de compressão maiores que 50%, a precisão diminui significativamente. No entanto, a eficiência de compressão máxima (75%) dentro deste algoritmo é obtida dentro de uma precisão razoável ($\overline{rms}\approx10^{-3}$).

4.4- Estudo quântico das interacções arginina-carboxilato.

As interacções arginina-carboxilato (Arg-Carb) foram tradicionalmente interpretadas com base na formação do complexo ligante duplo ião e na interacção coulômbica inerente à respectiva separação de carga. De facto, tanto as análises sistemáticas de estruturas cristalográficas [37-43], como a maioria dos cálculos teóricos [43-47] efectuados para este sistema foram interpretados como favorecendo o complexo duplo ião em detrimento do complexo neutro. No entanto, as estruturas de raios X não contêm informação explícita relativamente à posição dos átomos de hidrogénio e os cálculos teóricos foram efectuados apenas num conjunto restrito de conformações e utilizando bases de funções de pouca qualidade [48].

Neste trabalho, efectuou-se um estudo conformacional completo dos dois complexos ligantes associados às interacções Arg-Carb utilizando os modelos moleculares apropriados metilguanidínio-acetato (MGH^+:Ac^-) e metilguanidina-ácido acético (MG-HAc). Este estudo foi efectuado no vácuo o que permite reproduzir razoavelmente um ambiente proteico hidrofóbico [49], como por exemplo a cavidade ligante S_1 do factor β-XIIa onde ocorrem interacções Arg-Carb envolvendo inibidores com resíduo P_1 arginina.

A maioria das interacções Arg-Carb encontradas em proteínas estão associadas a conformações onde os átomos pesados de ambos os fragmentos se encontram no mesmo plano [37,43]. Nestas circunstâncias, o espaço conformacional pode ser definido fixando um

dos ângulos de rotação (α, α', ou α'') e optimizando os restantes parâmetros inter-fragmento (d e h) e intra-fragmento (ver figura 4.17).

Dentro do espaço conformacional definido na figura (4.17), podem ainda ser considerados dois tipos de conformações: conformações simétricas em que h=0 (ver figura 4.18) e conformações não simétricas em que h≠0.

Neste trabalho, foi efectuado inicialmente um estudo conformacional com modelos simétricos (ver figura 4.18) utilizando métodos *semi-empíricos* (AM1) e *ab initio* (6-31G** e MP2/6-31G**) [48,50]. Este estudo prévio, com um modelo molecular simplificado, destinou-se a determinar o nível de cálculo quântico mais baixo que permite uma descrição rigorosa das interacções Arg-Carb. Posteriormente, efectuou-se um estudo conformacional utilizando modelos não simétricos (ver figura 4.17) e o método quântico que, de acordo com os cálculos anteriores, apresentava um melhor compromisso entre a qualidade dos resultados obtidos e o tempo de computação gasto.

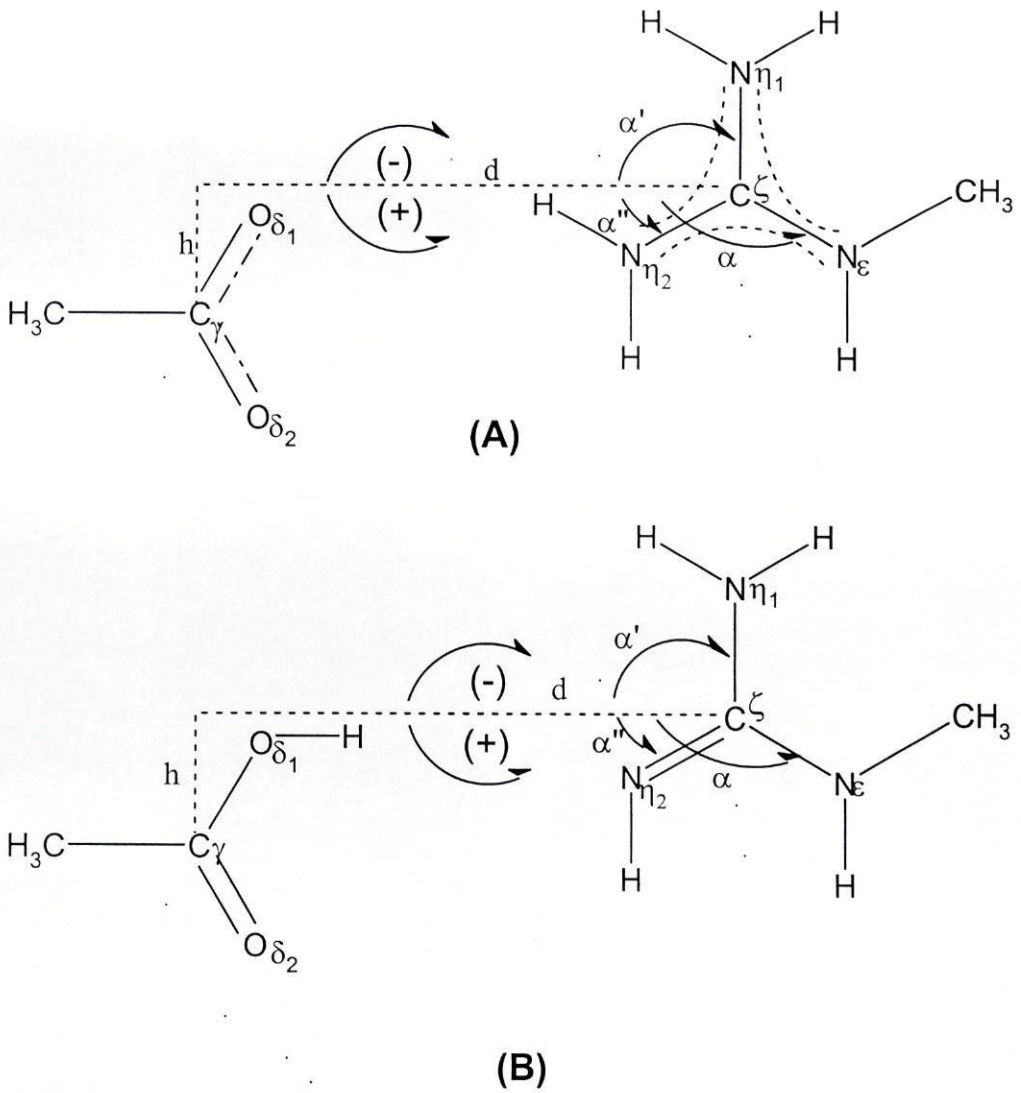


Figura 4.17. Representação esquemática dos modelos moleculares não simétricos metilguanidínio-acetato (A) e metilguanidina-ácido acético (B). Em ambos os sistemas, são apresentados o ângulo de rotação (α, α' ou α''), estudado nas análises conformacionais e o parâmetros inter-fragmento optimizáveis d e h.

Embora tenham sido utilizados ângulos de rotação diferentes em ambos os estudos (α' no estudo conformacional simétrico e α no estudo conformacional não simétrico), para permitir uma comparação mais imediata todos os resultados serão apresentados em termos do mesmo parâmetro de referência (α). A correspondência entre os parâmetros conformacionais α e α' é dada pela equação (4.3):

$$\begin{aligned} &\alpha = \alpha' - 120^{\circ}\,, \text{ se } \alpha' \in [-60^{\circ}:180^{\circ}] \\ &\alpha = \alpha' + 240^{\circ}\,, \text{ se } \alpha' \in [-180^{\circ}:-60^{\circ}]\,, \end{aligned} \tag{4.3}$$

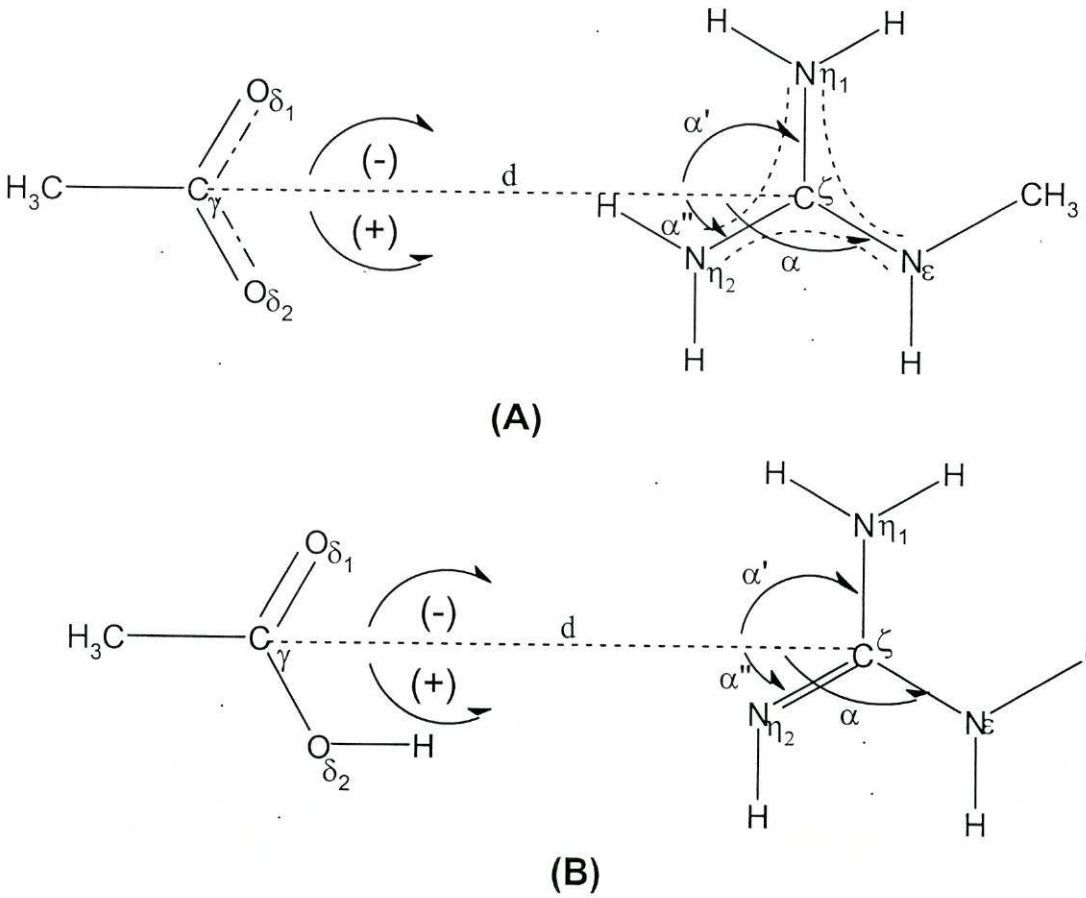


Figura 4.18. Representação esquemática dos modelos moleculares simétricos metilguanidínio-acetato (A) e metilguanidina-ácido acético (B). Em ambos os sistemas, são apresentados o ângulo de rotação (α, α' ou α''), estudado nas análises conformacionais e o parâmetro inter-fragmento optimizável d.

No estudo conformacional simétrico utilizou-se a seguinte metodologia:

Em cada ponto do espaço conformacional, o ângulo α' foi conservado constante enquanto os restantes parâmetros inter- e intra-fragmento foram optimizados (ver figura 4.18).

Os pontos iniciais da análise conformacional foram as conformações trans(H) (α=60º, α'=180º) e trans(CH_3). Estas conformações têm sido consideradas como as mais estáveis [38,42-48,50], pois são as únicas onde é possível estabelecer duas interacções do tipo N--H--C=O (ver figura 4.19).

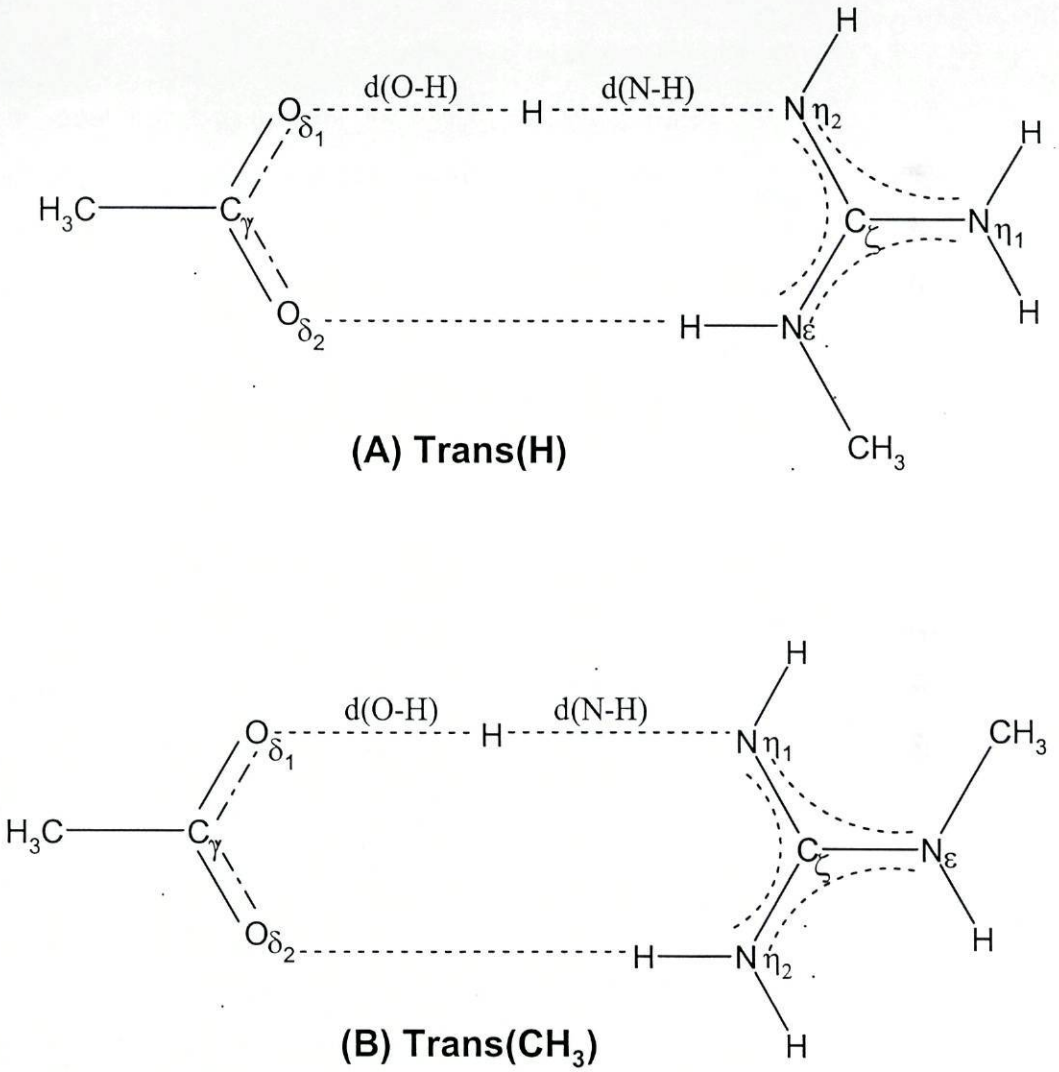


Figura 4.19. Representação esquemática das conformações trans(H) (α=60°, α'=180°) e trans(CH_3) (α=180°, α'= -60º). Em ambos os sistemas, são apresentadas duas escolhas possíveis (d_{N-H} e d_{O-H}) para coordenadas de reacção associadas à transferência protónica.

O modelo para o complexo neutro foi construído transferindo o protão mais favorável do metilguanidínio para o acetato. Em princípio, os valores de pK_a associados aos protões do

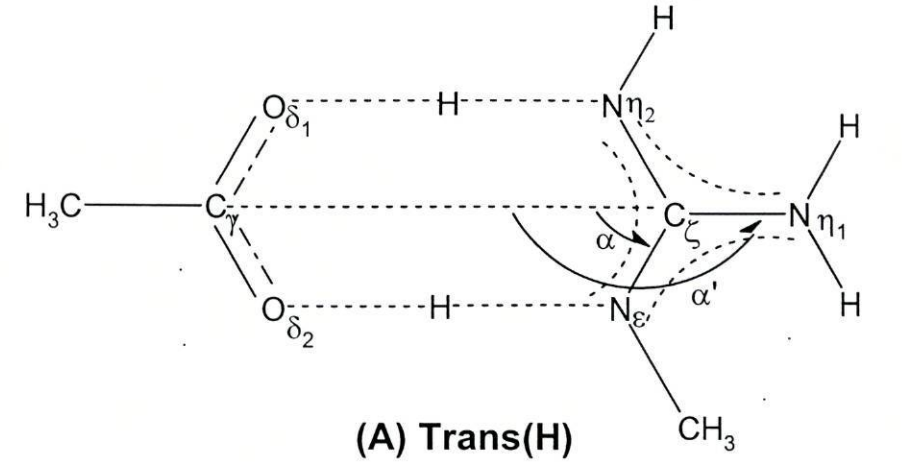
grupo guanidínio da arginina deveriam permitir escolher correctamente qual o protão a ser transferido em cada conformação. Os valores experimentais de pK_a foram determinados por Schmidt *et al* [51] como sendo iguais a 12,5, a 25 °C, quer para o protão do tipo N_ε-H quer para os protões do tipo N_η-H. No entanto, mesmo reconhecendo a quase-simetria do grupo em questão, parece altamente improvável que estes valores de pK_a sejam exactamente iguais e têm sido apresentados alguns resultados na literatura que confirmam esta hipótese [52]. Consequentemente, visto os resultados experimentais serem inconclusivos, nas conformações (trans(H), trans(CH_3), [α=-120°, α'=0°] e [α=0°, α'=120°]) onde dois protões tinham igual probabilidade de serem transferidos (ver figura 4.20), as duas geometrias foram consideradas nos cálculos. Para cada caso, foi escolhida a geometria com menor energia que foi considerada como modelo para o complexo neutro.

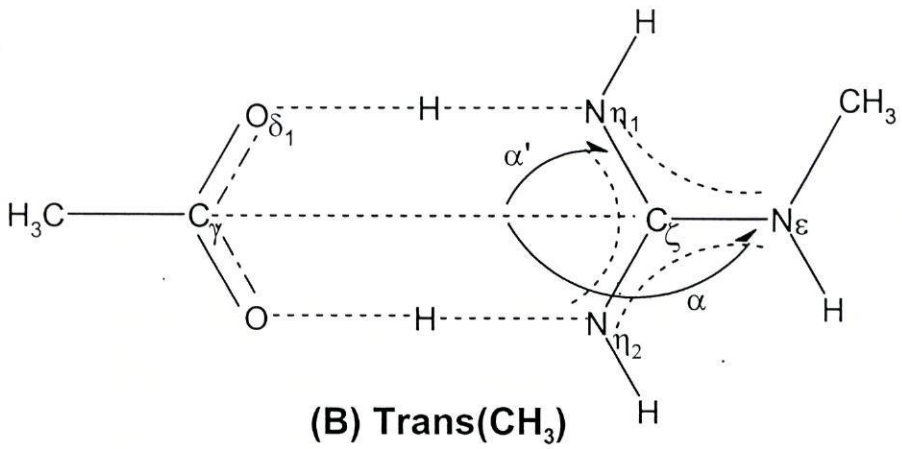
A geometria inicial de uma conformação foi obtida a partir da geometria previamente optimizada da conformação anterior e usando incrementos de 10° em α'. Nos cálculos AM1 e 6-31G** foram estudadas 36 conformações para cada complexo (72 conformações no total). Nos cálculos MP2/6-31G**, foram estudadas 6 conformações simetricamente distribuídas relativamente às conformações iniciais. Foram ainda efectuadas optimizações menos restritas das geometrias dos complexos duplo ião e neutro nas regiões trans(H) e trans(CH_3). Nestes cálculos, foram também utilizados os métodos AM1, 6-31G** e MP2/6-31G** e foi permitida a relaxação do ângulo α' durante o processo de optimização de forma a encontrar o mínimo global para cada complexo dentro do modelo utilizado.

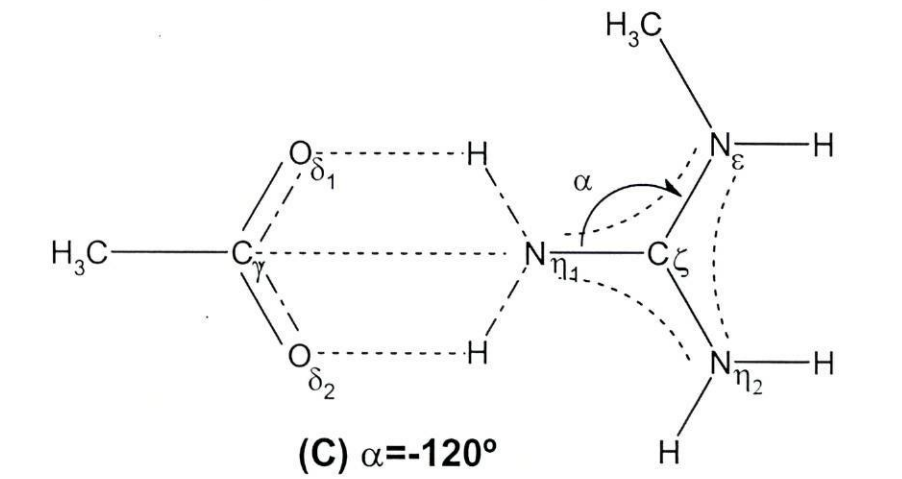
As energias de estabilização foram então calculadas utilizando o modelo da *super-molécula* descrito na secção 3.4. Nestes cálculos, utilizou-se o formalismo descrito na secção 3.5.7 em que o estado inicial não ligante (∞_1), constituído pelos fragmentos iónicos (MGH^+, Ac^-) isolados, como referência. As energias dos fragmentos iónicos e do estado inicial não ligante, calculadas por diversos métodos quânticos, serão apresentadas posteriormente na tabela 4.16.

As correcções de *counterpoise* foram calculadas para os mínimos globais obtidos nos cálculos *ab initio*. O *erro de sobreposição de bases* assim calculado era sempre inferior a 1,3 % das correspondentes energias de estabilização [50]. Nestas circunstâncias, esta correcção foi considerada de pouca importância nos sistemas em estudo e não foi efectuada nos restantes cálculos.

Os estudos conformacionais AM1 foram efectuados com o programa AMPAC [53]. Os estudos conformacionais *ab initio* forma efectuados com os programas GAUSSIAN 90 [54] e GAUSSIAN 92 [55]. Todos os cálculos foram efectuados num computador IBM Risc 6000.







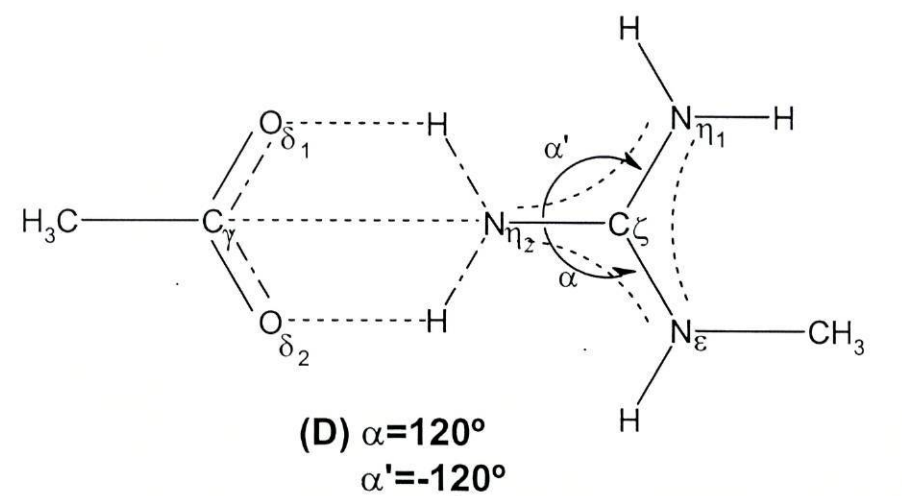


Figura 4.20. Representação esquemática das conformações onde, utilizando um modelo simétrico, existem dois protões com igual probabilidade de serem transferidos do metilguanidínio para o acetato.

As geometrias optimizadas dos complexos duplo ião e neutro são apresentadas nas figuras (4.21) e (4.22). Os parâmetros geométricos mais relevantes são apresentados nas tabelas 4.7 e 4.8.

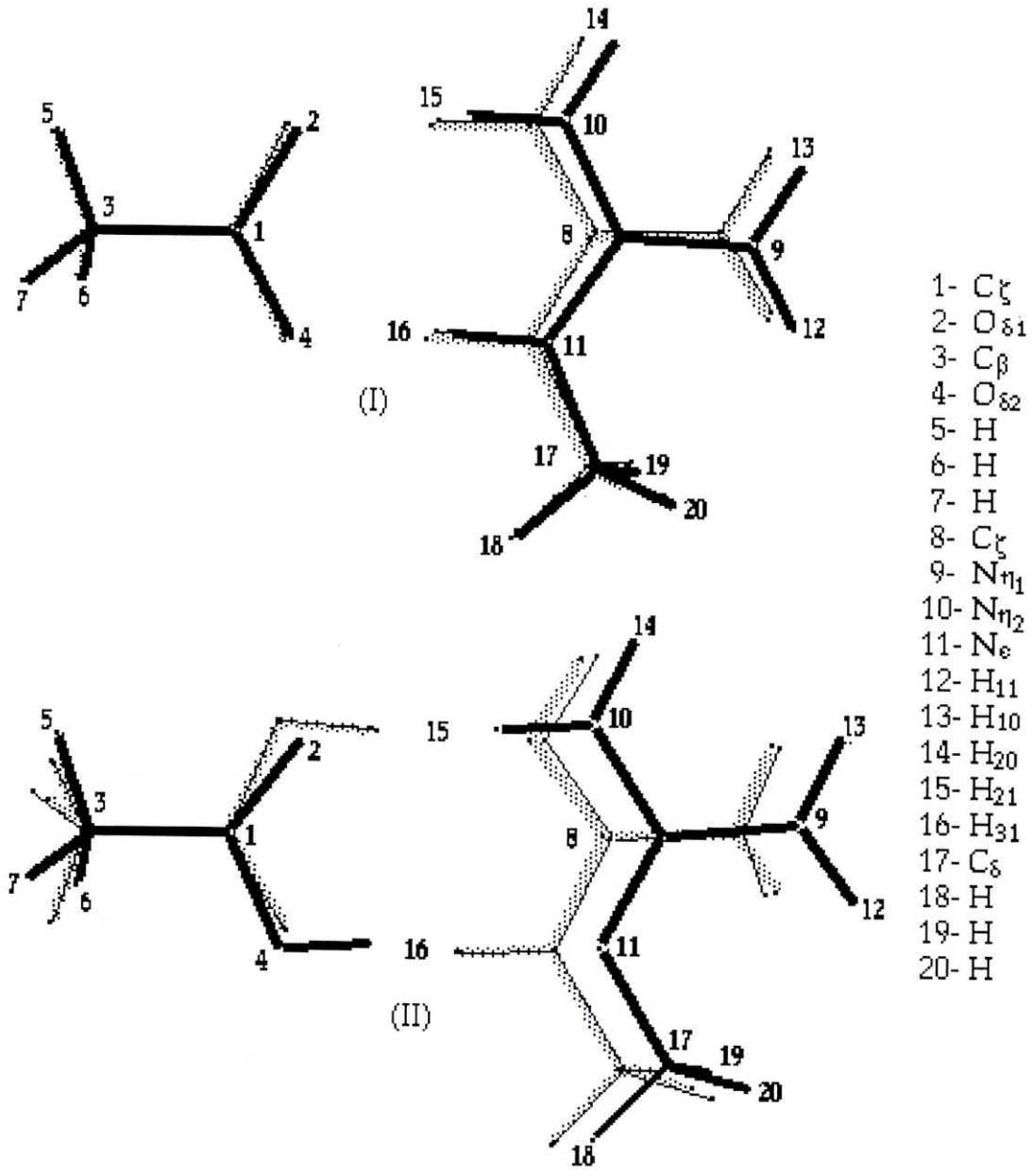


Figura 4.21. Geometrias optimizadas na região trans(H), utilizando modelos simétricos, dos complexos duplo ião (I) e neutro (II), obtidas com base em cálculos MP2/6-31G** (⁘), 6-31G** (—) e AM1 (▬).

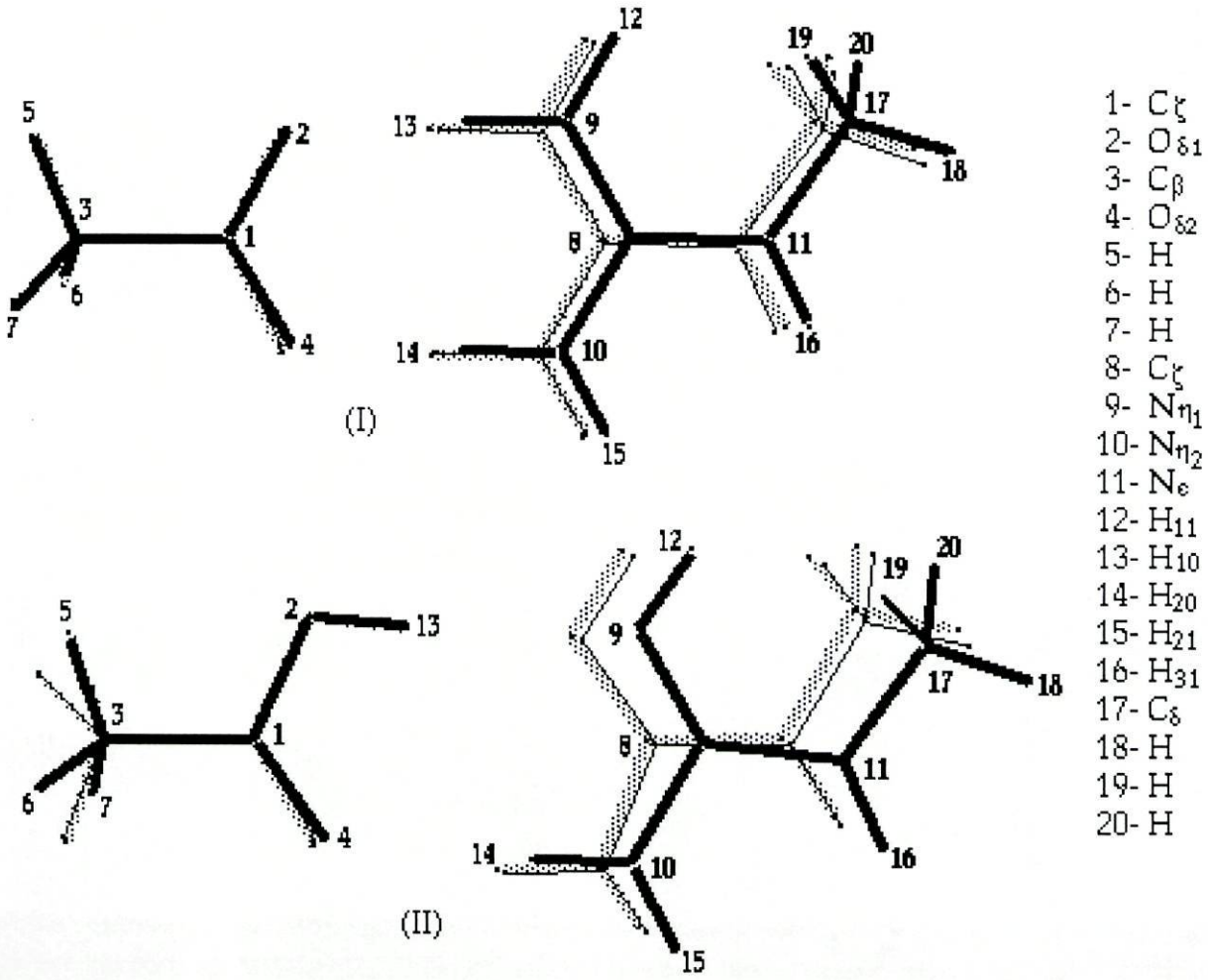


Figura 4.22. Geometrias optimizadas na região trans(CH_3), utilizando modelos simétricos, dos complexos duplo ião (I) e neutro (II), obtidas com base em cálculos MP2/6-31G** (⁘), 6-31G** (—) e AM1 (▬).

Tabela 4.7. Parâmetros geométricos optimizados[a,b], utilizando modelos simétricos, na região trans(H).

Método	Complexo	α	d	d(H_{21}-$N_{\eta 2}$)	d(H_{21}-$O_{\delta 1}$)	d(H_{31}-N_{ε})	d(H_{31}-$O_{\delta 2}$)	d($N_{\eta 2}$-$O_{\delta 1}$)	d(N_{ε}-$O_{\delta 2}$)
MP2/6-31G**	Duplo ião	60,00	3,83	1,08	1,51	1,09	1,50	2,60	2,59
MP2/6-31G**	Neutro	62,27	4,00	1,65	1,02	1,02	1,84	2,68	2,85
6-31G**	Duplo ião	60,00	3,88	1,04	1,62	1,04	1,62	2,66	2,66
6-31G**	Neutro	62,81	4,10	1,81	0,98	1,00	1,96	2,80	2,95
AM1	Duplo ião	57,10	4,12	1,02	1,83	1,05	1,69	2,85	2,74
AM1	Neutro	66,31	4,61	0,99	2,10	2,43	0,98	3,09	3,40

[a] Distâncias em angstroms, ângulos em graus.
[b] Parâmetros geométricos definidos de acordo com as figuras 4.18, 4.19 e 4.21.

Tabela 4.8. Parâmetros geométricos optimizados[a,b], utilizando modelos simétricos, na região trans(CH_3).

Método	Complexo	α	d	d(H_{21}-$N_{\eta 2}$)	d(H_{21}-$O_{\delta 1}$)	d(H_{31}-N_{ε})	d(H_{31}-$O_{\delta 2}$)	d($N_{\eta 2}$-$O_{\delta 1}$)	d(N_{ε}-$O_{\delta 2}$)
MP2/6-31G**	Duplo ião	-178,19	3,82	1,10	1,47	1,08	1,52	2,54	2,62
MP2/6-31G**	Neutro	-172,87	4,02	1,68	1,02	1,02	1,88	2,70	2,89
6-31G**	Duplo ião	-179,94	3,87	1,05	1,60	1,04	1,61	2,64	2,66
6-31G**	Neutro	-173,25	4,11	1,84	0,98	1,00	1,99	2,81	2,97
AM1	Duplo ião	-179,97	4,12	1,03	1,77	1,03	1,76	2,80	2,79
AM1	Neutro	179,96	4,57	2,37	0,98	0,99	2,10	3,34	3,09

[a] Distâncias em angstroms, ângulos em graus.
[b] Parâmetros geométricos definidos de acordo com as figuras 4.18, 4.19 e 4.22.

A análise destes resultados indica que a correlação electrónica não introduz nenhuma alteração significativa relativamente aos valores calculados ao nível 6-31G**. No entanto, os parâmetros geométricos AM1 apresentam algumas diferenças quando comparados com os valores obtidos ao nível *ab initio*. De facto, enquanto nos cálculos *ab initio* o protão é sempre transferido para o átomo $O_{\delta 1}$, nos cálculos AM1 a transferência protónica dá-se para o átomo $O_{\delta 2}$ na conformação trans(CH_3). As distâncias C_γ-C_δ (d) calculadas ao nível AM1 são sempre maiores que as previstas nos cálculos *ab initio*. Os valores obtidos para o ângulo de rotação (α) diferem também significativamente nos dois níveis de cálculo quântico, excepto para o complexo duplo ião na conformação trans(CH_3).

As distância N--O obtidas através de cálculos *ab initio* para o complexo duplo ião são menores que os valores médios determinados através de análises estatísticas das estruturas cristalográficas [35,40]. De facto, Singh *et al* [37] analisaram 37 estruturas proteicas de alta resolução e obtiveram uma distribuição normal para as distâncias N--O centrada em 2,8 Å. No seu estudo estatístico, Gorbitz [40] obteve valores médios para esta distância de 2,888 Å e 2,885 Å para as regiões trans(H) e trans(CH_3), respectivamente.

A maioria das interacções Arg-Carb, encontradas em estruturas cristalográficas de proteínas, ocorrem em ambientes hidrofílicos. Nestas circunstâncias, o complexo duplo ião deve ser mais estável que o complexo neutro devido às interacções com o solvente e com resíduos polares ou iónicos vizinhos. A razão dos valores teóricos *ab initio*, obtidos para as distâncias N--O, serem sistematicamente inferiores aos correspondentes valores experimentais deve-se certamente às interacções electrostáticas serem muito mais fortes no vácuo (ε=1) do que no meio cristalino (ε>>1). As distância N--O obtidas através de cálculos AM1 para o complexo duplo ião apresentam uma boa concordância com os valores experimentais. No entanto, esta concordância é provavelmente fortuita e deve resultar de uma descrição incorrecta das interacções electrostáticas no vácuo dada por este método *semi-empírico*.

As distâncias N--O calculadas para o complexo neutro não devem ser directamente comparadas com valores médios obtidos em análises estatísticas sistemáticas de estruturas cristalográficas. Estes valores teóricos deverão ser comparados com valores médios experimentais obtidos para interacções Arg-Carb que ocorram em ambientes hidrofóbicos.

Na figura (4.23) é apresentado um gráfico da energia de estabilização (ΔE) em função do ângulo de rotação (α) obtido com os três métodos teóricos acima mencionados. Estes resultados são também apresentados na tabela 4.9.

A análise destes resultados permite demonstrar que o complexo neutro é mais estável que o complexo duplo ião quando existe a possibilidade de um protão terminal N-H ser transferido do metilguanidínio para o acetato. Na região conformacional mais instável (α=-50° a α=0°), ocorre uma inversão a esta tendência geral. Nesta região, o grupo metilo interactua fortemente com o acetato e os únicos protões que podem ser transferidos pertencem a este grupo. Estes protões apresentam um carácter ácido muito pouco acentuado o que explica a inversão detectada.

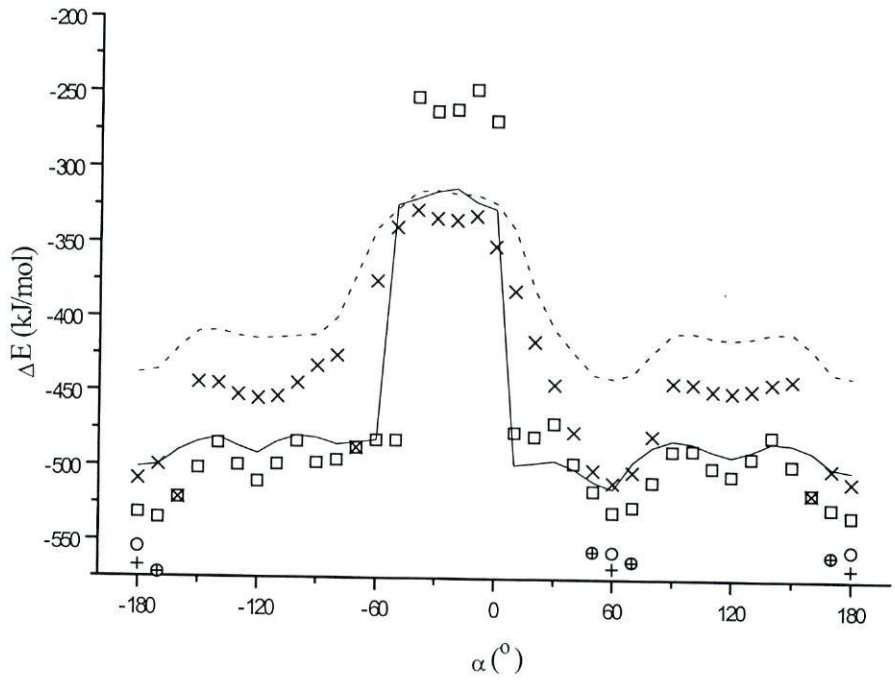


Figura 4.23. Gráfico da energia de estabilização (ΔE) em função do ângulo de rotação (α), obtido com modelos não simétricos, para o complexo metilguanidínio-acetato através de cálculos AM1 (--); o complexo metilguanidina-ácido acético através de cálculos AM1 (⁻) , o complexo metilguanidínio-acetato através de cálculos 6-31G** (□); o complexo metilguanidina-ácido acético através de cálculos 6-31G** (x), as conformações onde ocorreu transferência de protão durante a optimização de um complexo duplo ião num cálculo 6-31G** (⊠) ,o complexo metilguanidínio-acetato através de cálculos MP2/6-31G** (o); o complexo metilguanidina-ácido acético através de cálculos MP2/6-31G** (+) e as conformações onde ocorreu transferência de protão durante a optimização de um complexo duplo ião num cálculo MP2/6-31G** (⊕).

Tabela 4.9. Energia de estabilização (kJ/mol), obtida com modelos simétricos, em função do ângulo de rotação α (°).

	MP2/6-31G**		6-31G**		AM1	
α	zwitter.	neut.	zwitter.	neut.	zwitter.	neut.
-170	-572,22+	-572,22	-499,65	-535,27	-435,86	-499,73
-160	____	____	-521,40+	-521,40	-420,88	-489,86
-150	____	____	-444,33	-501,74	-409,53	-484,10
-140	____	____	-444,89	-484,62	-409,00	-480,65
-130	____	____	-452,33	-499,60	-412,76	-486,76
-120	____	____	-454,72	-510,37	-414,47	-491,39
-110	____	____	-453,54	-498,86	-413,87	-483,97
-100	____	____	-444,51	-483,39	-412,72	-479,45
-90	____	____	-432,81	-497,79	-411,66	-481,01
-80	____	____	-425,68	-495,93	-400,16	-485,29
-70	____	____	-487,38+	-487,38	-369,90	-484,09
-60	____	____	-375,36	-482,23	-339,74	-482,17
-50	____	____	-339,26	-252,76	-327,03	-324,10
-40	____	____	-327,64	-262,20	-315,34	-319,94
-30	____	____	-333,31	-260,73	-314,50	-315,34
-20	____	____	-334,29	-247,47	-316,57	-313,20
-10	____	____	-331,48	-267,95	-317,00	-321,94
0	____	____	-351,87	-351,87#	-322,66	-326,94
10	____	____	-381,49	-476,88	-340,35	-498,59
20	____	____	-415,34	-479,34	-380,70	-497,38
30	____	____	-444,53	-470,34	-406,88	-495,62
40	____	____	-476,35	-497,31	-422,92	-500,87
50	-556,57+	-556,57	-501,74	-515,69	-437,54	-509,28
60	-556,99	-567,81	-510,40	-530,06	-440,02	-514,48
opt*	-556,99	-568,25	-510,40	-530,71	-440,22	-514,87
70	-563,25+	-563,25	-502,95	-526,50	-436,62	-496,26
80	____	____	-478,63	-509,56	-421,85	-509,56
90	____	____	-442,92	-488,63	-408,96	-481,22
100	____	____	-443,67	-488,05	-408,78	-483,06
110	____	____	-448,02	-499,45	-412,01	-488,50
120	____	____	-449,11	-504,97	-413,03	-492,07
130	____	____	-447,53	-492,98	-411,84	-488,08
140	____	____	-442,92	-478,34	-408,53	-482,20
150	____	____	-440,85	-497,78	-408,49	-483,35
160	____	____	-516,83+	-516,83+	-419,65	-488,66
170	-558,63+	-558,63	-500,25	-526,45	-435,44	-498,83
180	-554,84	-567,35	-508,92	-531,68	-438,14	-501,30
opt**	-554,97	-573,15	-508,92	-536,35	-438,14	-501,30

zwitter. : complexo ligante duplo ião ; neut. : complexo ligante neutro.

* Valor de α optimizado na região trans(H). Os parâmetros geométricos são apresentados na tabela 4.7.

**Valor de α optimizado na região trans(CH_3). Os parâmetros geométricos são apresentados na tabela 4.8.

+ Conformações onde ocorreu uma transferência de protão durante a optimização de uma geometria inicial do tipo duplo ião.

Conformações onde ocorreu uma transferência de protão durante a optimização de uma geometria inicial do tipo neutro.

As diferenças de energia entre os dois complexos são significativamente superiores nos cálculos AM1 do que nos cálculos *ab initio*. Para ambos os mínimos conformacionais, as diferenças de energia obtidas pelos cálculos *ab initio* (6-31G** e MP2/6-31G**) são similares.

De acordo com os três métodos, os mínimos do espaço conformacional ocorrem nas regiões trans(H) e trans(CH_3), o que está de acordo com todos os cálculos teóricos [43-47] e análises estatísticas de estruturas cristalográficas [37,41,43] efectuadas para este sistema.

Nos cálculos AM1 e 6-31G**, foram também detectados dois mínimos locais pouco acentuados para os dois complexos ligantes em estudo. Estes pontos correspondem às conformações (α=-120° e α=0°) onde existem dois protões pertencentes ao mesmo átomo de azoto com igual probabilidade de serem transferidos do metilguanidínio para o acetato (ver figura 4.20).

Nos cálculos 6-31G**, foi detectada a transferência de protão, do metilguanidínio para o acetato, durante a optimização do complexo duplo ião. Este fenómeno foi observado em três conformações (α=-160°, α=-70° e α=160°). Nos cálculos MP2/6-31G**, esta transferência de protão foi detectada em todas a regiões vizinhas das conformações trans(H) e trans(CH_3).

Os cálculos efectuados com modelos simétricos indicam que os resultados obtidos aos níveis 6-31G** e MP2/6-31G** são bastante concordantes. Neste contexto, não foi considerado necessário incluir o tratamento da correlação electrónica para obter resultados de boa qualidade nos cálculos efectuados posteriormente com modelos não simétricos.

As diferenças obtidas entre os resultados dos cálculos *ab initio* e AM1 são bastante mais significativas. Consequentemente, decidiu-se não utilizar este método *semi-empírico* nos estudos deste sistema posteriormente efectuados. No entanto, este método poderá constituir uma primeira aproximação razoável no estudo de interacções Arg-Carb em sistemas de grande dimensão. De facto, a inclusão explícita do ambiente proteico e de uma esfera de solvatação em torno do par interactuante pode ser muito importante para obter uma descrição correcta destas interacções em ambientes hidrofílicos. Nestas circunstâncias, uma aproximação *ab initio* pode tornar-se incomportável em termos dos recursos computacionais actualmente disponíveis e os cálculos *semi-empíricos* podem constituir uma estratégia alternativa razoável.

Em seguida, efectuou-se uma análise conformacional dos dois complexos ligantes em estudo utilizando modelos não simétricos (ver figura 4.17). Neste estudo [33], foi utilizado o formalismo *ab initio* com uma base 6-31G**, visto os cálculos anteriormente efectuados com modelos simétricos terem demonstrado ser este o nível mais baixo de tratamento quântico que permite obter uma descrição adequada das interacções Arg-Carb.

Nestes cálculos, o ângulo de rotação α foi conservado constante enquanto os restantes parâmetros inter- (d e h) e intra-fragmentos foram optimizados.

As conformações trans(H) e trans(CH_3) foram também as geometrias iniciais escolhidas para iniciar esta análise conformacional. Cada conformação posterior foi obtida a partir da imediatamente anterior já optimizada e com incrementos de 10° em α. Foram estudadas 36 conformações para cada complexo, ou seja, 72 conformações no total.

A energia de estabilização foi calculada para os sistemas modelo em estudo utilizando o formalismo anteriormente descrito. Tendo em conta os resultados obtidos anteriormente com os modelos simétricos, as correcções de *counterpoise* não foram calculadas neste estudo.

Todos os cálculos foram efectuados com o programa GAUSSIAN 92 [55] num computador IBM Risc 6000.

Os resultados obtidos são apresentados na figura (4.24) e na tabela 4.10.

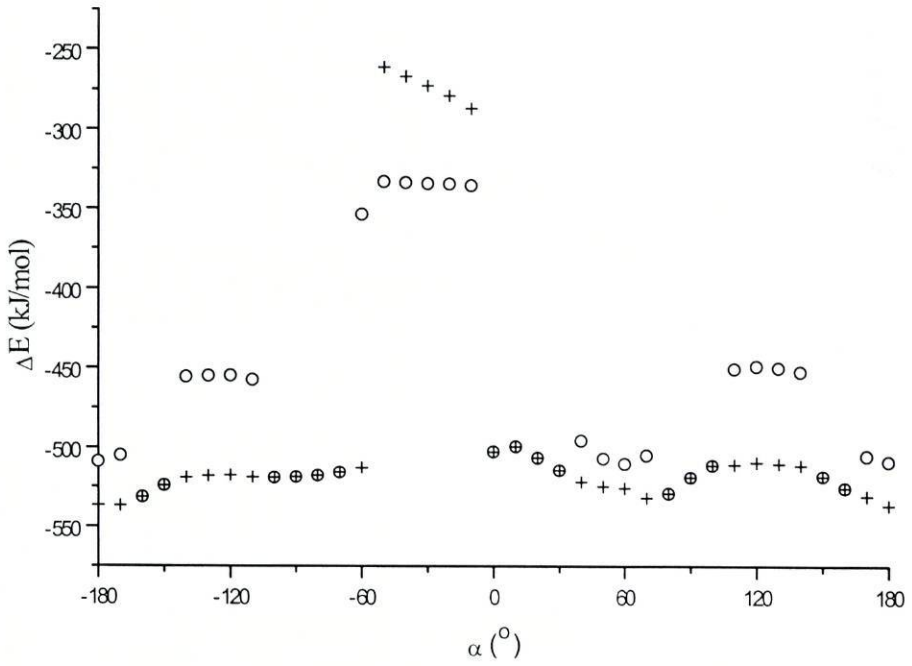


Figura 4.24. Gráfico da energia de estabilização (ΔE) em função do ângulo de rotação (α), obtido através de cálculos 6-31G** com modelos não simétricos, para o complexo metilguanidínio-acetato (o); o complexo metilguanidina-ácido acético (+) e as conformações onde ocorreu transferência de protão durante a optimização de um complexo duplo ião (⊕).

Tabela 4.10. Energia de estabilização ΔE e parâmetros inter-fragmento optimizados d e h, obtidos através de cálculos 6-31G** com modelos não simétricos, em função do ângulo de rotação α.

	ΔE(kJ/mol)		d(Å)		h(Å)	
α(°)	zwitter.	neut.	zwitter.	neut.	zwitter.	neut.
-170	-505,12	-537,12	3,86	4,12	0,47	0,53
-160+	-531,50	-531,50	4,18	4,18	0,92	0,92
-150+	-524,28	-524,28	4,28	4,28	1,31	1,31
-140	-455,83	-519,30	4,15	5,00	-0,57	-0,98
-130	-455,24	-518,13	4,19	4,53	-0,23	-0,64
-120	-454,97	-517,74	4,20	4,53	0,09	0,60
-110	-457,67	-519,06	4,16	4,50	0,45	0,94
-100+	-519,35	-519,35	4,44	4,44	1,29	1,29
-90+	-518,92	-518,92	4,36	4,36	1,66	1,66
-80+	-517,83	-517,83	4,24	4,24	2,06	2,06
-70+	-515,90	-515,90	4,08	4,08	2,50	2,50
-60	-353,82	-512,80	5,08	3,80	0,04	2,95
-50	-333,04	-261,10	5,44	5,20	-1,14	-3,38
-40	-333,87	-266,88	5,55	5,39	-0,80	-2,81
-30	-334,41	-272,76	5,63	5,49	-0,33	-2,21
-20	-334,61	-279,22	5,67	5,49	0,16	-1,61
-10	-335,46	-287,28	5,59	5,35	0,61	-1,01
0+	-502,85	-502,85	4,22	4,22	-2,53	-2,53
10+	-499,57	-499,57	4,25	4,25	-2,05	-2,05
20+	-506,67	-506,67	4,24	4,24	-1,61	-1,61
30+	-514,58	-514,58	4,17	4,17	-1,21	-1,21
40	-495,90	-522,10	3,85	4,10	-0,90	-0,80
50	-507,01	-524,94	3,88	4,08	-0,47	-0,37
60	-510,41	-526,09	3,88	4,13	-0,02	-0,46
opt*	-510,46	-532,21	3,88	4,10	-0,07	0,34
70	-505,09	-532,13	3,86	4,10	0,40	0,28
80+	-529,35	-529,35	4,13	4,10	0,70	0,70
90+	-519,10	-519,10	4,20	4,20	1,10	1,10
100+	-511,34	-511,34	4,46	4,46	-1,23	-1,23
110	-450,74	-510,96	4,18	4,51	-0,38	-0,88
120	-449,03	-509,40	4,21	4,54	-0,05	-0,54
130	-450,06	-510,26	4,20	4,52	0,26	0,70
140	-452,54	-511,24	4,15	4,48	0,60	1,05
150+	-518,34	-518,34	4,26	4,26	-1,23	-1,23
160+	-525,84	-525,84	4,16	4,16	-0,84	-0,84
170	-505,33	-530,92	3,86	4,11	-0,43	-0,45
180	-508,90	-536,72	3,87	4,11	0,00	0,14
opt**	-508,94	-537,89	3,87	4,11	-0,05	0,35

zwitter. : complexo ligante duplo ião ; neut. : complexo ligante neutro.

* Valor de α optimizado na região trans(H). Os parâmetros geométricos são apresentados na tabela 4.12.

**Valor de α optimizado na região trans(CH_3). Os parâmetros geométricos são apresentados na tabela 4.13.

+ Conformações onde ocorreu uma transferência de protão durante a optimização de uma geometria inicial do tipo duplo ião.

Os resultados obtidos com modelos não simétricos permitem confirmar que o complexo neutro é geralmente mais estável que o complexo duplo ião. Ocorrem excepções a esta tendência geral para as conformações, de α=-50° a α=-10°, onde os protões do grupo metilo são os únicos que podem ser transferidos para o acetato.

Os mínimos do espaço conformacional ocorrem nas regiões trans(H) e trans(CH_3), o que está de acordo com a análise conformacional efectuada neste trabalho com modelos simétricos [48,50], bem como com outros cálculos teóricos [43-47] e análises estatísticas de estruturas cristalográficas [37,41,43] anteriormente efectuadas para este sistema.

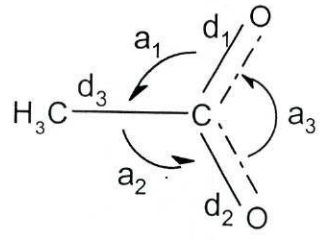
Nos resultados obtidos com modelos simétricos, foi detectada a transferência de protão do metilguanidínio para o acetato durante a optimização do complexo duplo ião. No entanto, este fenómeno foi somente observado num número limitado de conformações. Nos cálculos efectuados com modelos não simétricos, por serem possíveis deslocamentos horizontais durante o processo de optimização, o caminho para a transferência protónica é bastante mais fácil e realista. Consequentemente, a transferência de protão foi detectada num grande número de conformações (-160°, -150°, -100°, -90°, -80°, -70°, 0°, 10°, 20°, 30°, 80°, 90°, 100°, 150° e 160°). Estas conformações devem constituir canais privilegiados para a transferência de protão para os quais a energia de activação associada a este processo deve ser muito pequena.

As estruturas cristalográficas não fornecem geralmente qualquer tipo de informação relativamente às posições dos átomos de hidrogénio. Nestas circunstâncias, a descrição adequada das interacções Arg-Carb não pode ser efectuada se estes átomos forem restritos às cadeias laterais a que habitualmente pertencem. No entanto, a detecção de um complexo neutro numa interacção Arg-Carb pode ser efectuado tendo em conta as alterações induzidas pela transferência protónica nos parâmetros geométricos associados aos átomos pesados.

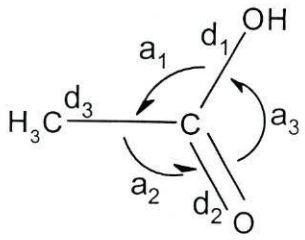
Neste trabalho, foi efectuado um estudo dos parâmetros geométricos relevantes para todos os fragmentos (metilguanidínio, acetato, metilguanidina e ácido acético) isolados e quando incluídos nos respectivos complexos ligantes.

Neste estudo, os parâmetros geométricos foram determinados usando orientações de referência dos fragmentos relativamente ao oxigénio aceitador no ácido acético; ao oxigénio que interactua mais proximamente com o metilguanidínio no acetato incluído no complexo duplo ião; ao grupo metilo no metilguanidínio isolado; ao protão removido na metilguanidina isolada; ao protão transferido na metilguanidina incluída no complexo neutro e ao hidrogénio que mais proximamente interactua com o acetato no metilguanidínio incluído no complexo duplo ião.

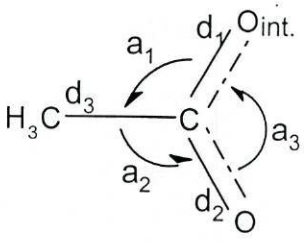
As orientações de referência e os parâmetros geométricos utilizados neste estudo são apresentados na figura (4.25).



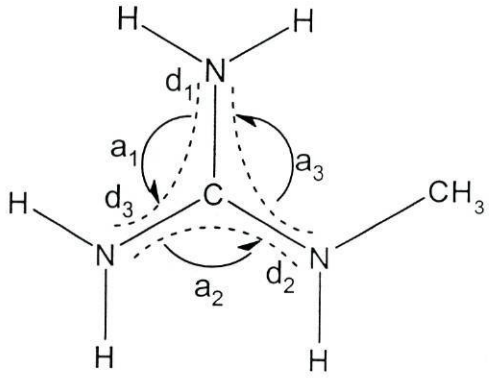
(A)



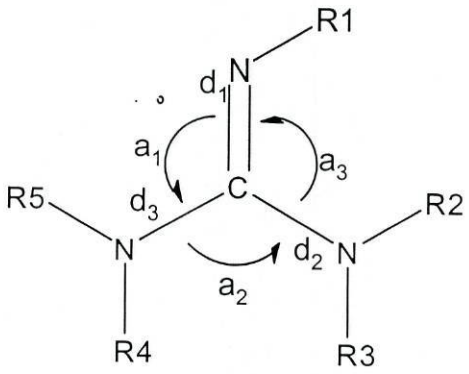
(B)



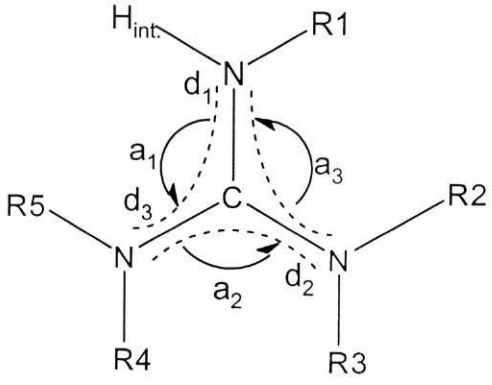
(C)



(D)



(E)



(F)

Figura 4.25. Orientações de referência utilizadas no estudo estatístico dos parâmetros geométricos (d_1, d_2, d_3, a_1, a_2 e a_3) associados à transferência protónica para o acetato isolado (A); o ácido acético (B); o acetato incorporado no complexo duplo ião (C); o metilguanidínio isolado (D); a metilguanidina (E) e o metilguanidínio incorporado no complexo duplo ião (F). Para os fragmentos iónicos incorporados no complexo duplo ião, na estrutura (C) $O_{int.}$ representa o átomo de oxigénio do acetato que interactua mais proximamente com o metilguanidínio e na estrutura (F) $H_{int.}$ representa o átomo de oxigénio do metilguanidínio que interactua mais proximamente com o acetato.

Para os fragmentos incluídos nos complexos ligantes, os parâmetros geométricos foram calculados como médias conformacionais no espaço estudado (α=-170° a α=180°). Para a metilguanidina isolada, estes valores foram obtidos como médias relativamente às estruturas possíveis (que diferem umas das outras pelo protão que foi removido). Para os outros fragmentos isolados, apenas existe uma estrutura possível e os parâmetros geométricos foram definidos de uma forma inequívoca. Os resultados obtidos são apresentados na tabela 4.11.

Tabela 4.11. Parâmetros geométricos[a,b] associados à transferência de protão, obtidos através de cálculos 6-31G**.

Fragmento	d_1	d_2	d_3	a_1	a_2	a_3
Ac^- isolado	1,24	1,24	1,55	115,25	115,25	129,50
HAc isolado	1,33	1,19	1,50	111,95	125,71	122,34
Ac^- na presença do MGH^+	1,2±0,8	1,2±0,5	1,5±0,1	116±2	118±2	126,1±0,6
HAc na presença do MG	1,3±0,5	1,2±0,3	1,5±0,9	112±1	124±1	124,1±0,2
MGH^+ isolado	1,32	1,32	1,33	119,28	120,06	120,66
MG isolado	1,26±0,01	1,38±0,01	1,38±0,01	120,2±0,7	112±1	127,0±0,7
MGH^+ na presença do Ac^-	1,30±0,01	1,34±0,01	1,33±0,01	120,4±0,9	119±1	120,9±0,9
MG na presença do HAc	1,27±0,01	1,38±0,01	1,37±0,01	120,6±0,7	114±1	125±1

Ac^-- acetato; HAc- ácido acético; MGH^+- metilguanidínio; MG- metilguanidina.

[a] Comprimentos de ligação em angstroms, ângulos em graus.

[b] Parâmetros geométricos definidos de acordo com as orientações de referência apresentadas na figura 4.25.

O grupo carboxilato do acetato isolado é um sistema conjugado π, onde as duas ligações C=O (d_1 e d_2) e os dois ângulos CH_3–C=O (a_1 e a_2) são equivalentes. No ácido

acético isolado, esta simetria é quebrada, pelo protão transferido para um átomo de oxigénio, dando origem a uma ligação simples C–OH (d_1) e a uma ligação dupla C=O (d_2) localizadas. Consequentemente, d_1 aumenta de 1.24 Å para 1.33 Å e d_2 diminui de 1.24 Å para 1.19 Å. Os ângulos de ligação associados CH_3–C–OH (a_1) e CH_3–C=O (a_2) deixam também de ser equivalentes. Por isso, enquanto a_1 diminui de 115.25° para 111.95°, a_2 aumenta de 115.25° para 125.71°. Os restantes ângulo O=C–O (a_3) e comprimento de ligação C–CH_3 (d_3) diminuem de 129.50° para 122.34° e de 1.55 Å para 1.50 Å, respectivamente.

O metilguanidínio isolado é também um sistema conjugado π onde as três ligações C=N (d_1, d_2 e d_3) e ângulos N=C=N (a_1, a_2 e a_3) são aproximadamente equivalentes. Na metilguanidina isolada, esta quase-simetria é quebrada devido à remoção de um protão de um átomo de azoto. Isto dá origem a duas ligações simples C–N (d_2 e d_3) e uma ligação dupla C=N, localizada. Consequentemente, enquanto d_1 diminui de 1.32 Å para 1.26 Å, d_2 e d_3 aumentam para 1.38 Å. Esta quase-simetria é também quebrada relativamente aos ângulos de ligação. Assim, enquanto a_1 é aproximadamente conservado a 120°, a_2 diminui de 120.06° para 112° e a_3 aumenta de 120.60° para 127.0°.

Quando os fragmentos são incluídos nos respectivos complexos ligantes, as tendências acima referidas são normalmente conservadas. No entanto, as interacções fragmento-fragmento originam diferenças menos acentuadas entre os parâmetros geométricos dos fragmentos neutros e iónicos.

Para os complexos duplo ião e neutro foram ainda efectuadas optimizações 6-31G** menos restritas nas regiões trans(H) e trans(CH_3). Nestes cálculos, foi permitido o relaxamento do ângulo α durante o processo de optimização de forma a encontrar o mínimo global para cada complexo dentro do modelo utilizado. Os resultados obtidos são apresentados nas tabelas 4.12 e 4.13.

As distâncias N--O, obtidas com estes modelos moleculares mais elaborados para o complexo duplo ião, são menores que os valores médios determinados através de análises estatísticas das estruturas cristalográficas [35,40]. As razões para estas discrepâncias são as mesmas que já anteriormente foram apresentadas na discussão dos resultados obtidos com os modelos simétricos.

Tabela 4.12. Parâmetros geométricos optimizados[a,b], através de cálculos 6-31G**, utilizando modelos não simétricos, na região trans(H).

Complexo	α	d	h	d(H_{21}-$N_{\eta 2}$)	d(H_{21}-$O_{\delta 1}$)	d(H_{31}-N_{ε})	d(H_{31}-$O_{\delta 2}$)	d($N_{\eta 2}$-$O_{\delta 1}$)	d(N_{ε}-$O_{\delta 2}$)
Duplo ião	58,90	3,88	-0,07	1,04	1,61	1,04	1,61	2,66	2,65
Neutro	71,39	4,10	0,34	1,78	0,98	1,00	2,01	2,76	3,01

[a] Distâncias em angstroms, ângulos em graus.
[b] Parâmetros geométricos definidos de acordo com as figuras 4.17, 4.19 e 4.21.

Tabela 4.13. Parâmetros geométricos optimizados[a,b], através de cálculos 6-31G**, utilizando modelos não simétricos, na região trans(CH_3).

Complexo	α	d	h	d(H_{21}-$N_{\eta 2}$)	d(H_{21}-$O_{\delta 1}$)	d(H_{31}-N_{ε})	d(H_{31}-$O_{\delta 2}$)	d($N_{\eta 2}$-$O_{\delta 1}$)	d(N_{ε}-$O_{\delta 2}$)
Duplo ião	178,92	3,87	-0,05	1,05	1,61	1,05	1,61	2,65	2,65
Neutro	-174,61	4,11	0,35	1,80	0,98	1,00	2,05	2,77	3,04

[a] Distâncias em angstroms, ângulos em graus.
[b] Parâmetros geométricos definidos de acordo com as figuras 4.17, 4.19 e 4.22.

Nos cálculos anteriormente apresentados para os complexos ligantes associados às interacções arginina-carboxilato, introduziram-se modelos moleculares apropriados (os modelos não simétricos metilguanidínio-acetato e metilguanidina-ácido acético), estabeleceu-se um nível adequado de tratamento quântico do problema (*ab initio* com uma base 6-31G**) e determinaram-se os respectivos mínimos conformacionais (trans(H) e trans(CH_3)).

Um problema que ainda não foi abordado foi o do mecanismo da reacção de transferência protónica. Neste contexto, foram então determinadas superfícies de energia potencial, associadas a esta reacção, em função da coordenada de reacção d(N-H) para ambos os mínimos conformacionais (ver figura 4.19). Estes cálculos foram efectuados, usando um formalismo *ab initio* com uma base 6-31G**, de acordo com a seguinte metodologia:

Para cada ponto da superfície de energia potencial, d(N-H) foi mantido constante enquanto os restantes parâmetros intra- e inter-fragmento (α, d e h) foram optimizados (ver figura 4.17). Os incrementos em d(N-H) foram de 0,1 Å nas regiões distantes do estado de transição (de d(N-H)=1,3 Å a d(N-H)=1,7 Å) e de 0,01 Å nas regiões próximas do estado de transição (de d(N-H)=1,1 Å a d(N-H)=1,3 Å).

Para ambas as superfícies de energia potencial, os estados de transição foram localizados explicitamente verificando a existência de um único valor próprio negativo da matriz das constantes de força a que corresponde uma frequência imaginária.

Todos os cálculos foram efectuados com o programa GAUSSIAN 92 [55] num computador IBM Risc 6000.

Os resultados obtidos são apresentados nas figuras (4.26) e (4.27) e nas tabelas 4.14 e 4.15.

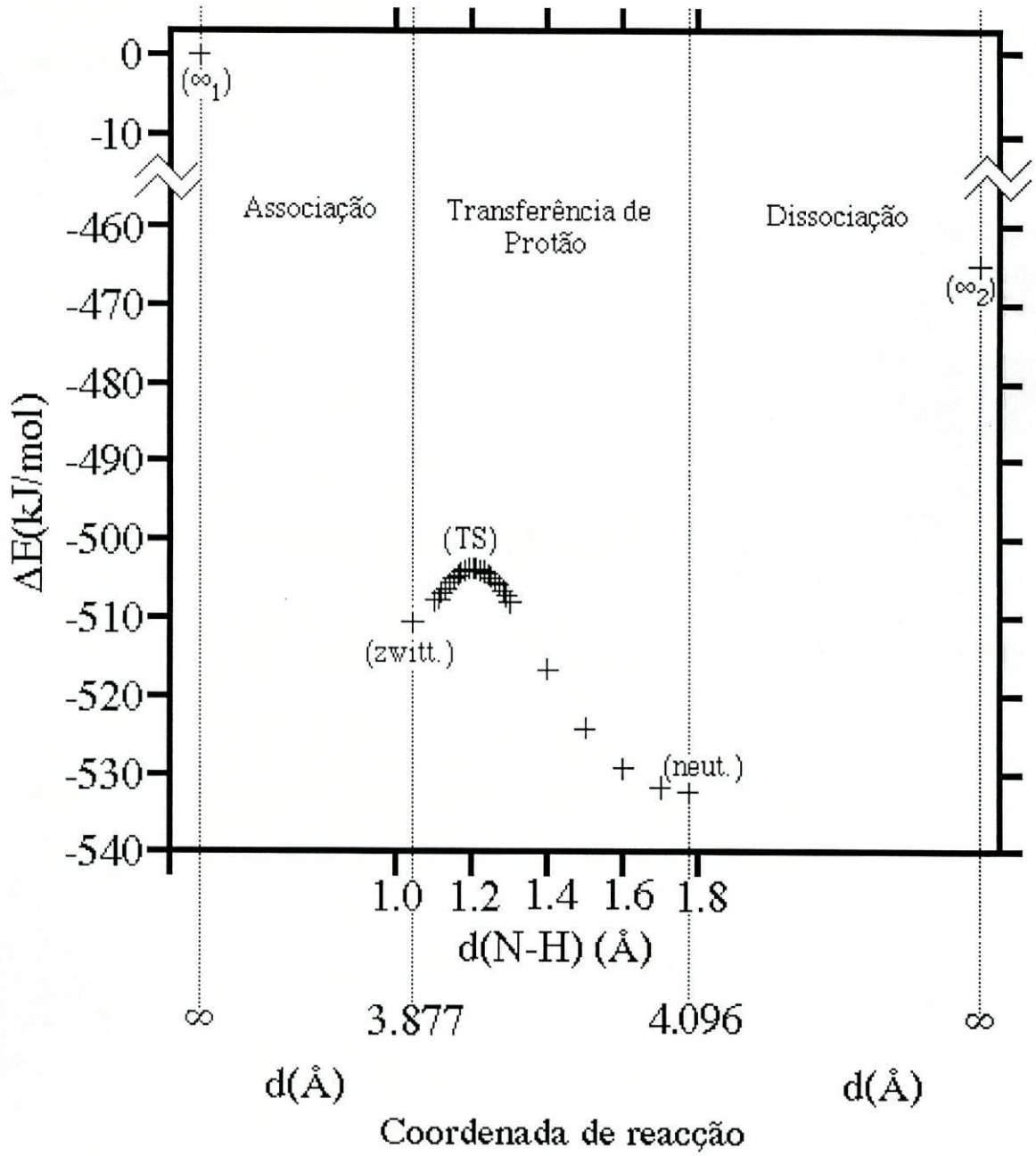


Figura 4.26. Superfície de energia potencial, obtida através de cálculos 6-31G**, para a reacção de transferência protónica do metiliguanidínio para o acetato na região trans(H). Na figura, ∞_1 representa o estado inicial não ligante constituído pelos fragmentos iónicos isolados, zwitt. o complexo ligante duplo ião, neut. o complexo ligante neutro, TS o estado de transição e ∞_2 o estado final não ligante constituído pelos fragmentos neutros isolados.

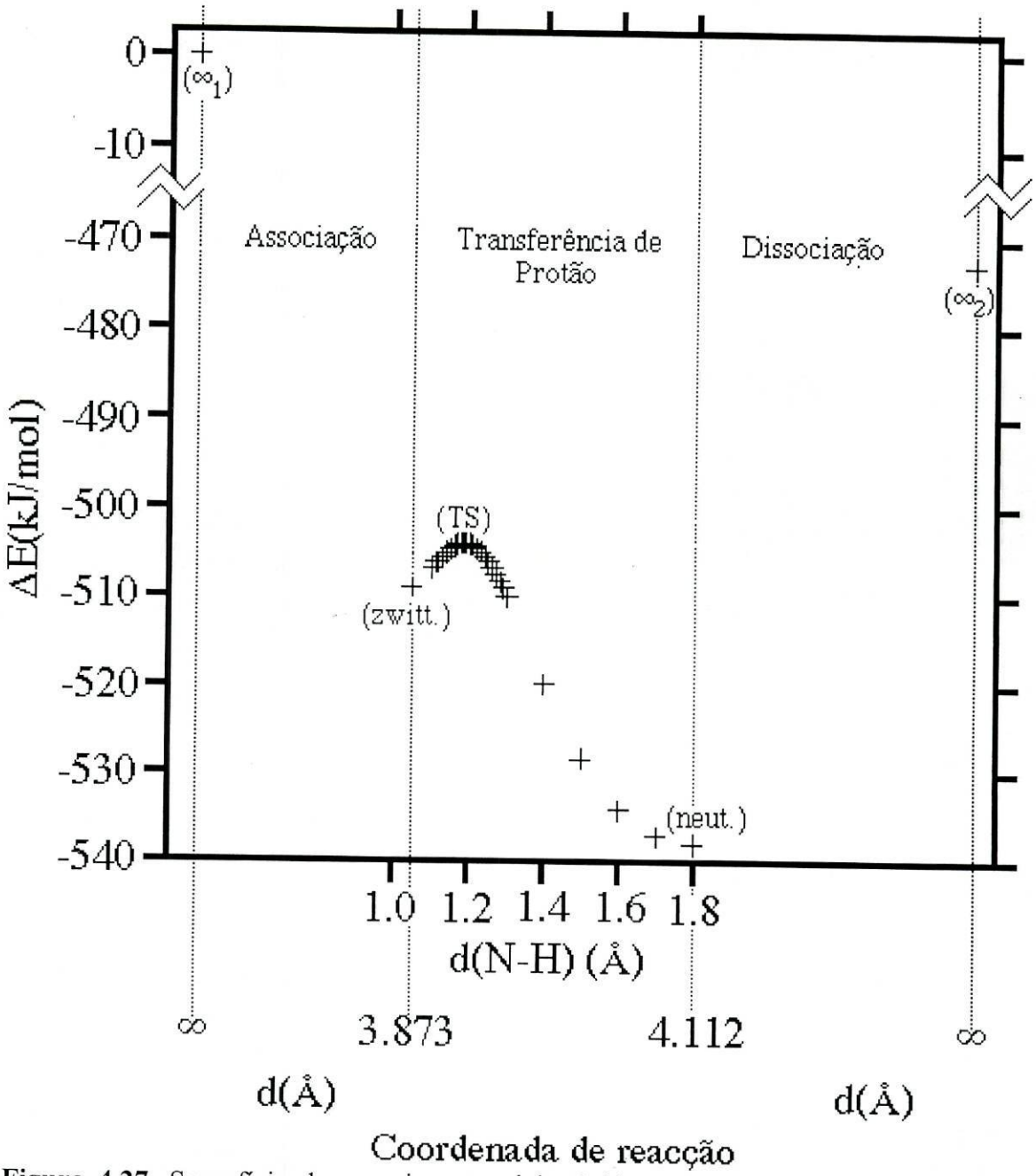


Figura 4.27. Superfície de energia potencial, obtida através de cálculos 6-31G**, para a reacção de transferência protónica do metiliguanidínio para o acetato na região trans(CH_3). Na figura, ∞_1 representa o estado inicial não ligante constituído pelos fragmentos iónicos isolados, zwitt. o complexo ligante duplo ião, neut. o complexo ligante neutro, TS o estado de transição e ∞_2 o estado final não ligante constituído pelos fragmentos neutros isolados.

Tabela 4.14. Energia de estabilização ΔE, obtida através de cálculos 6-31G**, em função da coordenada de reacção d(N-H) para a superfície de energia potencial associada à transferência protónica na região trans(H). Os parâmetros geométricos optimizados d(O-H), α, d e h são também apresentados.

d(N-H)[a] (Å)	d(O-H)[a] (Å)	ΔE(kJ/mol)	α(°)	d(Å)	h(Å)
1,042 (zwitt.)	1,606	-510,461	58,896	3,877	-0,071
1,100	1,487	-507,713	62,527	3,863	0,040
1,110	1,456	-507,441	64,895	3,856	0,169
1,120	1,433	-506,756	65,026	3,853	0,191
1,130	1,408	-506,089	66,264	3,850	0,211
1,140	1,387	-505,471	66,906	3,846	0,234
1,150	1,364	-504,915	67,545	3,843	0,249
1,160	1,344	-504,440	68,086	3,841	0,267
1,170	1,323	-504,054	68,676	3,838	0,284
1,180	1,303	-503,769	69,247	3,836	0,300
1,190	1,285	-503,591	69,795	3,834	0,315
1,200	1,266	-503,522	70,311	3,833	0,330
1,201 (TS)	1,264	-503,521	70,376	3,833	0,333
1,210	1,250	-503,560	70,791	3,832	0,346
1,220	1,240	-503,706	71,230	3,832	0,354
1,230	1,219	-503,952	71,607	3,831	0,367
1,240	1,205	-504,294	71,987	3,831	0,376
1,250	1,191	-504,724	72,352	3,832	0,382
1,260	1,179	-505,234	72,664	3,833	0,392
1,270	1,167	-505,812	72,991	3,835	0,404
1,280	1,156	-506,455	73,312	3,836	0,415
1,290	1,145	-507,156	73,481	3,839	0,414
1,300	1,136	-507,901	73,680	3,841	0,418
1,400(TP)	1,067	-516,443	74,684	3,878	0,438
1,500	1,028	-524,062	74,536	3,929	0,433
1,600	1,003	-529,140	73,660	3,987	0,407
1,700	0,988	-531,686	74,469	4,048	0,373
1,776 (neut.)	0,980	-532,213	71,391	4,096	0,339

zwitt.- complexo ligante duplo ião ;TS- estado de transição; TP- ponto particular da superfície de energia potencial onde ocorre a transferência de protão; neut.- complexo ligante neutro.
[a] Parâmetros geométricos definidos de acordo com a figura 4.19.

Tabela 4.15. Energia de estabilização ΔE, obtida através de cálculos 6-31G**, em função da coordenada de reacção d(N-H) para a superfície de energia potencial associada à transferência protónica na região trans(CH_3). Os parâmetros geométricos optimizados d(O-H), α, d e h são também apresentados.

d(N-H)[a] (Å)	d(O-H)[a] (Å)	ΔE(kJ/mol)	α(°)	d(Å)	h(Å)
1,045 (zwitt.)	1,605	-508,945	178,920	3,873	-0,046
1,100	1,472	-506,778	-178,212	3,860	0,067
1,110	1,452	-506,131	-177,682	3,857	0,089
1,120	1,423	-506,084	-175,651	3,849	0,220
1,130	1,399	-505,555	-175,093	3,846	0,240
1,140	1,378	-505,084	-174,645	3,843	0,260
1,150	1,357	-504,676	-174,184	3,839	0,277
1,160	1,334	-504,352	-173,690	3,837	0,300
1,170	1,315	-504,125	-173,284	3,834	0,315
1,180	1,295	-503,997	-172,892	3,832	0,330
1,186 (TS)	1,278	-503,978	-172,588	3,831	0,342
1,190	1,277	-503,986	-172,470	3,831	0,346
1,200	1,259	-504,081	-172,112	3,829	0,361
1,210	1,241	-504,285	-171,777	3,828	0,373
1,220	1,227	-504,595	-171,472	3,827	0,384
1,230	1,213	-505,000	-171,231	3,828	0,396
1,240	1,198	-505,508	-170,955	3,828	0,404
1,250	1,185	-506,094	-170,789	3,828	0,414
1,260	1,172	-506,762	-170,553	3,830	0,421
1,270	1,161	-507,496	-170,352	3,831	0,427
1,280	1,151	-508,290	-170,218	3,833	0,433
1,290	1,140	-509,130	-170,024	3,835	0,439
1,300	1,131	-510,021	-169,979	3,838	0,442
1,400(TP)	1,064	-519,765	-169,674	3,876	0,463
1,500	1,026	-528,261	-170,315	3,928	0,458
1,600	1,001	-533,986	-171,404	3,987	0,440
1,700	0,987	-537,024	-172,926	4,050	0,397
1,798 (neut.)	0,976	-537,890	-174,608	4,112	0,349

zwitt.- complexo ligante duplo ião ;TS- estado de transição; TP- ponto particular da superfície de energia potencial onde ocorre a transferência de protão; neut.- complexo ligante neutro.
[a] Parâmetros geométricos definidos de acordo com a figura 4.19.

Através da análise dos resultados obtidos, pode constatar-se que as barreiras de energia potencial associadas à reacção de transferência protónica são muito pequenas (6,94 kJ/mol e 4,97 kJ/mol para as regiões trans(H) e trans(CH_3), respectivamente). Estes valores são significativamente inferiores aos determinados por Zheng e Ornstein [56] (de 14,6 kJ/mol a 33,1 kJ/mol) usando um modelo guanidínio-formato simétrico (h=0) e diversos métodos *ab initio*. A principal razão desta discrepância resulta provavelmente da flexibilidade introduzida

no presente modelo não simétrico que permite um caminho muito mais fácil e realista para a transferência protónica.

Analisando a tabela 4.14, pode verificar-se que na região trans(H) a reacção de transferência protónica implica a transição entre o estado inicial complexo duplo ião (α=58,896°, d=3,877 Å, h=-0,071 Å, d(N-H)=1,042 Å e d(O-H)=1,606 Å) e o estado final complexo neutro (α=71,391°, d=4,096 Å, h=0,339 Å, d(N-H)=1,776 Å e d(O-H)=0,980 Å). Analisando a tabelas 4.15, pode verificar-se que na região trans(CH_3) a mesma reacção envolve a transição entre o estado inicial complexo duplo ião (α = 178,920°, d=3,873 Å, h= -0,046 Å, d(N-H) = 1,045 Å e d(O-H) = 1,605 Å) e o estado final complexo neutro (α=-174,391°, d=4,112 Å, h=0,349 Å, d(N-H)=1,798 Å e d(O-H)=0,976 Å).

Analisando estes resultados com o auxílio da figura (4.17), pode constatar-se que a transição do estado inicial para o estado final nestas superfícies de energia potencial implica uma rotação de α no sentido anti-horário, um aumento das distâncias vertical (d) e horizontal (h) e a transferência do protão de um átomo de azoto do metilguanidínio para um átomo de oxigénio do acetato. Se todas estas alterações ocorressem conjugadamente as barreiras de activação para a transferência protónica seriam bastante maiores. De facto, num mecanismo conjugado deste tipo o protão teria de percorrer uma grande distância, durante a transferência entre os referidos átomos de azoto e oxigénio, sem poder ser estabilizado significativamente por interacções com os fragmentos envolvidos. Por outro lado, a rigidez do modelo associado a este mecanismo conduziria a um aumento significativo das repulsões fragmento-fragmento.

No entanto, analisando a figura (4.24) e a tabela 4.10, podem encontrar-se canais privilegiados para a transferência protónica rodando o ângulo α no sentido anti-horário a partir dos valores característicos do complexo neutro. Estes canais ocorrem para valores α próximos de 80° na região trans(H) e de -160° na região trans(CH_3). Nestas conformações particulares, o parâmetro h apresenta valores maiores relativamente aos característicos dos complexos ligantes nas suas geometrias optimizadas.

Nestas circunstâncias, é possível propor para a reacção de transferência protónica, em estudo, um mecanismo fundamentalmente diverso que pode ser ilustrado pelas figuras (4.28) a (4.31).

Ao longo de todo o caminho de reacção, a distância d(N-H) diminui e a distância d(O-H) aumenta o que reflecte o processo de transferência de protão entre os dois átomos pesados envolvidos (ver figuras 4.28 e 4.30).

Inicialmente, os dois fragmentos aproximam-se um do outro (ver figuras 4.28 e 4.30) até que na região do estado de transição (d(N-H)≈1,20 Å) as distâncias d(N-H) e d(O-H) são praticamente iguais. Isto permite a estabilização do estado de transição devido às interacções do protão com ambos os fragmentos envolvidos no processo. Simultaneamente, o ângulo α roda no sentido anti-horário e o parâmetro h aumenta (ver figuras 4.29 e 4.31).

A partir do estado de transição, os dois fragmentos começam a afastar-se, acompanhando o protão o movimento do acetato (ver figuras 4.28 e 4.30) e, simultaneamente, o metilguanidínio continua a rodar no sentido anti-horário e a deslocar-se horizontalmente relativamente ao acetato (ver figuras 4.29 e 4.31) até atingir uma conformação favorável à transferência protónica (α = 74,684°, d = 3,878 Å, h = 0,438 Å, d(N-H) = 1,400 Å e d(O-H)=1,067 Å, na região trans(H)) e (α= -169,674°, d = 3,876 Å, h = 0,463 Å, d(N-H)=1,400 Å e d(O-H)=1,064 Å, na região trans(CH_3)). Neste ponto da superfície de energia potencial, dá-se a transferência de protão do metilguanidínio para o acetato . De facto, a população electrónica de sobreposição entre o azoto do primeiro fragmento e o protão tende para zero e este encontra-se somente ligado ao oxigénio do segundo fragmento.

A partir deste ponto, o complexo neutro já se encontra formado pelo que todas alterações posteriores correspondem ao relaxamento da sua estrutura. Isto implica a rotação de α no sentido horário, a diminuição de h e o aumento de d (ver figuras 4.28 a 4.31).

O mecanismo proposto para a reacção de transferência de protão em estudo pode ser facilmente visualizado nas figuras (4.32) e (4.33).

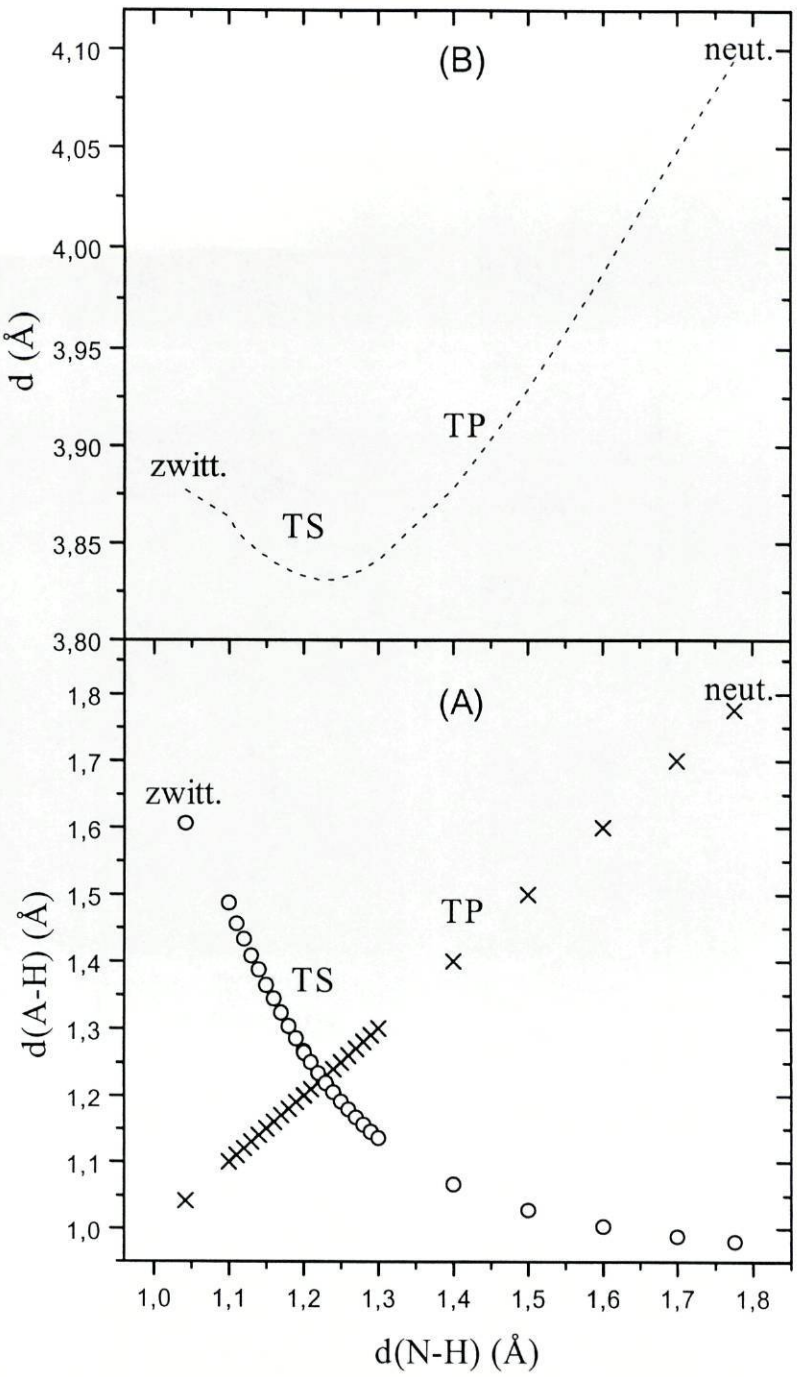


Figura 4.28. Gráficos (A) das distâncias d(N-H) (x) e d(O-H) (o) e (B) do parâmetro inter-fragmento d (---) em função da coordenada de reacção d(N-H), obtidos para a superfície de energia potencial 6-31G** da reacção de transferência protónica do metiliguanidínio para o acetato na região trans(H). Na figura, d(A-H) representa a distância entre o protão transferido e um dos átomos pesados (N ou O) envolvidos nesta reacção, zwitt. o complexo ligante duplo ião, TS o estado de transição, TP o ponto particular da superfície de energia potencial onde ocorre a transferência de protão e neut. o complexo ligante neutro.

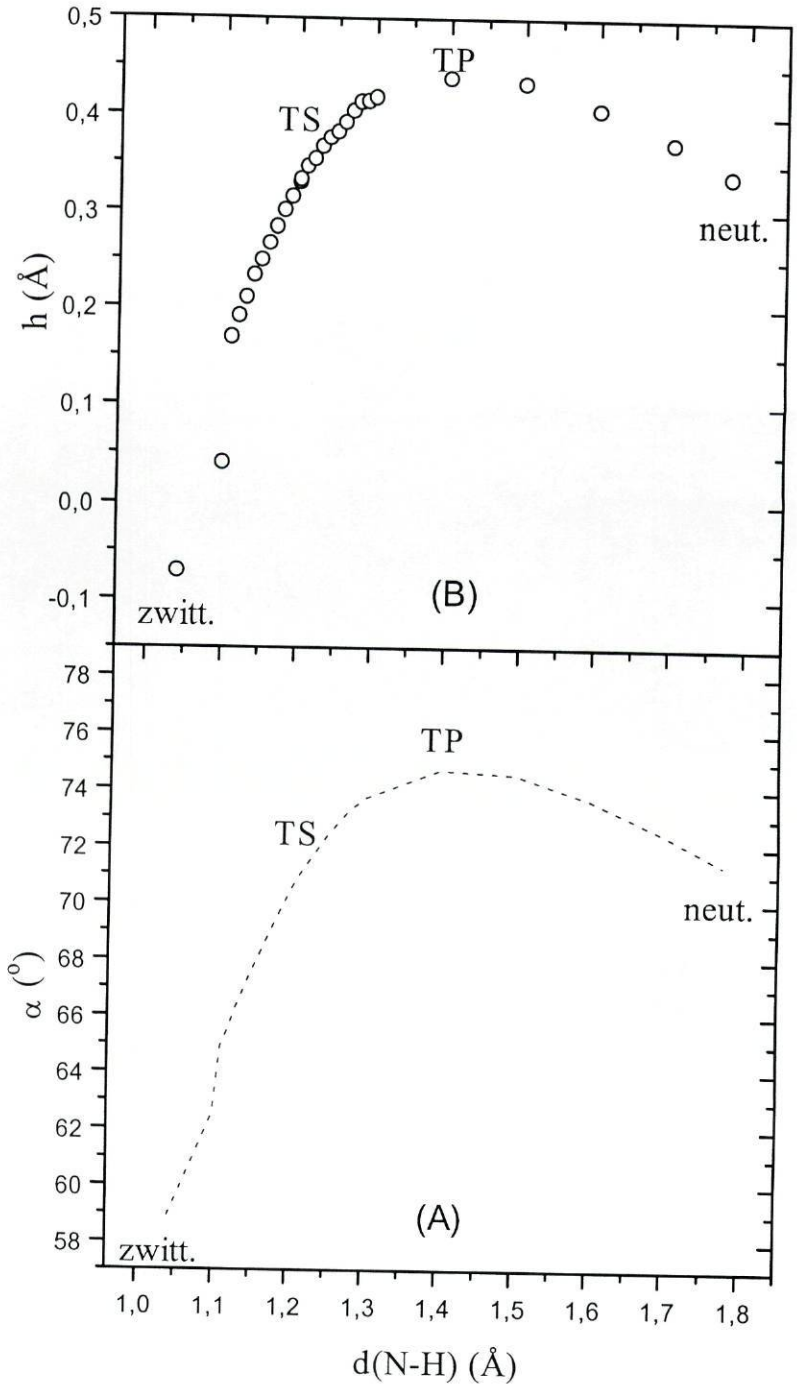


Figura 4.29. Gráficos (A) do ângulo de rotação α(---) e (B) do parâmetro inter-fragmento h (o) em função da coordenada de reacção d(N-H), obtidos para a superfície de energia potencial 6-31G** da reacção de transferência protónica do metiliguanidínio para o acetato na região trans(H). Na figura, zwitt. representa o complexo ligante duplo ião, TS o estado de transição, TP o ponto particular da superfície de energia potencial onde ocorre a transferência de protão e neut. o complexo ligante neutro.

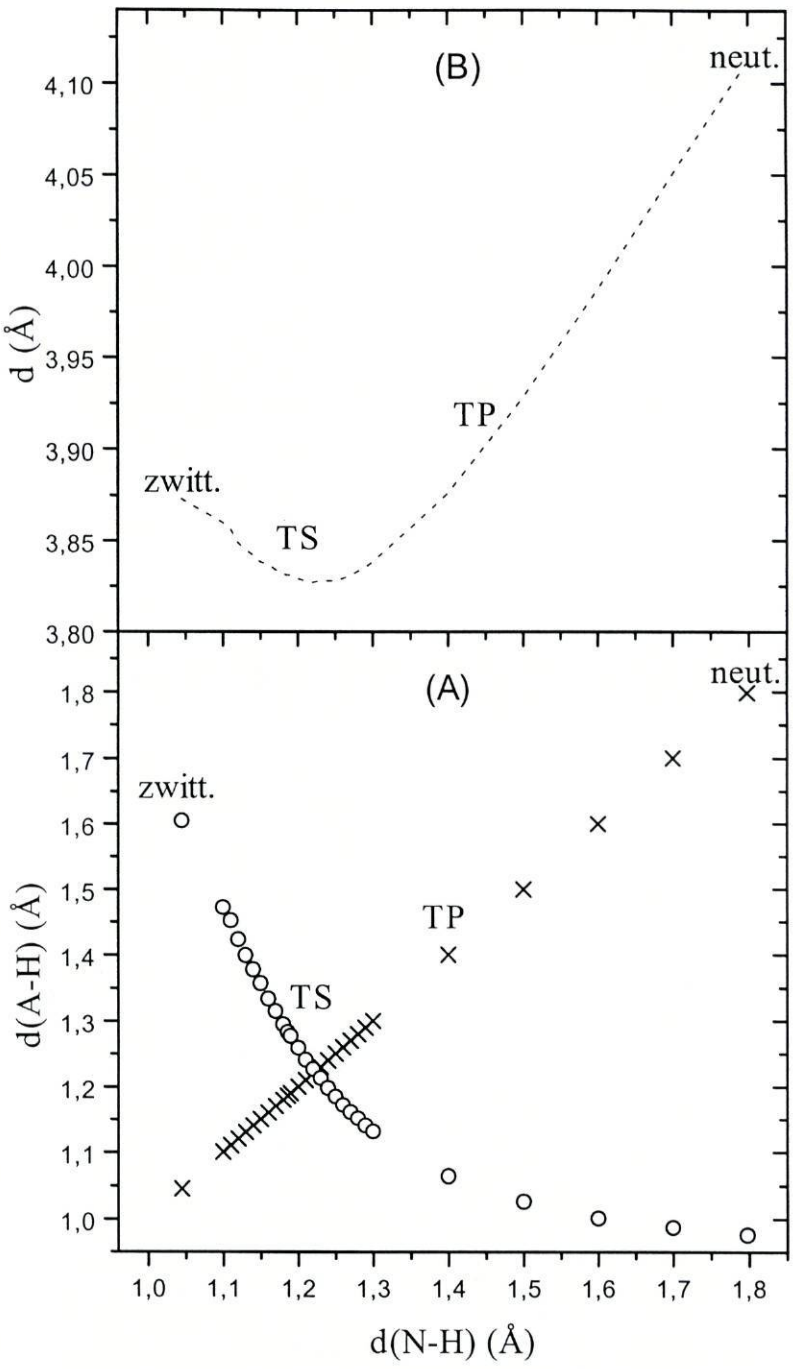


Figura 4.30. Gráficos (A) das distâncias d(N-H) (x) e d(O-H) (o) e (B) do parâmetro inter-fragmento d (---) em função da coordenada de reacção d(N-H), obtidos para a superfície de energia potencial da reacção de transferência protónica do metiliguanidínio para o acetato na região trans(CH_3). Na figura, d(A-H) representa a distância entre o protão transferido e um dos átomos pesados (N ou O) envolvidos nesta reacção, zwitt. o complexo ligante duplo ião, TS o estado de transição, TP o ponto particular da superfície de energia potencial onde ocorre a transferência de protão e neut. o complexo ligante neutro.

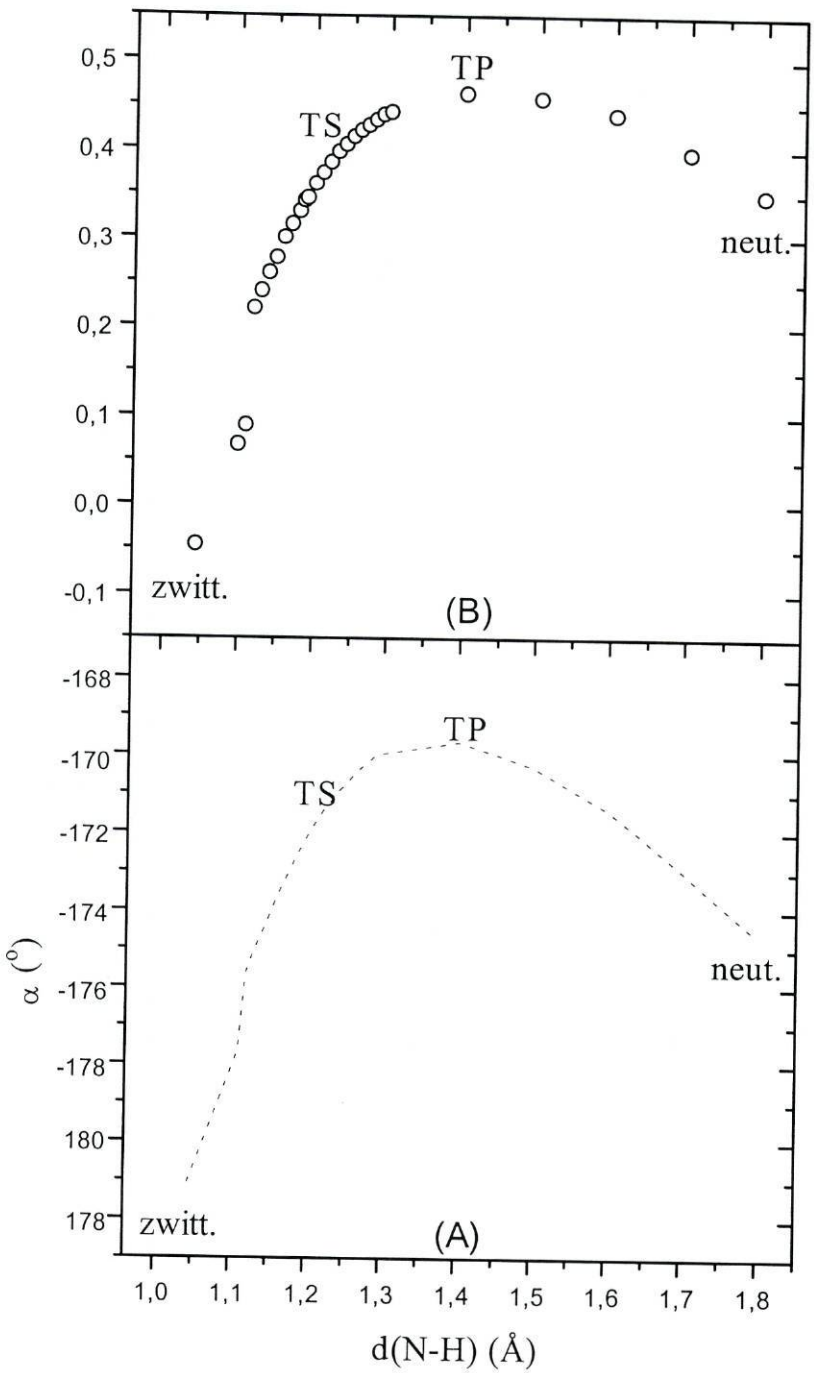


Figura 4.31. Gráficos (A) do ângulo de rotação α(---) e (B) do parâmetro inter-fragmento h (o) em função da coordenada de reacção d(N-H), obtidos para a superfície de energia potencial 6-31G** da reacção de transferência protónica do metiliguanidínio para o acetato na região trans(CH_3). Na figura, zwitt. representa o complexo ligante duplo ião, TS o estado de transição, TP o ponto particular da superfície de energia potencial onde ocorre a transferência de protão e neut. o complexo ligante neutro.

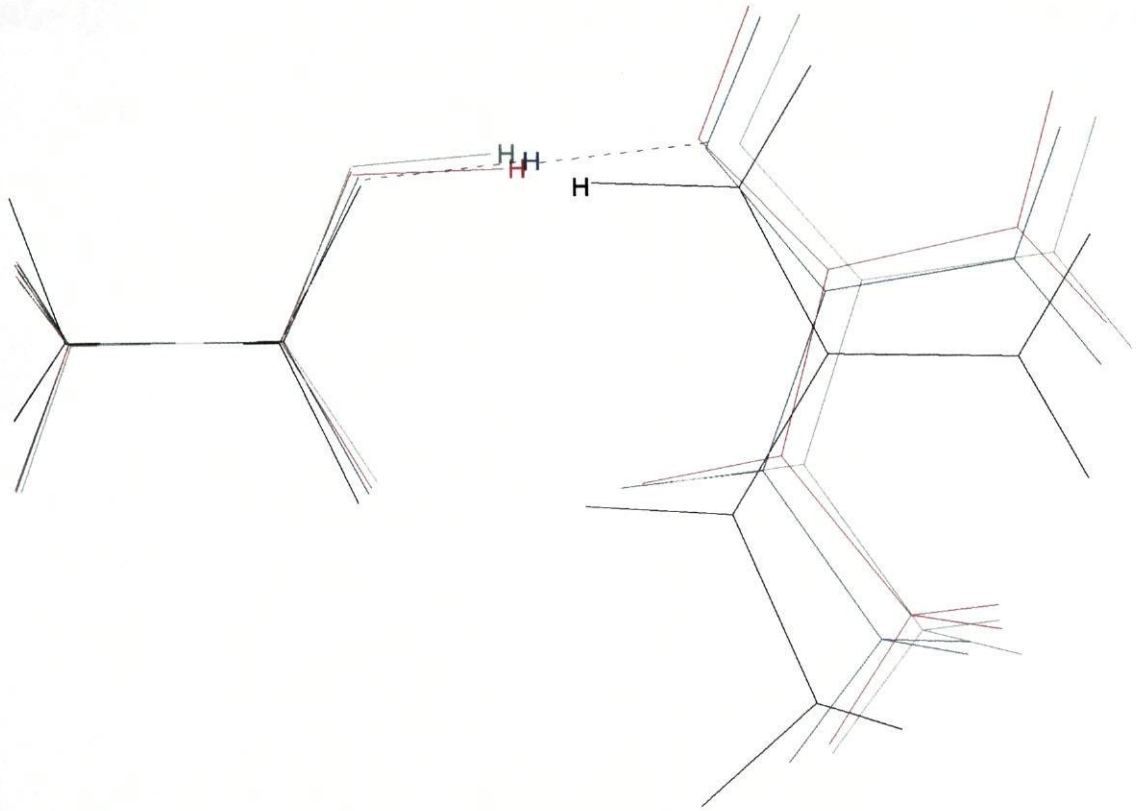


Figura 4.32. Modelo molecular para a reacção de transferência protónica do metiliguanidínio para o acetato na região trans(H). Na figura são apresentadas as estruturas relevantes da respectiva superfície de energia potencial obtida através de cálculos 6-31G*: a preto o complexo ligante duplo ião, a azul o estado de transição, a vermelho o ponto particular da superfície de energia potencial onde ocorre a transferência de protão e a verde o complexo ligante neutro.

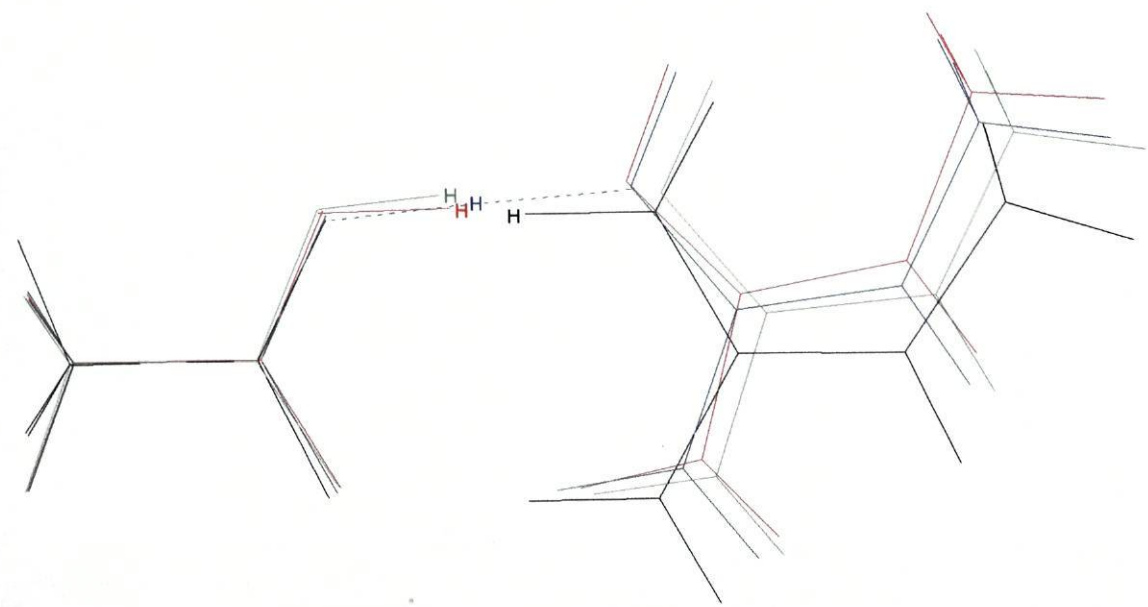


Figura 4.33. Modelo molecular para a reacção de transferência protónica do metiliguanidínio para o acetato na região trans(CH_3). Na figura são apresentadas as estruturas relevantes da respectiva superfície de energia potencial obtida através de cálculos 6-31G*: a preto complexo ligante duplo ião, a azul o estado de transição, a vermelho o ponto particular da superfície de energia potencial onde ocorre a transferência de protão e a verde o complexo ligante neutro.

Como já foi referido, embora as interacções Arg-Carb tenham sido estudadas extensivamente quer por experimentalistas [37-43] quer por teóricos [33,43-50,56], não foi apresentada nenhuma explicação sistemática dos efeitos responsáveis pela estabilização preferencial de cada um dos complexos ligantes.

Para permitir uma interpretação molecular da natureza destas interacções, o método de partição da energia de estabilização de Melo e Ramos, descrito na secção 3.5.7, foi aplicado aos mínimos relevantes das superfícies de energia potencial (4.26) e (4.27) [57].

Estas superfícies de energia dependem bastante das bases de funções utilizadas nos cálculos *ab initio* [58-66]. Para avaliar a importância deste facto na estabilidade relativa dos complexos ligantes duplo ião e neutro, a metodologia de decomposição atrás referida foi aplicada usando cálculos *ab initio* com uma série de bases diferentes. Nestes cálculos, os complexos ligantes foram completamente optimizados sem qualquer restrição relativa à sua planaridade. Todos os cálculos foram efectuados com o programa GAUSSIAN 92 [55] e com um programa especialmente elaborado para efectuar a partição de acordo com o formalismo atrás referido num computador IBM Risc 600.

As energias do estado inicial não ligante de referência, obtidas por diferentes métodos quânticos, são apresentadas na tabela 4.16. Os resultados da partição da energia de estabilização efectuada são apresentados nas tabelas 4.17 e 4.18.

Tabela 4.16. Energias (u.a.) dos fragmentos iónicos e do estado inicial não ligante de referência.

Método	$E(MGH^+)$	$E(Ac^-)$	E^{∞_1}
AM1	-36,05	-34,54	-70,59
6-31G	-243,46	-227,12	-470,58
6-31G*	-243,56	-227,23	-470,79
6-31G**	-243,58	-227,23	-470,81
6-311G	-243,52	-227,19	-470,71
6-311G*	-243,61	-227,28	-470,89
6-311G**	-243,63	-227,29	-470,92
D95	-243,50	-227,17	-470,67
D95*	-243,59	-227,28	-470,87
D95**	-243,62	-227,29	-470,91
6-31+G	-243,47	-227,15	-470,62
6-31+G*	-243,56	-227,25	-470,81
6-31+G**	-243,58	-227,25	-470,83
MP2/6-31G**	-244,35	-227,86	-472,21

MGH^+- metilguanidínio; Ac^-acetato; ∞_1 - estado inicial não ligante de referência.

Tabela 4.17. Decomposição da energia de estabilização (kJ/mol) para o mínimo conformacional trans(H) ($\Delta E^{zwitt.}$, energia de estabilização do complexo ligante duplo ião; $\Delta E^{neut.}$, energia de estabilização do complexo ligante neutro; $\Delta\Delta E^{zwitt.\rightarrow neut.}$ energia de estabilização relativa; ΔE^{∞_2} energia de estabilização do estado final não ligante).

Base	$\Delta E^{zwitt.}$			$\Delta E^{neut.}$			$\Delta\Delta E^{zwitt.\rightarrow neut.}$			ΔE^{∞_2}	
	Total	Intra	Inter	Total	Intra	Inter	Total	Intra	Inter	Total	Intra
6-31G	-518,40	216,64	-735,04	-501,00	-225,93	-275,08	17,40	-442,57	459,96	-403,88	-403,88
6-31G*	-507,95	196,83	-704,79	-528,23	-265,37	-262,86	-20,28	-462,20	441,93	-461,23	-461,23
6-31G**	-510,47	123,76	-634,23	-532,15	-324,91	-207,24	-21,68	-448,67	426,99	-464,82	-464,82
6-311G	-512,26	45,30	-557,56	-497,28	-374,72	-122,57	14,98	-420,02	434,99	-403,94	-403,94
6-311G*	-499,38	68,14	-567,52	-517,67	-339,08	-178,59	-18,29	-407,22	388,93	-453,14	-453,14
6-311G**	-502,91	48,50	-551,41	-527,88	-337,43	-190,45	-24,97	-385,93	360,96	-463,42	-463,42
D95	-506,89	25,24	-532,13	-490,03	-475,96	-14,07	16,86	-501,20	518,06	-389,90	-389,90
D95*	-491,13	112,81	-603,94	-507,54	-372,30	-135,25	-16,41	-485,11	468,69	-451,44	-451,44
D95**	-495,36	-34,46	-460,90	-516,08	-492,11	-23,97	-20,72	-457,65	436,93	-452,98	-452,98
6-31+G	-491,85	196,80	-688,65	-470,14	-198,72	-271,42	21,71	-395,52	417,23	-381,67	-381,67
6-31+G*	-479,49	289,50	-768,99	-493,05	-185,83	-307,22	-13,56	-475,33	461,77	-431,25	-431,25
6-31+G**	-480,96	26,25	-507,21	-497,61	-448,61	-49,00	-16,65	-474,86	458,21	-435,89	-435,89

Tabela 4.18. Decomposição da energia de estabilização (kJ/mol) para o mínimo conformacional trans(CH_3) ($\Delta E^{zwitt.}$, energia de estabilização do complexo ligante duplo ião; $\Delta E^{neut.}$, energia de estabilização do complexo ligante neutro; $\Delta\Delta E^{zwitt.\rightarrow neut.}$ energia de estabilização relativa; ΔE^{∞_2} energia de estabilização do estado final não ligante).

Base	$\Delta E^{zwitt.}$			$\Delta E^{neut.}$			$\Delta\Delta E^{zwitt.\rightarrow neut.}$			ΔE^{∞_2}	
	Total	Intra	Inter	Total	Intra	Inter	Total	Intra	Inter	Total	Intra
6-31G	-516,03	205,61	-721,64	-502,89	-245,58	-257,31	13,14	-451,19	464,33	-408,55	-408,55
6-31G*	-506,28	193,94	-700,22	-534,56	-289,15	-245,42	-28,28	-483,09	454,80	-469,17	-469,17
6-31G**	-508,94	115,73	-624,67	-537,84	-348,17	-189,67	-28,90	-463,90	435,00	-472,78	-472,78
6-311G	-509,44	29,55	-538,98	-499,11	-387,38	-111,73	10,33	-416,93	427,25	-409,80	-409,80
6-311G*	-497,17	74,88	-572,04	-524,56	-356,53	-168,03	-27,39	-431,41	404,01	-461,51	-461,51
6-311G**	-501,53	44,17	-545,70	-534,44	-359,99	-174,46	-32,91	-404,16	371,24	-471,55	-471,55
D95	-503,76	15,57	-519,33	-491,29	-492,00	0,71	12,47	-507,57	520,04	-394,71	-394,71
D95*	-489,15	114,06	-603,22	-518,83	-407,95	-110,88	-29,68	-522,01	492,34	-458,60	-458,60
D95**	-493,65	-36,68	-456,97	-522,13	-504,79	-17,34	-28,48	-468,11	439,63	-460,39	-460,39
6-31+G	-488,76	148,61	-637,37	-472,60	-98,75	-373,84	16,16	-247,36	263,53	-386,90	-386,90
6-31+G*	-477,55	272,57	-750,12	-499,51	-152,15	-347,36	-21,96	-424,72	402,76	-439,75	-439,75
6-31+G**	-479,06	-42,21	-436,85	-503,80	-456,82	-46,98	-24,74	-414,61	389,87	-444,47	-444,47

As optimizações globais dos complexos ligantes originaram estruturas em que os átomos pesados não apresentam grandes desvios relativamente à planaridade. Isto confirma que os modelos planares utilizados nos cálculos anteriores eram apropriados.

As duas últimas colunas das tabelas 4.17 e 4.18 demonstram que, no estado final não ligante, onde os fragmentos neutros se encontram a distância infinita, a única componente da energia de estabilização do sistema tem de ser de natureza intra-fragmento.

Através da análise dos resultados contidos nestas tabelas, é possível constatar que as piores bases favorecem o complexo duplo ião. No entanto, a inclusão de funções de polarização inverte estes resultados favorecendo o complexo neutro. Isto reflecte-se no valor da energia de estabilização relativa ($\Delta\Delta E^{zwitt.\rightarrow neut.}$) que se torna consistentemente mais negativa à medida que se aumenta o número de funções de polarização para um determinado tipo de base.

No complexo duplo ião, a componente inter-fragmento da energia de estabilização é sempre claramente negativa enquanto a componente intra-fragmento é normalmente positiva. A estabilização global deste complexo ligante está pois associada à componente inter-fragmento que deriva da interacção electrostática bastante forte que se estabelece entre dois fragmentos de carga oposta.

No complexo neutro, a componente inter-fragmento da energia de estabilização é também habitualmente negativa embora inferior, em valor absoluto, aos valores característicos do complexo duplo ião. No entanto, a componente intra-fragmento contribui consistentemente para a estabilização global deste complexo ligante por apresentar valores sempre negativos. Isto indica que o acetato tem uma afinidade protónica intrínseca maior que o metilguanidínio.

Analisando as componentes da energia de estabilização relativa ($\Delta\Delta E^{zwitt.\rightarrow neut.}$), todos os cálculos indicam que a componente inter-fragmento favorece o complexo duplo ião enquanto o complexo neutro é favorecido pela maior afinidade protónica intrínseca do acetato relativamente ao metilguanidínio.

Referências

[1] C. F. Wong e J. A. McCammon, *J. Am. Che. Soc.*, **108**, 3830 (1986).

[2] P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge e P. A. Kollman, *Science*, **235**, 574 (1987).

[3] D. L. Beveridge e F. M. Dicapua, *Annu. Rev. Biophys. Biophys. Chem.*, **18**, 431 (1989).

[4] M. C. Menziani, C. A. Reynolds e W. G. Richards, *J. Chem. Soc. Chem. Commun.*, 853 (1989).

[5] L. X. Dang, K. M. Merz e P. A. Kollman, *J. Am. Chem. Soc.*, **111**, 8505 (1989).

[6] J. Gao, K. Kuczera, B. Tidor e M. Karplus, *Science*, **244**, 1069 (1989).

[7] K. Kuczera, J. Gao, B. Tidor e M. Karplus, *Proc. Natl. Acad. Sci. USA*, **87**, 8481 (1990).

[8] M. Prevost, S. J. Wodak, B. Tidor e M. Karplus, *Proc. Natl. Acad. Sci. USA*, **88**, 10880 (1991).

[9] B. Tidor e M. Karplus, *Biochemistry*, **30**, 3217 (1991).

[10] T. Simonson e A. T. Brünger, *Biochemistry*, **31**, 8661 (1992).

[11] B. Prod'hom e M. Karplus, *Protein Eng.*, **6**, 585 (1993).

[12] S. Yun-yu, A. E. Mark, W. Cun-xin, H. Fuhua, H. J. C. Berendsen e W. F. van Gunsteren, *Protein Eng.*, **6**, 289 (1993).

[13] M. Zacharias, T. P. Straatsma, J. A. McCammon e F.A. Qiocho, *Biochemistry*, **32**, 7428 (1993).

[14] F. T. K. Lau e M. Karplus, *J. Mol. Biol.*, **236**, 1049 (1994).

[15] P. L. Cummins e J. E. Gready, *J. Comp. Chem.*, **15**, 704 (1994).

[16] A. E. Mark e W. F. van Gunsteren, *J. Mol. Biol.*, **240**, 167 (1994).

[17] D. A. Pearlman e P. R. Connelly, *J. Mol. Biol.*, **248**, 696 (1995).

[18] G. P. Brady e K. A. Sharp, *J. Mol. Biol.*, **254**, 77 (1995).

[19] S. Boresh e M. Karplus. ., *J. Mol. Biol.*, **254**, 801 (1995).

[20] G. Liang, R. K. Schmidt, H. A. Yu, D. A. Cumming e J.W. Brady, *J. Phys. Chem.*, **100**, 2528 (1996).

[21] J. P. Kocher, M. Prevost, S. J. Wodak e B. Lee, *Structure*, **4**, 1517 (1996).

[22] T. Simonson, G. Archontis e M. Karplus, *J. Phys. Chem.*, B **101**, 8349 (1997).

[23] J. K. Gilson, J. A. Given, B. L. Bush e J. A. McCammon, *Biophys. J.*, **72**, 1047 (1997).

[24] T. Fox, T. S. Scanlan e P. A. Kollman, *J. Am. Chem. Soc.*, **119**, 11571 (1997).

[25] D. L. Veenstra e P. A. Kollman, *Protein Eng.*, **10**, 789 (1997).

[26] P. A. Kollman, *Pharm. Res.*, **15**, 368 (1998).

[27] G. Archontis, T. Simonson, D. Moras e M. Karplus, *J. Mol. Biol.*, **275**, 823 (1998).

[28] A. Hermone e K. Kuczera, *Biochemistry*, **37**, 2843 (1998).

[29] H. Kono, N. Nishiyama, M. Tokonura e J. Doi, *Protein Eng.*, **11**, 47 (1998).

[30] M. L. Lamb e W. L. Jorgensen, *Curr. Opin. Chem. Biol.*, **1**, 449 (1998).

[31]W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Inpey e M. L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).

[32] A. Melo e M. J. Ramos, *J. Peptide Res.*, **50**, 239 (1997).

[33] Maria João Ramos, André Melo, Elsa S. Henriques, José A. N. F. Gomes, Nathalie Reuter, Bernard Maigret, Wely B. Floriano e Marco A. C. Nascimento, *Int. J. Quantum Chem., submetido para publicação.*

[34] B. R. Brooks, R. F. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan e M. Karplus, *J. Comp. Chem.*, **4**, 187 (1983).

[35] A. D. MacKerell Jr., J. Wiórkiwicz-Kuczera e M. Karplus, *J. Am. Chem. Soc.*, **117**, 11946 (1995).

[36] President and Fellows of Harvard College, CHARMM (versão 2.2), Molecular Simulations Inc., Burlington, USA..

[37] J. Singh, J. M. Thornton, M. Snarey e S. F. Campbell, *FEBS Lett.*, **224**, 161 (1987).

[38] D. S. Eggleston e D. J. Hodgson, *Int. J. Peptide Protein Res.*, **25**, 242 (1985).

[39] D. M. Salunke e M. Vijayan, *Int. J. Peptide Protein Res.*, **18**, 348 (1981).

[40] C. H Görbitz, *Acta Cryst.*, **B45**, 390 (1989).

[41] J. A. Ippolito, R. S. Alexander e D. W. Christianson, *J. Mol. Biol.*, **215**, 457 (1990).

[42] P. Chakrabarti, *Int. J. Peptide Protein Res.*, **43**, 284 (1994).

[43] J. B. O Mitchell, J. M. Thornton e J. Singh, *J. Mol. Biol.*, **226**, 251 (1992).

[44] A. M. Sapse e C. S. Russell, *Int. J. Quantum Chem.*, **26**, 91 (1984).

[45] A. M. Sapse e C. S. Russell, *J. Mol. Struct. (Theochem)*, **137**, 43 (1986).

[46] D. W. Deerfield, H. B. Nicholas Jr., R. G. Hiskey e L. G. Pedersen, *Proteins Struct. Fun. Genetics*, **6**, 168 (1989).

[47] J. B. O. Mitchell, C. L. Nandi, J. M. Thornton, S. L. Price, J. Singh e M. Snarey, *J. Chem. Soc. Faraday Tans.*, **89**, 2619 (1993).

[48] A. Melo e M. J. Ramos, *Chem. Phys. Lett.*, **245**, 498 (1995).

[49]V. Thery, D. Rinaldi, J. L. Rivail, B. Maigret e G. G. Ferenczy, *J. Comp. Chem.*, **15**, 269 (1994).

[50] A. Melo, M. J. Ramos, W. B. Floriano, J. A. N. F. Gomes, J. F. R. Leão, A. L. Magalhães, B. Maigret, M. C. Nascimento e N. Reuter, *J. Mol. Struct. (Theochem), aceite para publicação.*

[51] C. L. A. Schmidt, P. L. Kirk e W. K. Appleman, *J. Biol. Chem.*, **88**, 285 (1930).

[52] G. D. Henry e B. D. Sykes, *J. Biomol. NMR*, 6, 59 (1995).

[53] M. J. S. Dewar, E. F. Healy, D. A. Liotard e J. M. Ruiz, AMPAC, QCPE, Programa no. 506, (1989).

[54] M. J. Frisch, M. Head-Gordon, G. W. Trucks, J. B. Foresman, H. B. Schlegel, K. Raghavachari, M. A. Robb, J. S. Binkley, C. Gonzalez, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. L. Martin, L. R. Kahn, , J. J. P. Stewart, S. Topiol e J. A. Pople, Gaussian 90, Revision A, Gaussian, Inc., Pittsburgh, PA, (1990).

[55] M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, e J. A. Pople, Gaussian 92, Revision A, Gaussian, Inc., Pittsburgh, PA, (1992).

[56] J. Zheng e R. L. Ornstein, *J. Am. Chem. Soc.*, **118**, 11237 (1996).

[57] A. Melo e M. J. Ramos, *Int. J. Quantum Chem., aceite para publicação.*

[58] J. P. Dijkman, R. Osman e M. Weinstein, *Int. J. Quantum Chem., Quantum Biol. Symp.*, **14**, 211 (1987).

[59] D. Hadzi, *J. Mol. Struct. (Theochem)*, **177**, 1 (1988).

[60] J. P. Dijkman, R. Osman e M. Weinstein, *Int. J. Quantum Chem.*, **35**, 241 (1989).

[61] S. Scheiner, *Acc. Chem. Res.*, **18**, 174 (1985).

[62] L. Pardo, A. P. Mazurek e R. Osman, *Int. J. Quantum Chem.*, **37**, 701 (1990).

[63] C. J: Cassaday, S. R. Carr, K. Zhang e A. Chung-Philiphs, *J. Org. Chem.*, **60**, 1704 (1995).

[64] V. Hrouda, J. Florián, M. Polásek e P. Hobza, *J. Phys. Chem.*, **98**, 4742 (1994).

[65] S. Scheiner, *J. Mol. Struct. (Theochem)*, **307**, 65 (1994).

[66] A.L. Sobolewsky e L. Adamowicz, *Chem. Phys. Lett.*, **234**, 94 (1995).

Conclusões

A Ciência no seu melhor não nos permite obter respostas absolutas, mas sim melhores perguntas.

Norman Cousins (1976)

O presente trabalho insere-se num projecto de estudo mais vasto relativo às interacções protease serina:inibidor e ao desenho de novos inibidores para esta classe de enzimas.

Neste trabalho efectuou-se o estudo das interacções ligantes responsáveis pela estabilidade do complexo tripsina:PTI utilizando mutações de resíduos específicos na respectiva região de contacto e métodos de simulação molecular apropriados. As variações da energia de Helmholtz (ΔΔA) obtidas através desta metodologia permitem concluir que uma interacção ligante é tanto mais forte quanto maiores forem a separação de carga entre os grupos de átomos nela envolvidos e a hidrofilicidade do meio em que ocorre. Neste contexto, a maioria das interacções que ocorrem na região de contacto do complexo tripsina:PTI são fracas pois são do tipo polar-polar e ocorrem na região adjacente da cavidade ligante S_1 que é bastante inacessível ao solvente. As interacções mais fortes são a Tyr^{39}-Cys^{14}, que é do tipo polar-polar mas ocorre num meio muito acessível ao solvente e claramente hidrofílico, e as interacções His^{40}-Arg^{17} e Asp^{189}-Lys^{15}, que embora ocorram em meios hidrofóbicos, envolvem uma (no primeiro caso) ou duas (no segundo caso) cadeias laterais carregadas.

Os resultados obtidos neste trabalho são ainda consistentes com a especificidade apresentada pela tripsina relativamente a inibidores com resíduos P_1 em que as cadeias laterais estão carregadas positivamente (lisina ou arginina) [1-2]. De facto, o valor de ΔΔA associado à mutação Asp^{189}→Ala é bastante superior aos associados às mutações utilizadas no estudo das restantes interacções.

No futuro, este estudo pode ser melhorado em diversas vertentes, nomeadamente:

Os modelos moleculares podem ser melhorados através da solvatação de todo o sistema proteico.

As simulações MD podem ser aperfeiçoadas utilizando *condições periódicas fronteira*, não impondo restrições aos movimentos atómicos, trabalhando em *ensembles* mais próximos das condições fisiológicas (por exemplo o *ensemble isobárico-isotérmico*), aumentando o tempo de amostragem, recorrendo a formas mais elaboradas da função energia potencial (por exemplo incluindo efeitos de polarização) e utilizando técnicas mais rigorosas de controle da temperatura e da pressão (como por exemplo o método de Nosé-Hoover). Estas melhorias implicam um aumento significativo dos recursos computacionais disponíveis. No entanto, este aumento pode ser minorado tendo em conta os resultados obtidos neste trabalho. De facto, o número de janelas pode ser significativamente reduzido, dentro do formalismo da integração

termodinâmica, recorrendo ao método de extrapolação de Brooks III que se provou ser bastante eficiente relativamente aos *pontos centrais* da mutação. No entanto, deve evitar-se o uso deste método nas extrapolação para os *pontos extremos* especialmente quando ocorre a aniquilação de átomos.

Os ficheiros de trajectórias produzidos nas simulações MD foram de dimensão apreciável e colocaram bastantes problemas de armazenamento. O algoritmo BS-VLC, proposto neste trabalho, apresenta uma eficiência bastante melhor na compressão destes ficheiros que os algoritmos conservativos clássicos. No entanto, este algoritmo pode ainda ser melhorado adoptando estratégias de compressão que tenham em conta, de uma forma mais explícita, a distribuição estatística associada às coordenadas diferenciais.

Numa segunda vertente, efectuou-se um estudo quântico das interacções arginina-carboxilato que desempenham um papel muito importante no mecanismo de inibição do factor β-XII a. Neste estudo, foram utilizados os modelos moleculares metilguanidínio-acetato e metilguanidina-ácido acético para simular os complexos ligantes duplo ião e neutro, respectivamente. Todos os cálculos foram efectuados no vazio, o que permite reproduzir de uma forma razoável o ambiente hidrofóbico da cavidade ligante S_1 do factor β-XII a.

Estudos conformacionais prévios, efectuados com modelos simétricos simplificados, permitiram estabelecer que o formalismo *ab initio* com uma base 6-31G** constitui o nível mais baixo de cálculo quântico que permite descrever adequadamente as interacções arginina-carboxilato. Neste contexto, o uso da correlação electrónica não introduziu alterações significativas relativamente aos resultados obtidos ao nível 6-31G**. Por isso, este formalismo foi utilizado nos estudos conformacionais posteriormente efectuados com modelos moleculares não simétricos mais realistas.

Com base nestes estudos conformacionais foi possível concluir que as conformações mais estáveis eram aquelas (trans(H) e trans(CH_3)) em que se estabeleciam duas interacções do tipo (N--H--C=O). Os resultados do presente estudo quântico permitiram ainda indicar que o complexo neutro é mais estável que o complexo duplo ião para dímeros arginina-carboxilato que ocorram em ambientes hidrofóbicos. Isto está em desacordo com a maioria dos cálculos teóricos e análises estatísticas de estruturas cristalográficas efectuados anteriormente para este sistema. No entanto, este resultado foi confirmado por outros cálculos teóricos efectuados para interacções arginina-carboxilato que ocorrem no peptído Gly-Arg-Glu-Gly [3], em tetrâmeros na superfície de contacto do complexo Fab-Lisozima [4], e em vários solventes

orgânicos [6]. Foi ainda possível detectar várias conformações vizinhas dos mínimos conformacionais onde ocorre a transferência de um protão entre o metilguanidínio e o carboxilato durante a optimização da geometria do complexo duplo ião. Estas conformações constituem canais privilegiados para a transferência protónica e parecem desempenhar um papel importante no mecanismo desta reacção.

As posições dos átomos de hidrogénio não pode ser normalmente observada nas estruturas obtidas por difracção de raios X. Neste contexto, foi proposta uma metodologia para detectar a ocorrência do complexo neutro arginina-carboxilato em estruturas cristalográficas, que se baseia na *quebra de simetria* induzida pela transferência protónica em determinados parâmetros associados a átomos pesados.

As superfícies de energia potencial, incluindo a localização explícita dos respectivos estados de transição, foram determinadas para ambos os mínimos conformacionais do dímero arginina-carboxilato dentro do formalismo *ab initio* 6-31G**. Os resultados obtidos permitiram estabelecer que as barreiras de activação para a transferência protónica são bastante pequenas. Com base na variação de alguns parâmetros geométricos inter-fragmento em função da coordenada da reacção foi ainda proposto um mecanismo para esta reacção. Este mecanismo baseia-se no relaxamento da estrutura do complexo duplo ião de forma a adquirir a geometria característica de um dos canais privilegiados para a transferência protónica e na diminuição da distância percorrida pelo protão envolvido neste processo.

No âmbito deste trabalho, foi ainda proposto um novo método de decomposição da energia baseado no formalismo dos integrais indistinguíveis e na partição espacial das matrizes densidade e densidade de par. De acordo com este método, as energias molecular e de estabilização são decompostas em componentes associadas aos fragmentos que constituem o sistema em estudo. Este esquema de partição é definido de uma forma completamente flexível, permitindo a troca de átomos entre os fragmentos, e é especialmente apropriado ao estudo de sistemas onde ocorre uma reacção de transferência protónica.

Este método de partição foi aplicado ao estudo das interacções arginina-carboxilato, usando cálculos *ab initio* e 12 bases de funções diferentes, com o objectivo de permitir uma interpretação dos principais efeitos responsáveis pela estabilização relativa dos dois complexos ligantes a elas associados. As melhores bases favorecem o complexo neutro, enquanto as piores bases favorecem o complexo duplo ião. Apesar destas discrepâncias, todos os cálculos indicam que o complexo duplo ião é favorecido pela componente inter-fragmento

da energia de estabilização, enquanto a estabilização preferencial do complexo neutro deriva da maior afinidade protónica intrínseca do acetato relativamente ao metilguanidínio.

O estudo quântico das interacções arginina-carboxilato apresenta bastante interesse, não só na compreensão do mecanismo de inibição factor β-XII a, mas também na interpretação de outros mecanismos biológicos importantes. Consequentemente, poderá ter interesse alargar o âmbito do presente trabalho incluindo no estudo destas interacções o ambiente proteico e o solvente. Isto pode ser efectuado utilizando métodos híbridos clássico/quânticos como por exemplo o LSCF [3]. O solvente pode ser tratado de uma forma discreta para a primeira esfera de solvatação do par interactuante e por um modelo contínuo para as restantes moléculas.

Neste contexto, terá bastante interesse alargar o formalismo de partição aqui proposto ao tratamento explícito das interacções com o meio (proteico e solvente, quer ele seja descrito quantica ou classicamente e independentemente de ele ser tratado por um modelo continuo ou discreto. A *correcção de counterpoise* pode ser introduzida sem dificuldade no método de decomposição proposto neste trabalho. No entanto, a interpretação das componentes resultantes não é absolutamente clara. Esta deverá também ser uma linha de investigação a ser desenvolvida no futuro com vista a alargar o âmbito de aplicação do presente formalismo.

Referências

[1] R. M. Stroud, *Sci. Amer.*, **231**, 74 (1974).

[2] J. Kaut, *Ann. Rev. Biochem.*, **46**, 331 (1977).

[3] V. Thery, D. Rinaldi, J. L. Rivail, B. Maigret e G. G. Ferenczy, *J. Comp. Chem.*, **15**, 269 (1994).

[4] A. L. Magalhães, B. Maigret e J. A. N. F. Gomes, *E.C.C.I. Computational Chemistry - F.E.C.S. Conference*, **330**, 419 (1995).

[5] J. Zheng e R. L. Ornstein, *J. Am. Chem. Soc.*, **118**, 11237 (1996).

Apêndice I

Variações de energia de Helmholtz obtidas nas simulações computacionais das mutações estudadas na região de contacto tripsina:PTI

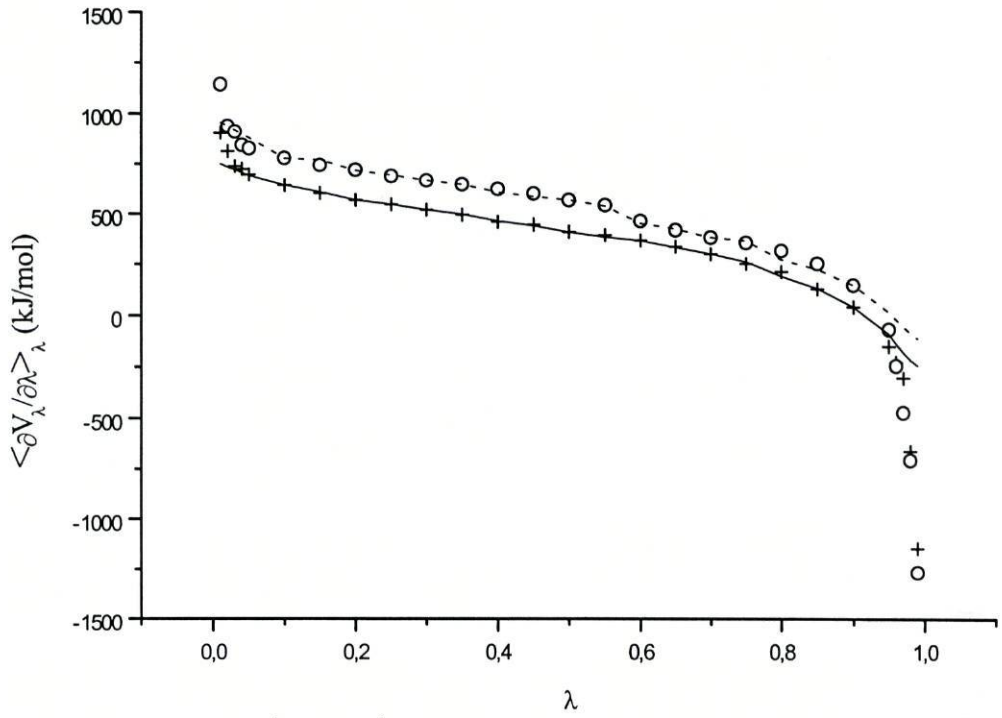


Figura I.1. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Asp^{189} \rightarrow Ala$, para (+) o processo 3 (Tripsina→MTripsina); (−) o processo 3 (Tripsina→MTripsina) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→MTripsina:PTI); (- -) o processo 4 (Tripsina:PTI→MTripsina:PTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

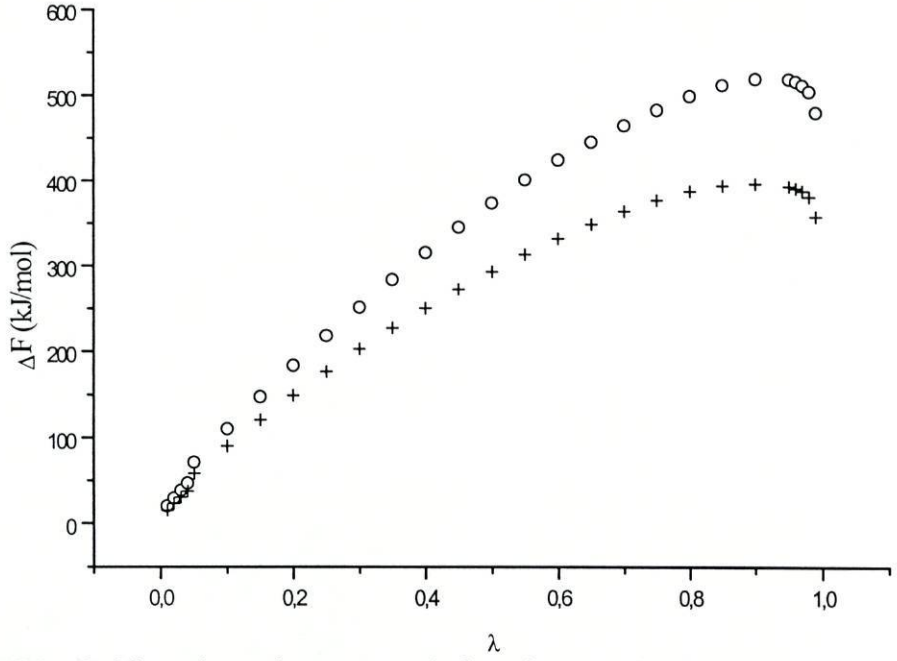


Figura I.2. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (Tripsina→MTripsina); (o) o processo 4 (Tripsina:PTI→MTripsina:PTI), na mutação $Asp^{189} \rightarrow Ala$.

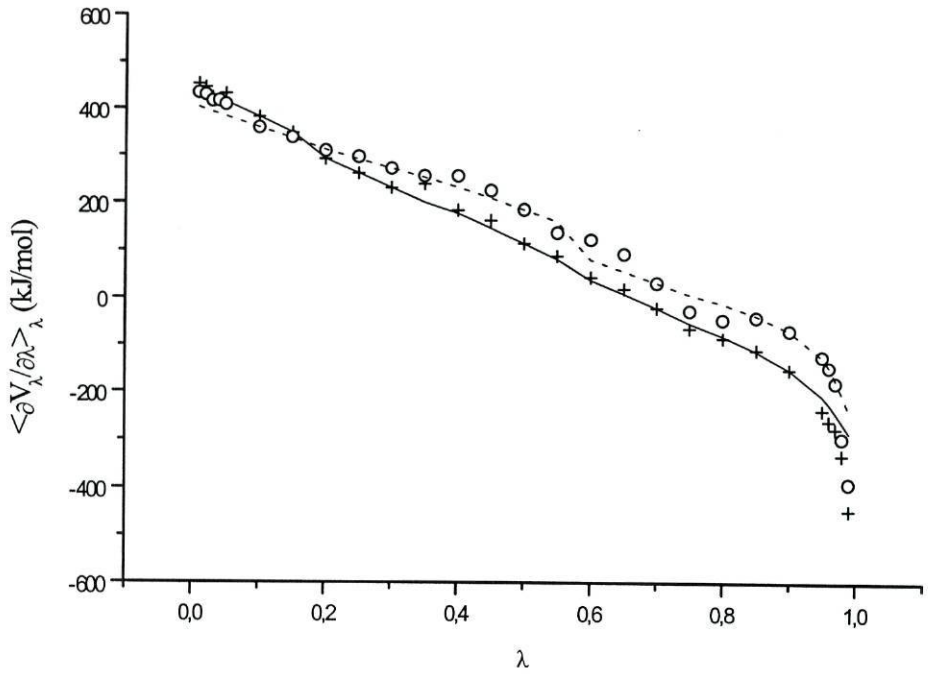


Figura I.3. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Arg^{17} \rightarrow Argn$, para (+) o processo 3 (PTI→MPTI); (-) o processo 3 (PTI→MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI); (- -) o processo 4 (Tripsina:PTI→Tripsina:MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

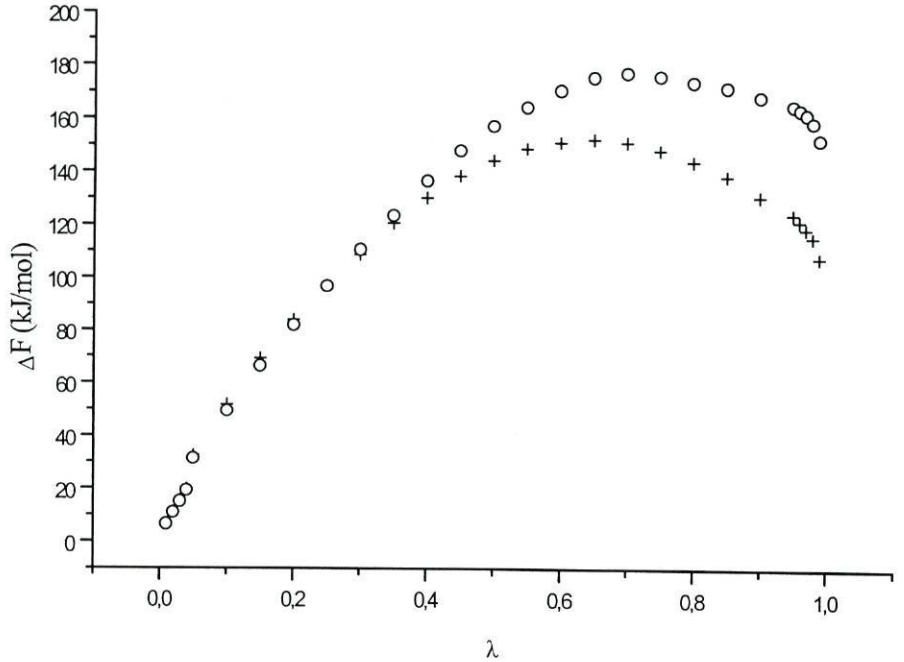


Figura I.4. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (PTI→MPTI); (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI), na mutação $Arg^{17} \rightarrow Argn$.

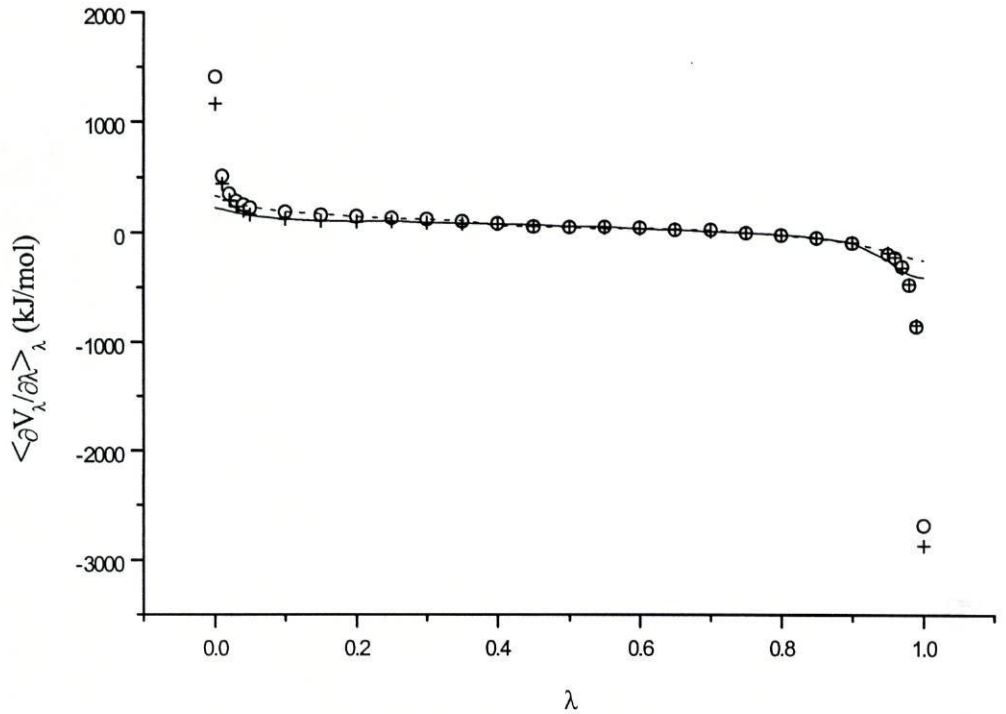


Figura I.5. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Tyr^{39} \rightarrow Phe$, para (+) o processo 3 (Tripsina→MTripsina); (−) o processo 3 (Tripsina→MTripsina) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→MTripsina:PTI); (- -) o processo 4 (Tripsina:PTI→MTripsina:PTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

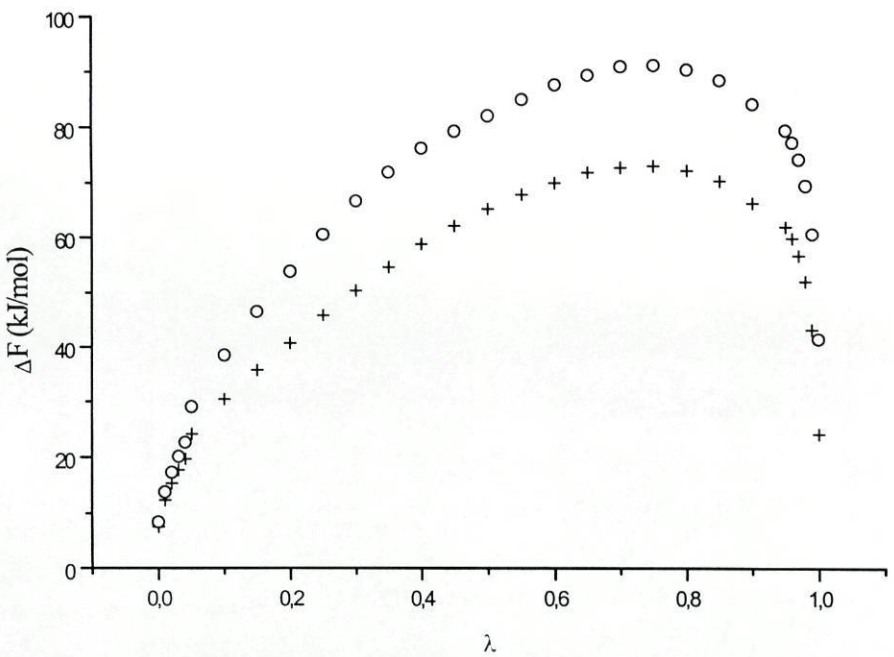


Figura I.6. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (Tripsina→MTripsina); (o) o processo 4 (Tripsina:PTI→MTripsina:PTI), na mutação $Tyr^{39} \rightarrow Phe$.

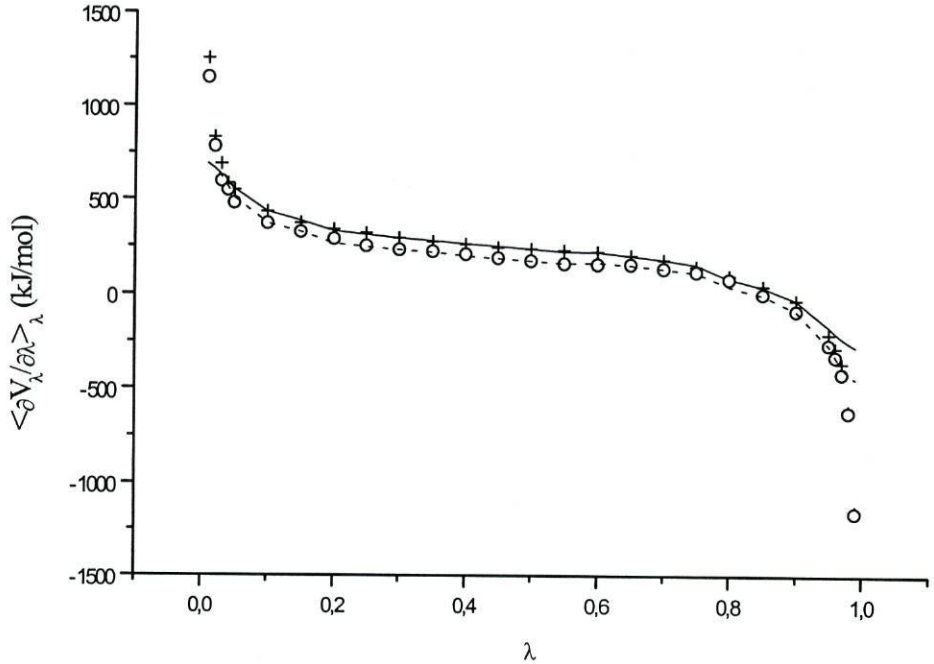


Figura I.7. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Gln^{192} \rightarrow Leu$, para (+) o processo 3 (Tripsina→MTripsina); (−) o processo 3 (Tripsina→MTripsina) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→MTripsina:PTI); (- -) o processo 4 (Tripsina:PTI→MTripsina:PTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

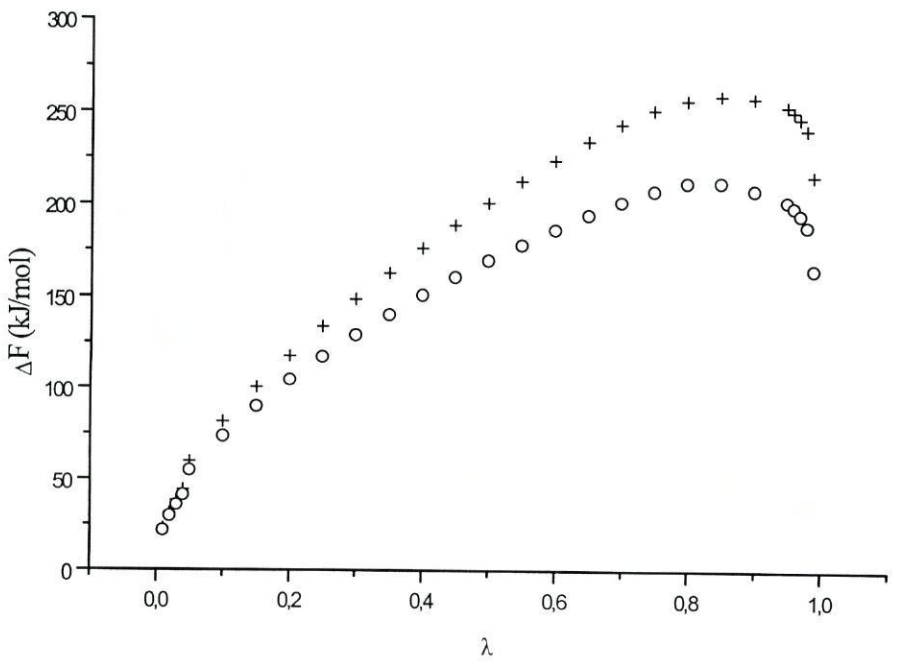


Figura I.8. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (Tripsina→MTripsina); (o) o processo 4 (Tripsina:PTI→MTripsina:PTI), na mutação $Gln^{192} \rightarrow Leu$.

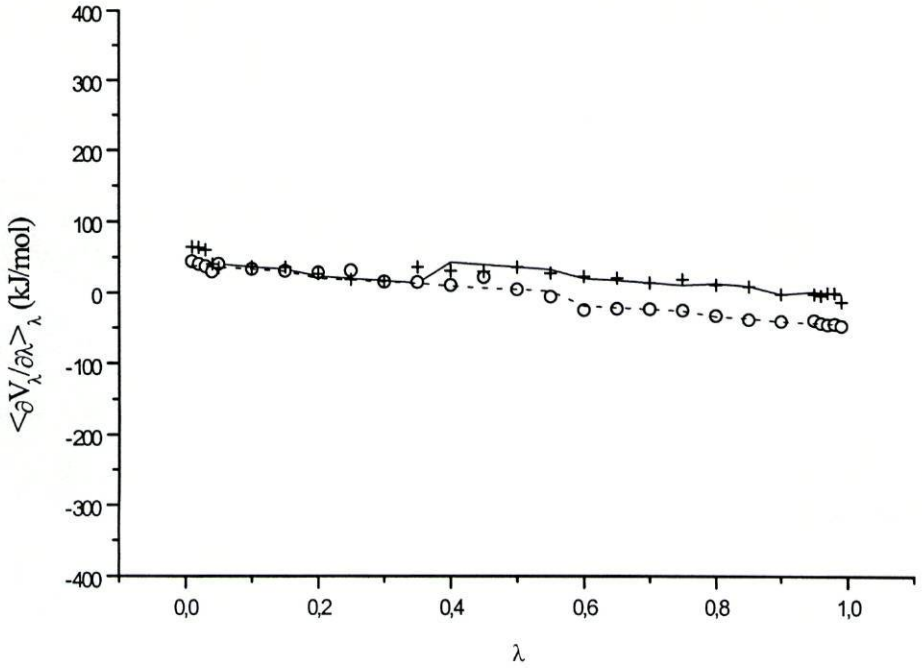


Figura I.9. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Gln^{192} \rightarrow Gln_D$, para (+) o processo 3 (Tripsina→MTripsina); (−) o processo 3 (Tripsina→MTripsina) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→MTripsina:PTI); (- -) o processo 4 (Tripsina:PTI→MTripsina:PTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

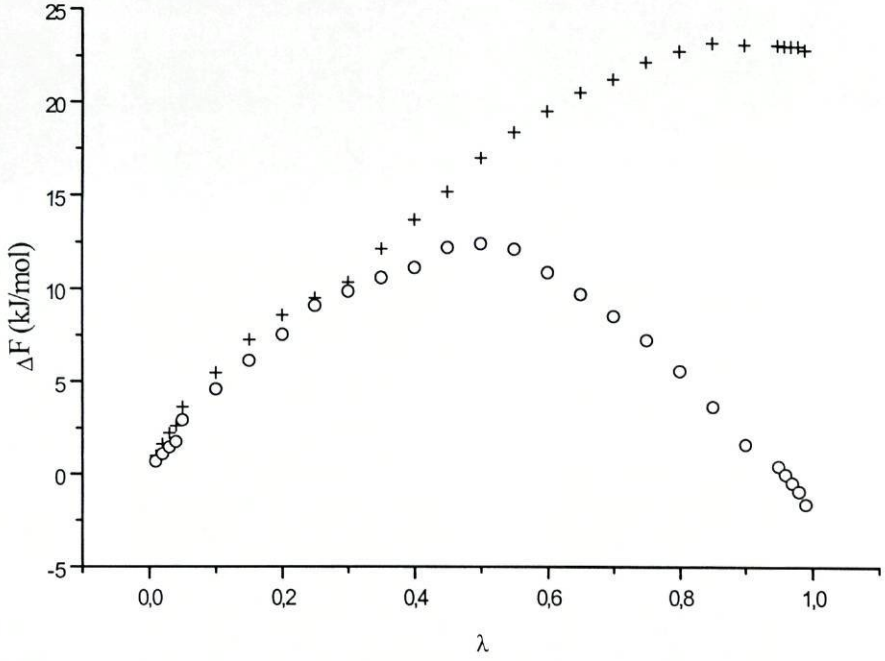


Figura I.10. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (Tripsina→MTripsina); (o) o processo 4 (Tripsina:PTI→MTripsina:PTI), na mutação $Gln^{192} \rightarrow Gln_D$.

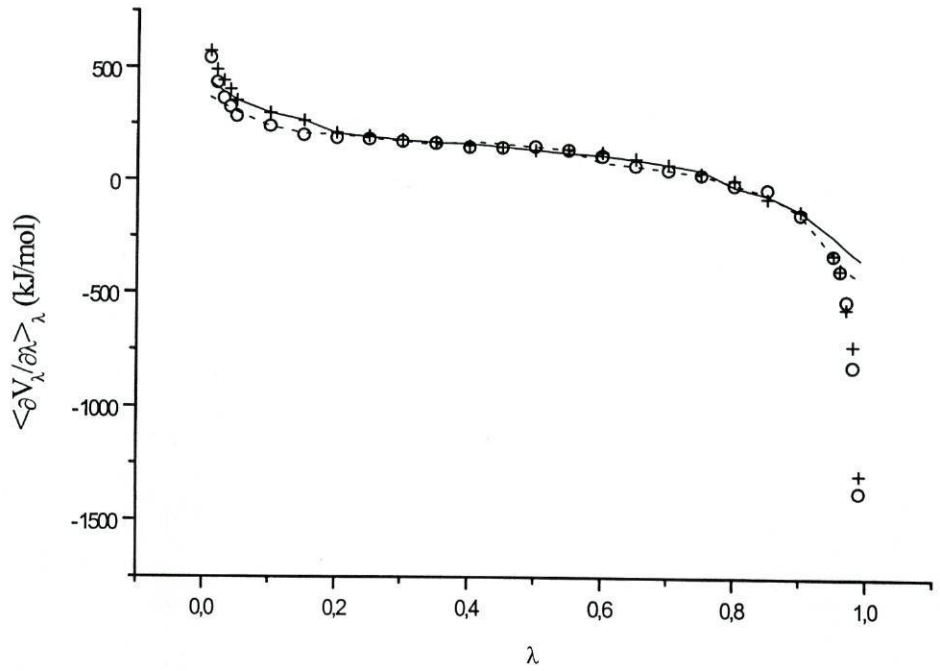


Figura I.11. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação Gln^{192}→Ser, para (+) o processo 3 (Tripsina→MTripsina); (−) o processo 3 (Tripsina→MTripsina) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→MTripsina:PTI); (- -) o processo 4 (Tripsina:PTI→MTripsina:PTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

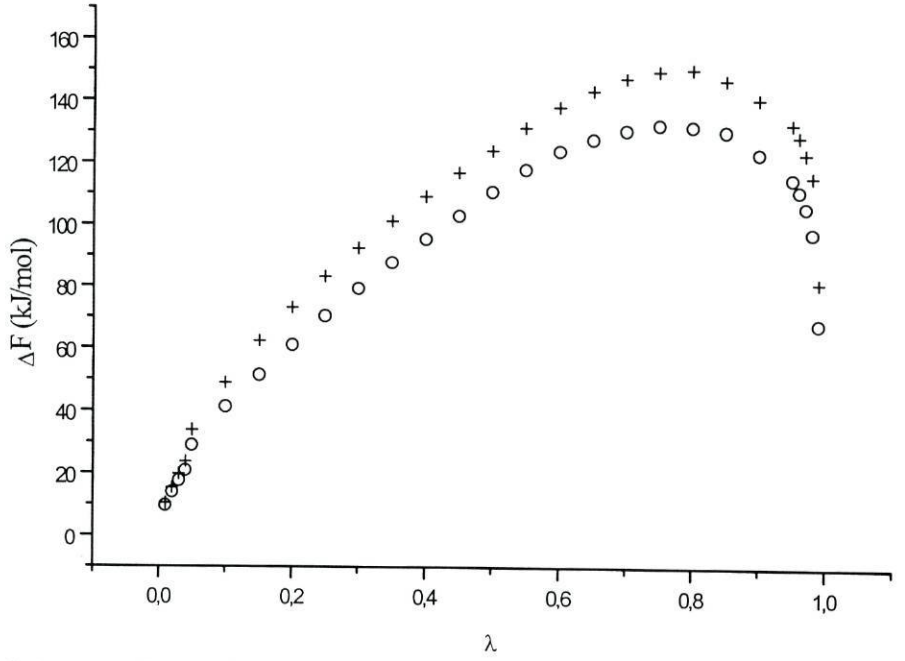


Figura I.12. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (Tripsina→MTripsina); (o) o processo 4 (Tripsina:PTI→MTripsina:PTI), na mutação Gln^{192}→Ser.

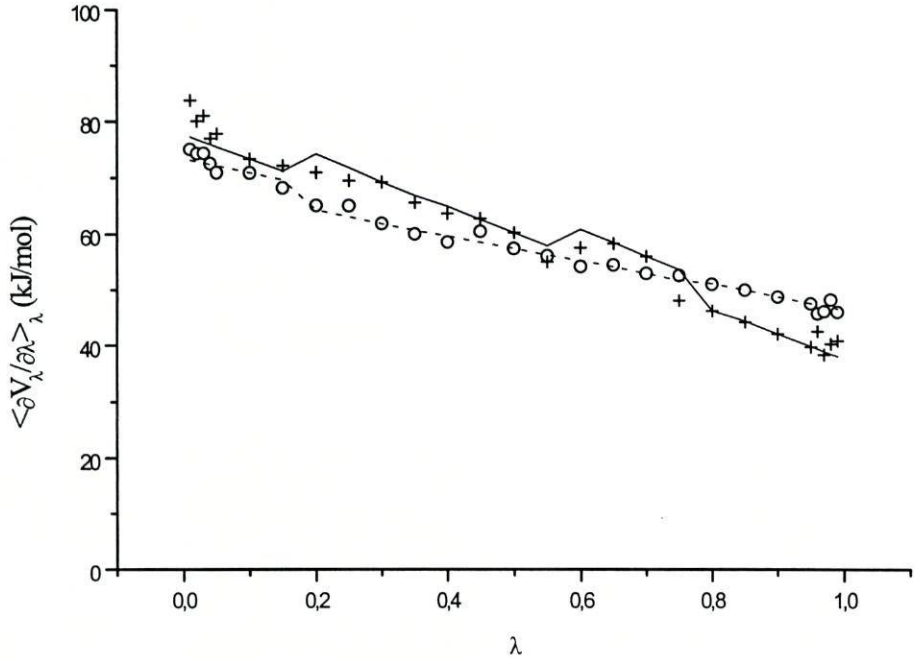


Figura I.13. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Lys^{17} \rightarrow Lys_{D1}$, para (+) o processo 3 (PTI→MPTI); (-) o processo 3 (PTI→MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI); (- -) o processo 4 (Tripsina:PTI→Tripsina:MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

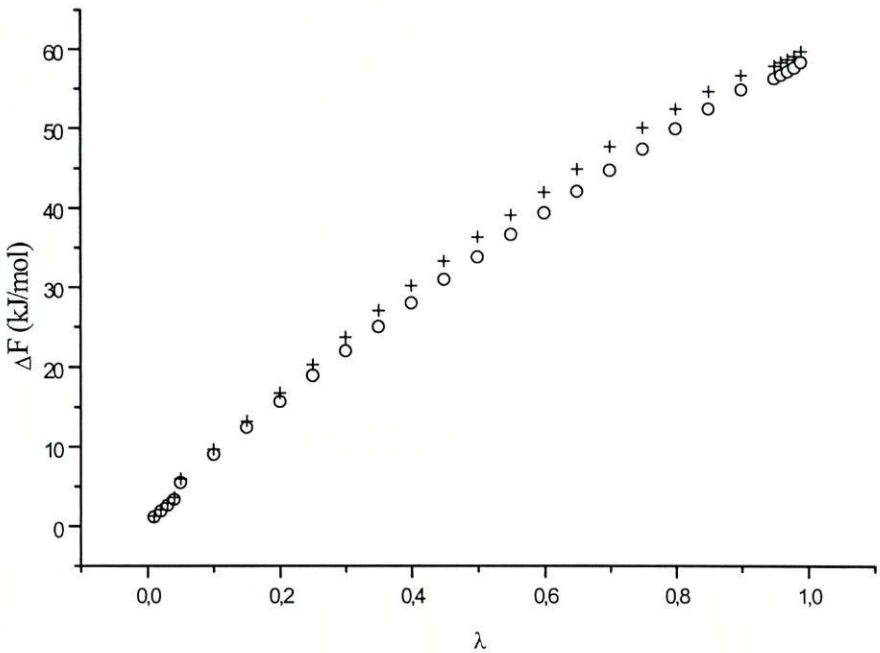


Figura I.14. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (PTI→MPTI); (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI), na mutação $Lys^{17} \rightarrow Lys_{D1}$.

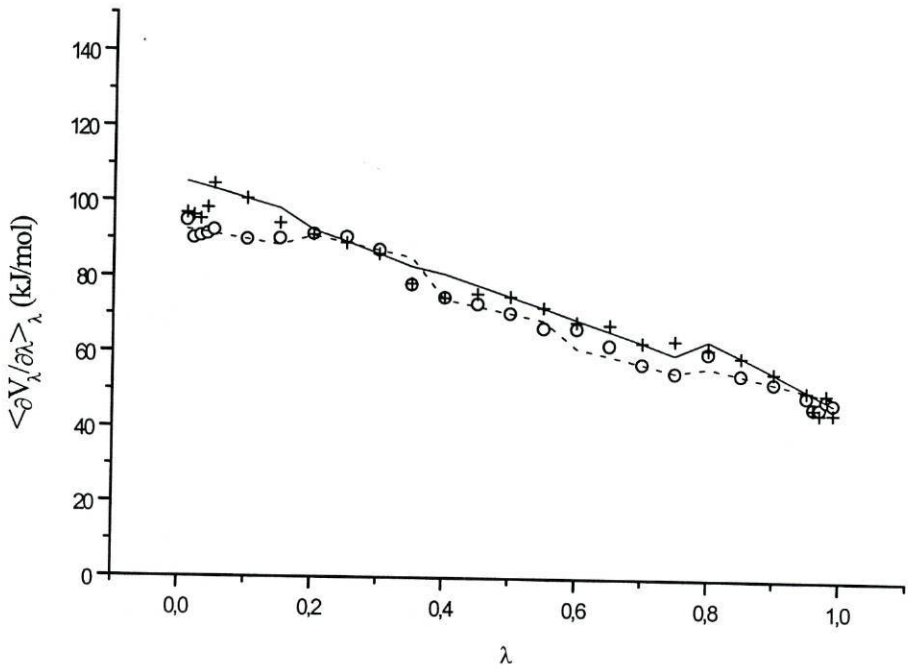


Figura I.15. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Arg^{17} \rightarrow Arg_D$, para (+) o processo 3 (PTI→MPTI); (-) o processo 3 (PTI→MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI); (- -) o processo 4 (Tripsina:PTI→Tripsina:MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

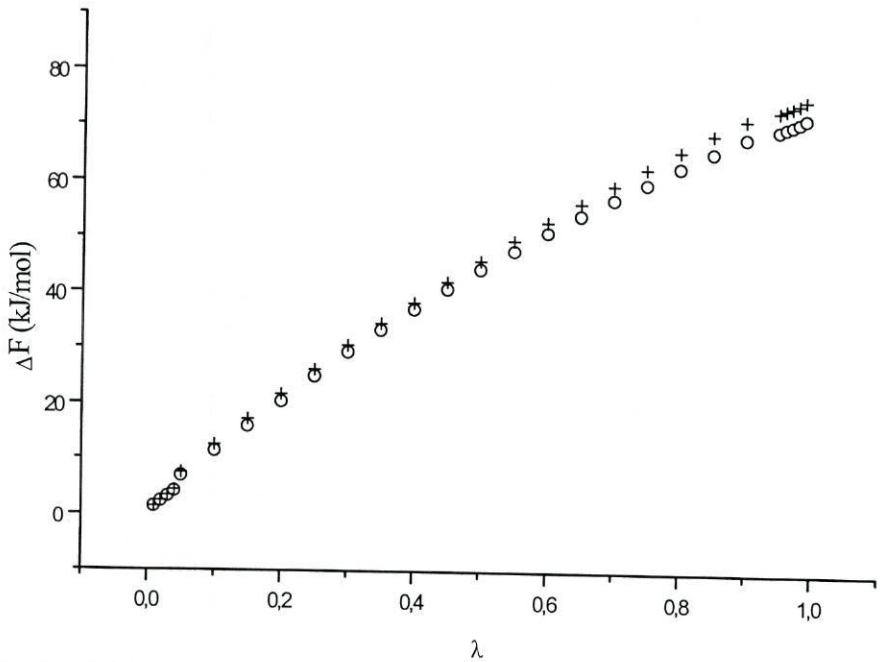


Figura I.16. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (PTI→MPTI); (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI), na mutação $Arg^{17} \rightarrow Arg_D$.

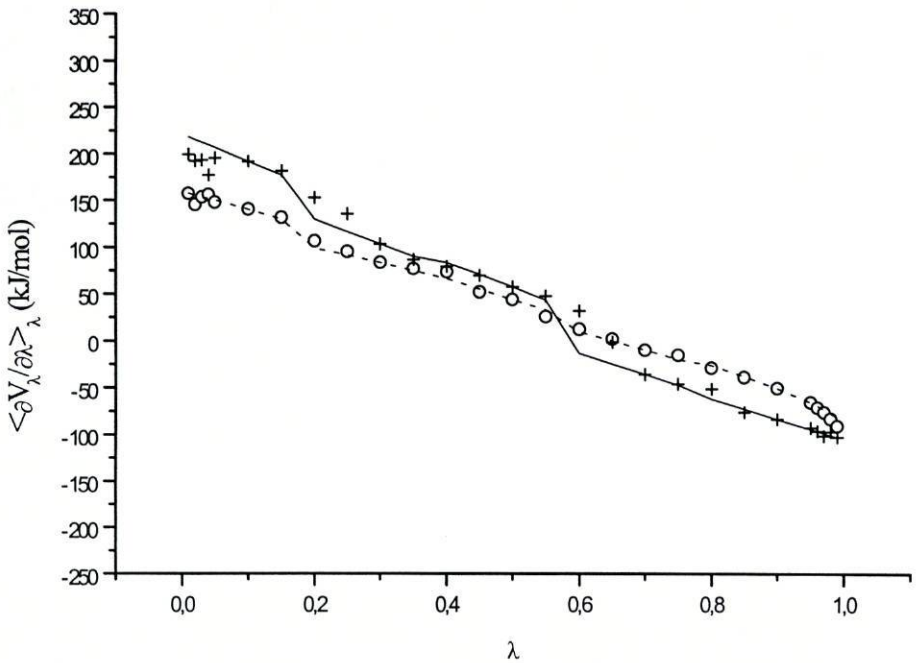


Figura I.17. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Lys^{17} \rightarrow Lys_{D2}$, para (+) o processo 3 (PTI→MPTI); (-) o processo 3 (PTI→MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI); (- -) o processo 4 (Tripsina:PTI→Tripsina:MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

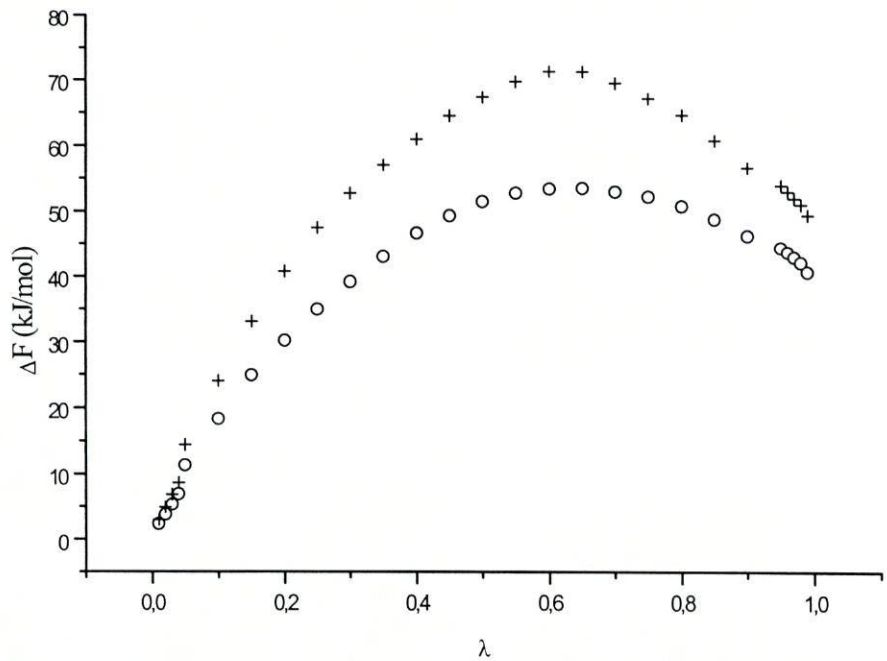


Figura I.18. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (PTI→MPTI); (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI), na mutação $Lys^{17} \rightarrow Lys_{D2}$.

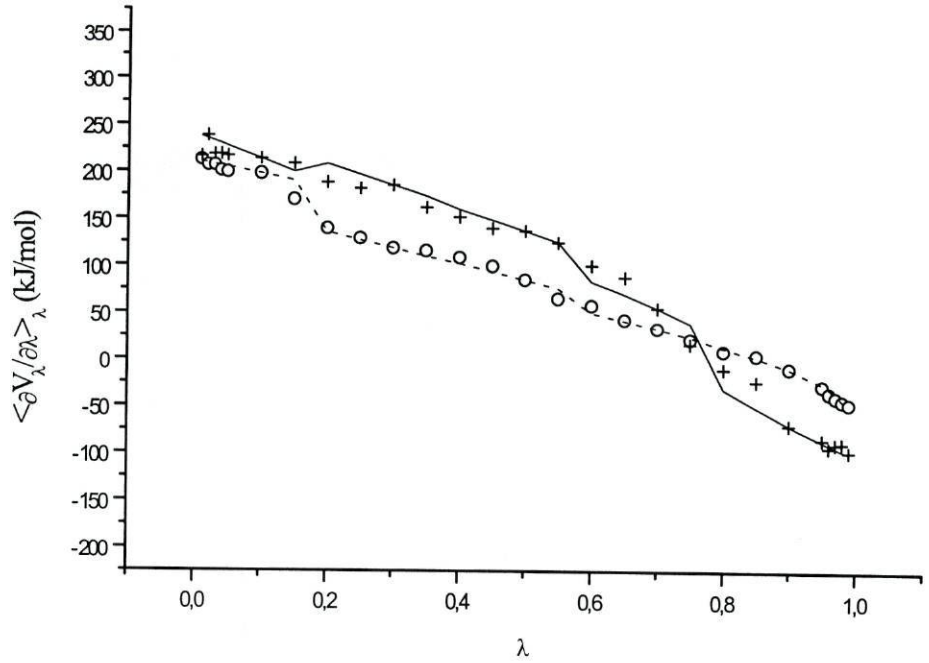


Figura I.19. Gráfico de $\langle \partial V_\lambda / \partial \lambda \rangle_\lambda$ em função de λ, na mutação $Pro^{13} \rightarrow Pro_D$, para (+) o processo 3 (PTI→MPTI); (-) o processo 3 (PTI→MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9; (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI); (- -) o processo 4 (Tripsina:PTI→Tripsina:MPTI) usando o método de extrapolação de Brooks III a partir dos estados de referência λ=0,1, 0,3, 0,5, 0,7 e 0,9.

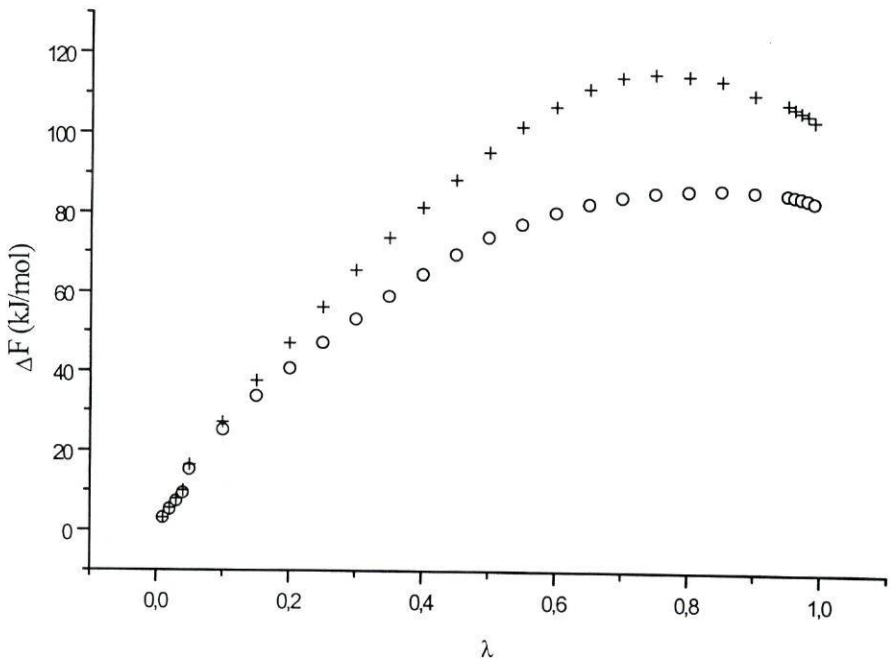


Figura I.20. Gráfico da variação cumulativa da energia de Helmholtz conformacional em função de λ calculada pelo método da perturbação termodinâmica para (+) o processo 3 (PTI→MPTI); (o) o processo 4 (Tripsina:PTI→Tripsina:MPTI), na mutação $Pro^{13} \rightarrow Pro_D$.

Apêndice II

Desvios quadráticos médios totais obtidos nas simulações computacionais das mutações estudadas na região de contacto tripsina:PTI

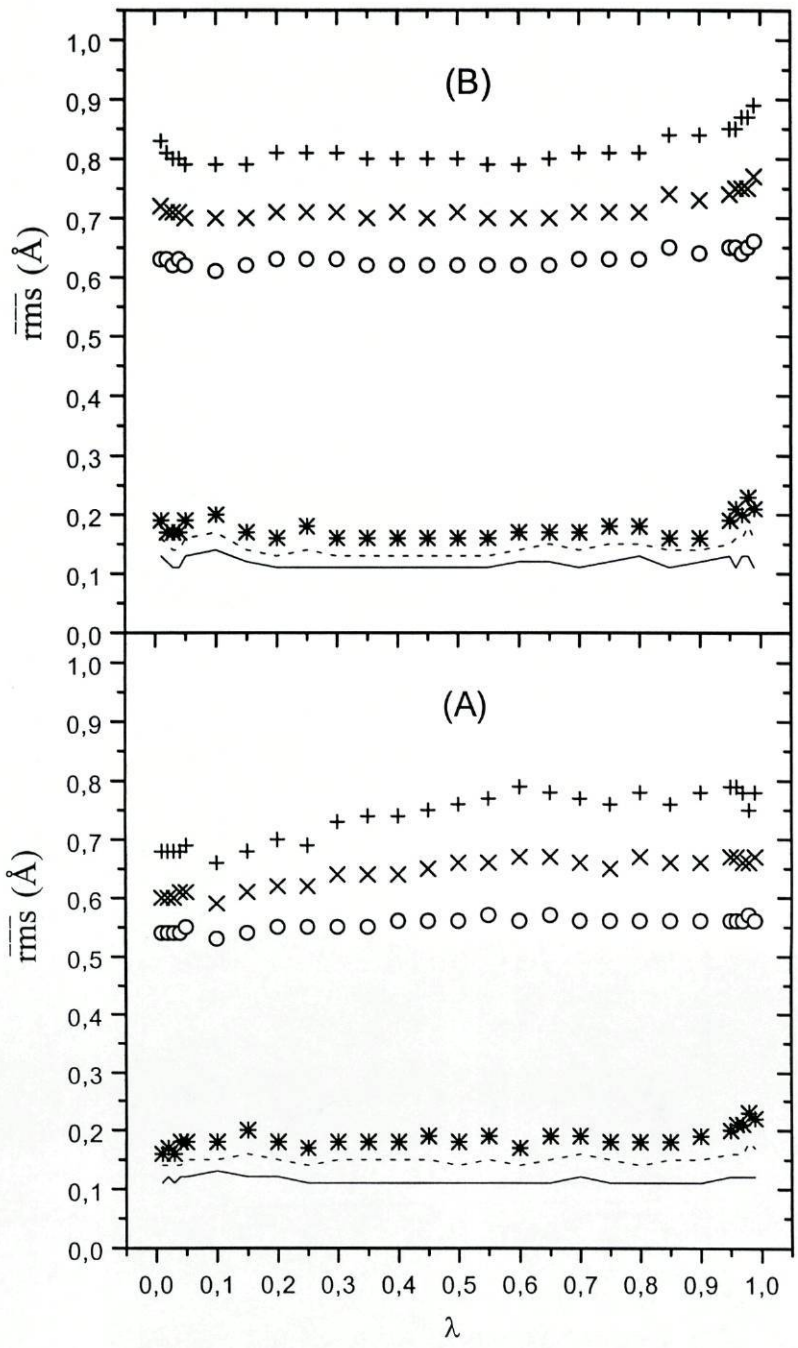


Figura II.1. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (Tripsina→MTripsina) e (B) o processo 4 (Tripsina:PTI→MTripsina:PTI) na mutação Asp^{189}→Ala, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (−); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

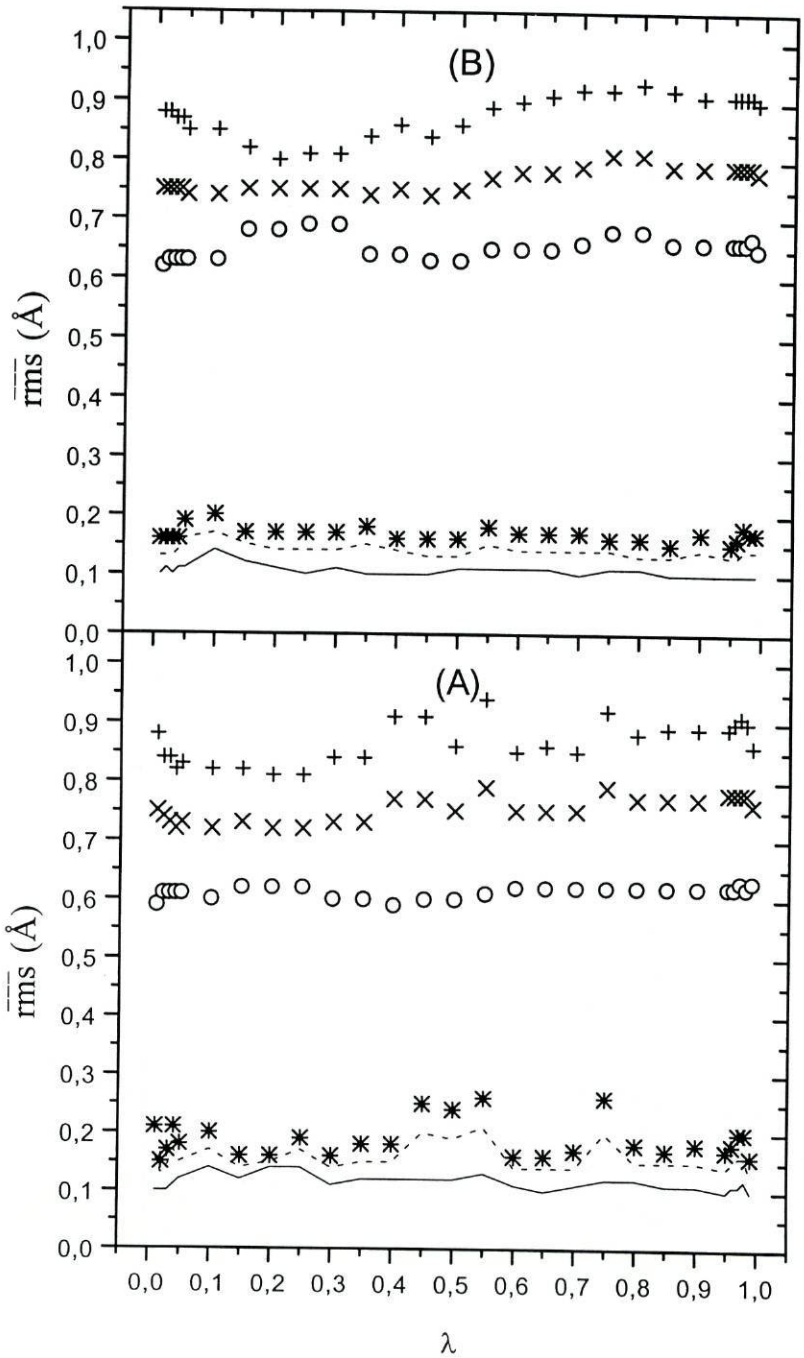


Figura II.2. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (PTI→MPTI) e (B) o processo 4 (Tripsina:PTI→Tripsina:MPTI) na mutação Arg^{17}→Argn, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

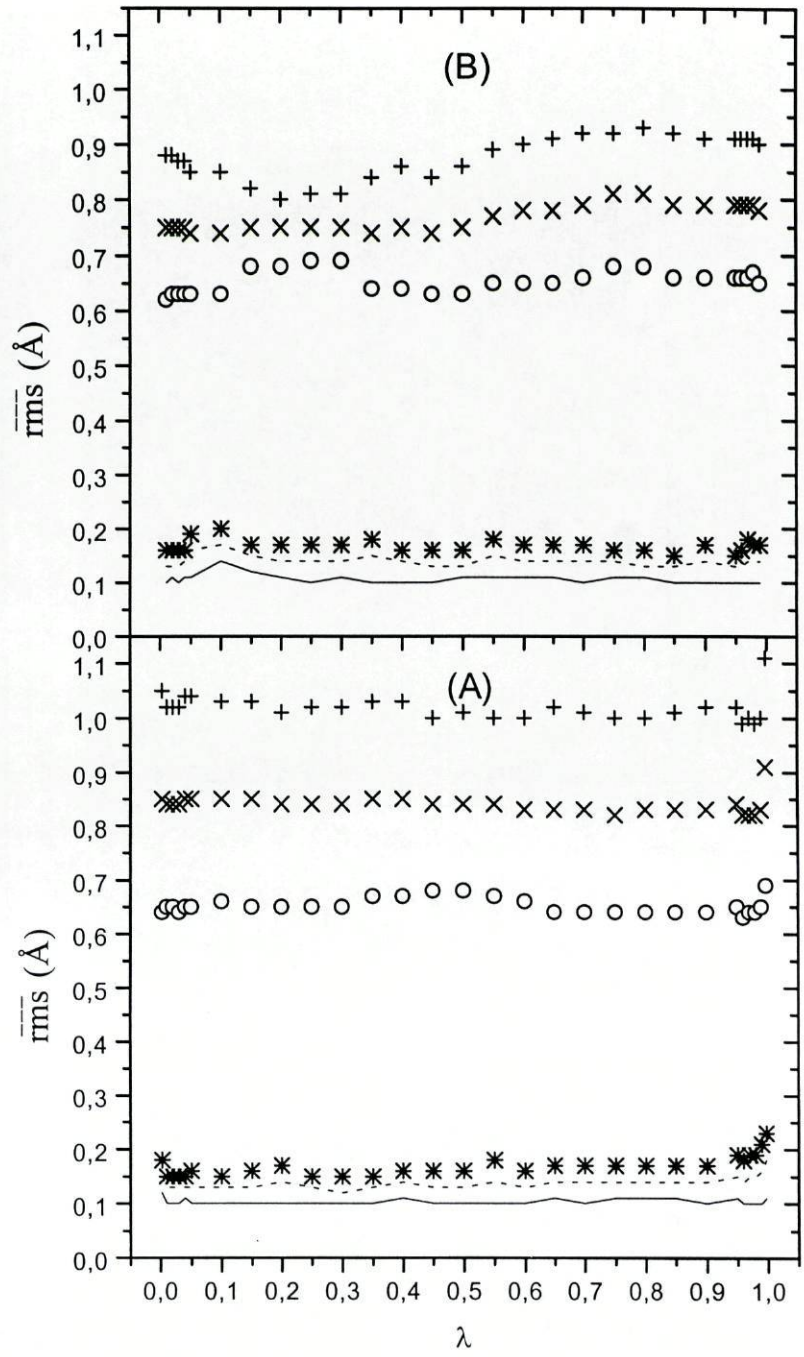


Figura II.3. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (Tripsina→MTripsina) e (B) o processo 4 (Tripsina:PTI→MTripsina:PTI) na mutação Tyr^{39}→Phe, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

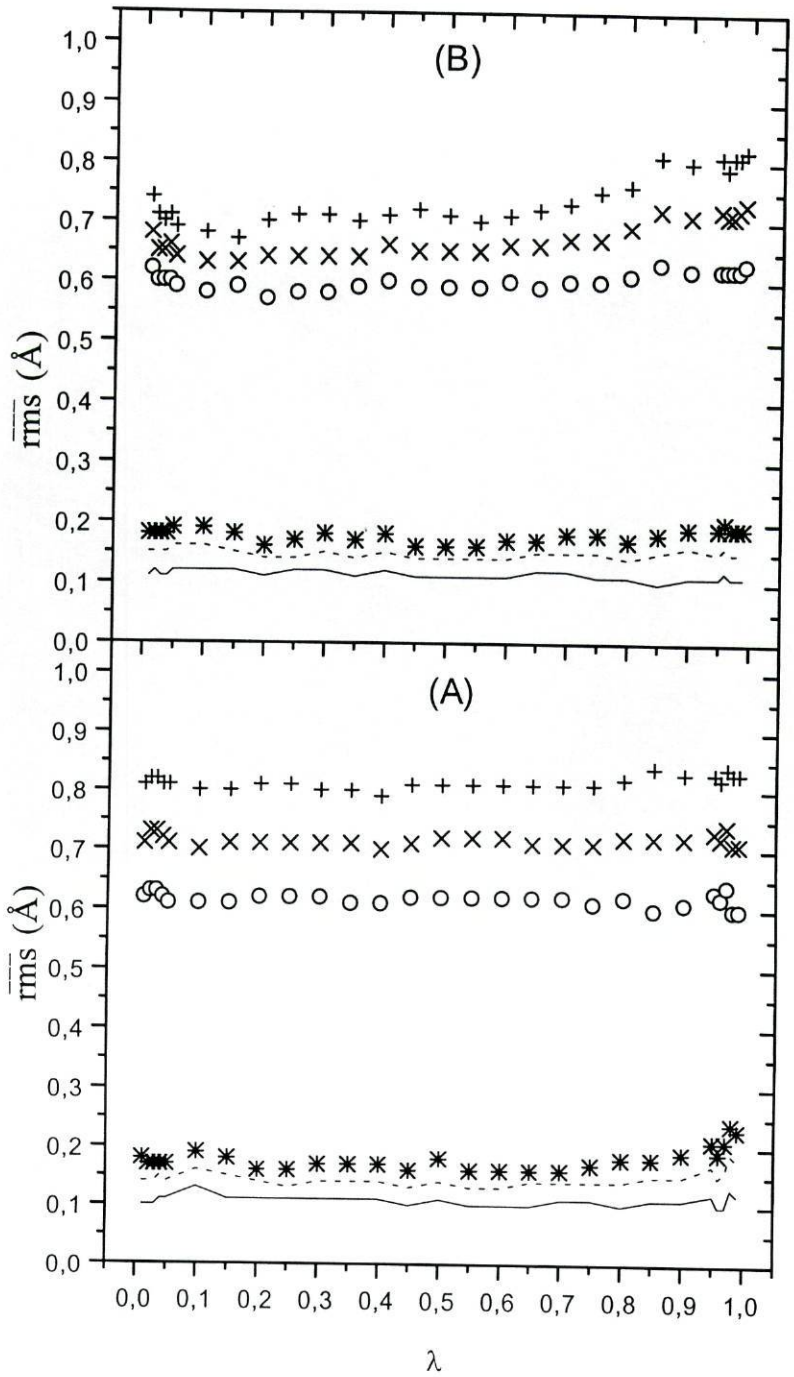


Figura II.4. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (Tripsina→MTripsina) e (B) o processo 4 (Tripsina:PTI→MTripsina:PTI) na mutação Gln^{192}→Leu, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

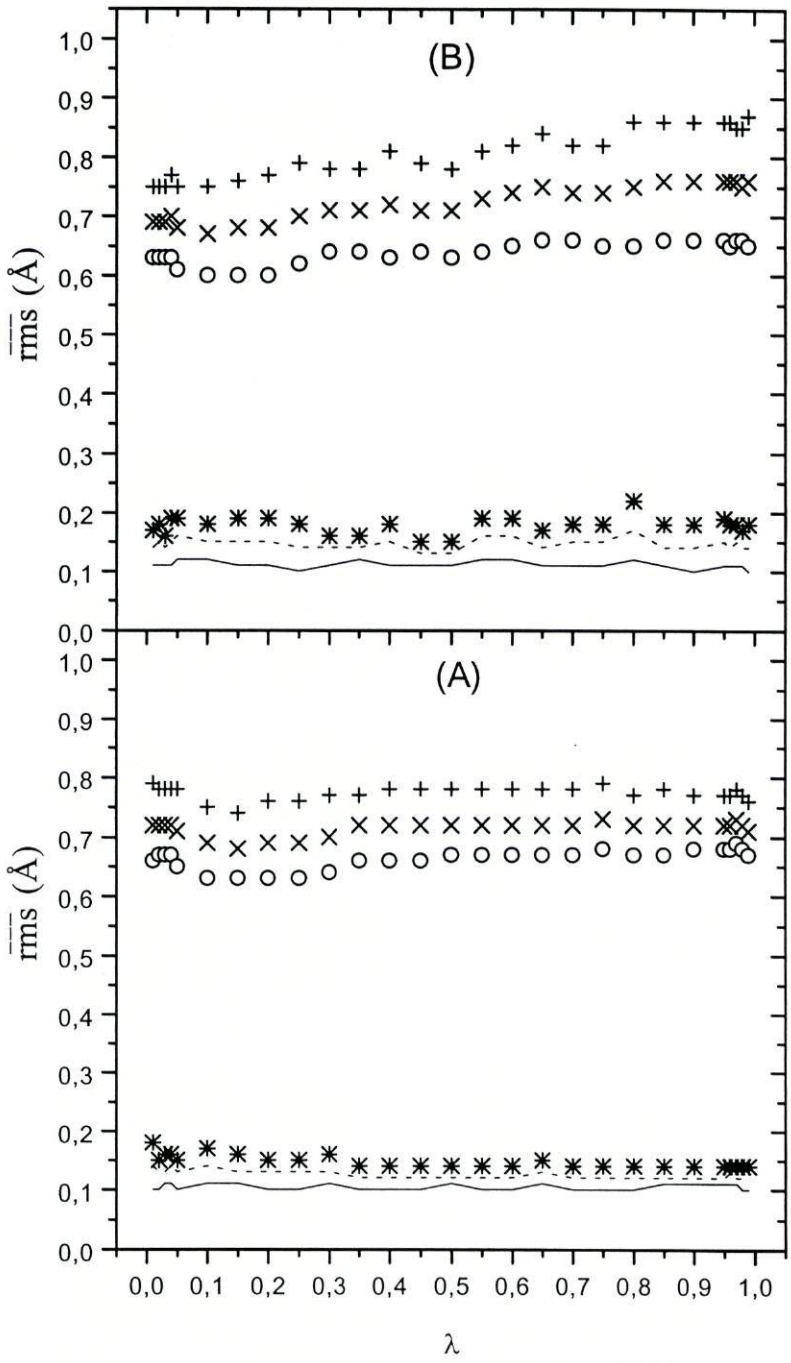


Figura II.5. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (Tripsina→MTripsina) e (B) o processo 4 (Tripsina:PTI→MTripsina:PTI) na mutação Gln^{192}→Gln_D, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

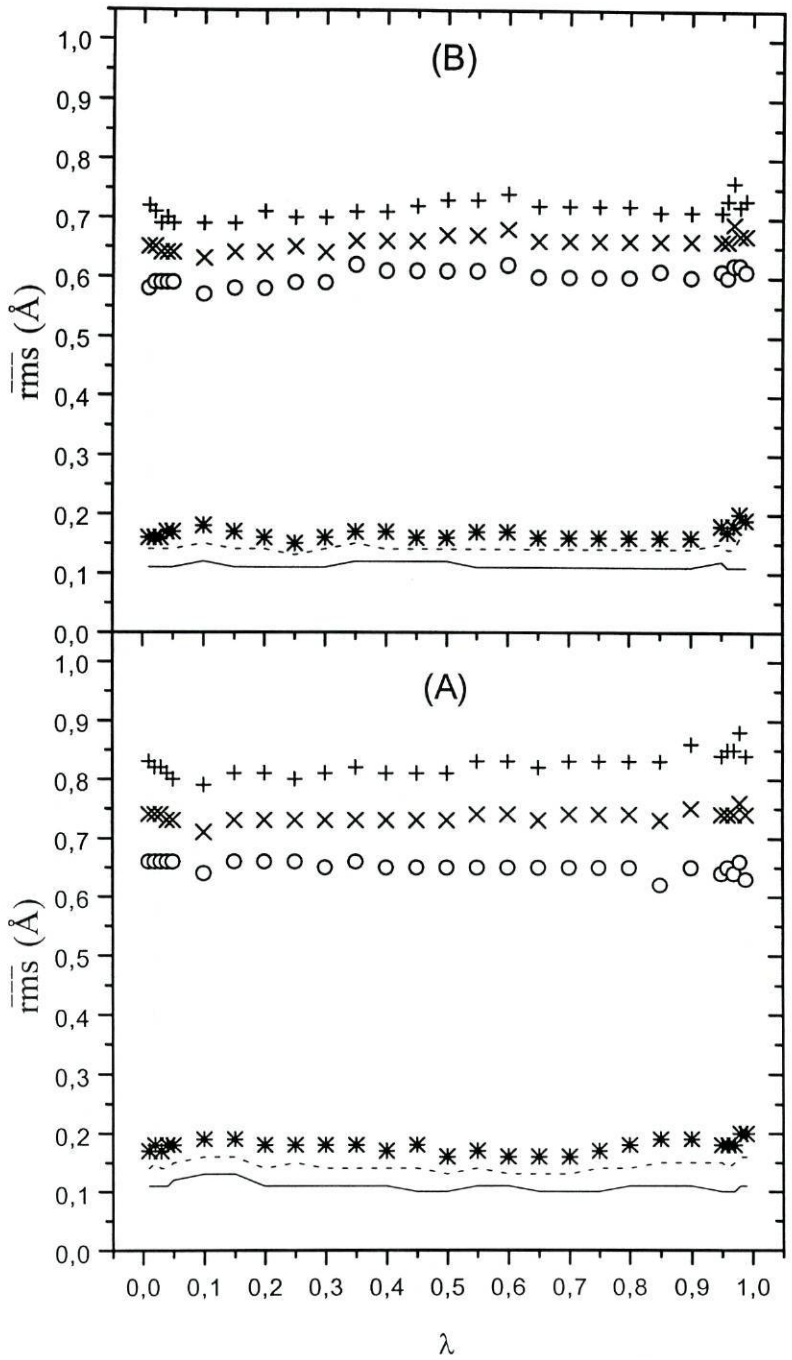


Figura II.6. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (Tripsina→MTripsina) e (B) o processo 4 (Tripsina:PTI→MTripsina:PTI) na mutação Gln^{192}→Ser, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

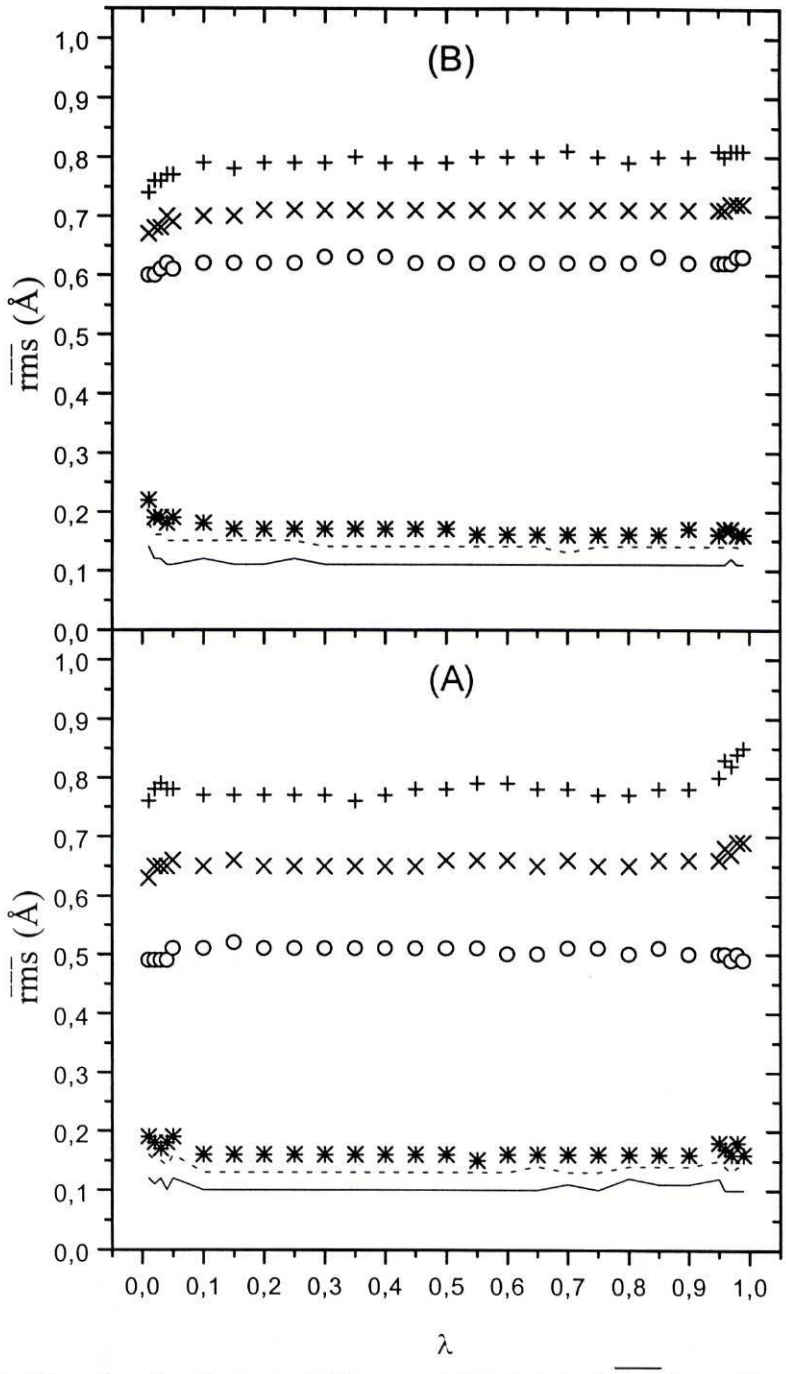


Figura II.7. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (PTI→MPTI) e (B) o processo 4 (Tripsina:PTI→Tripsina:MPTI) na mutação Lys^{15}→Lys_{D1}, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

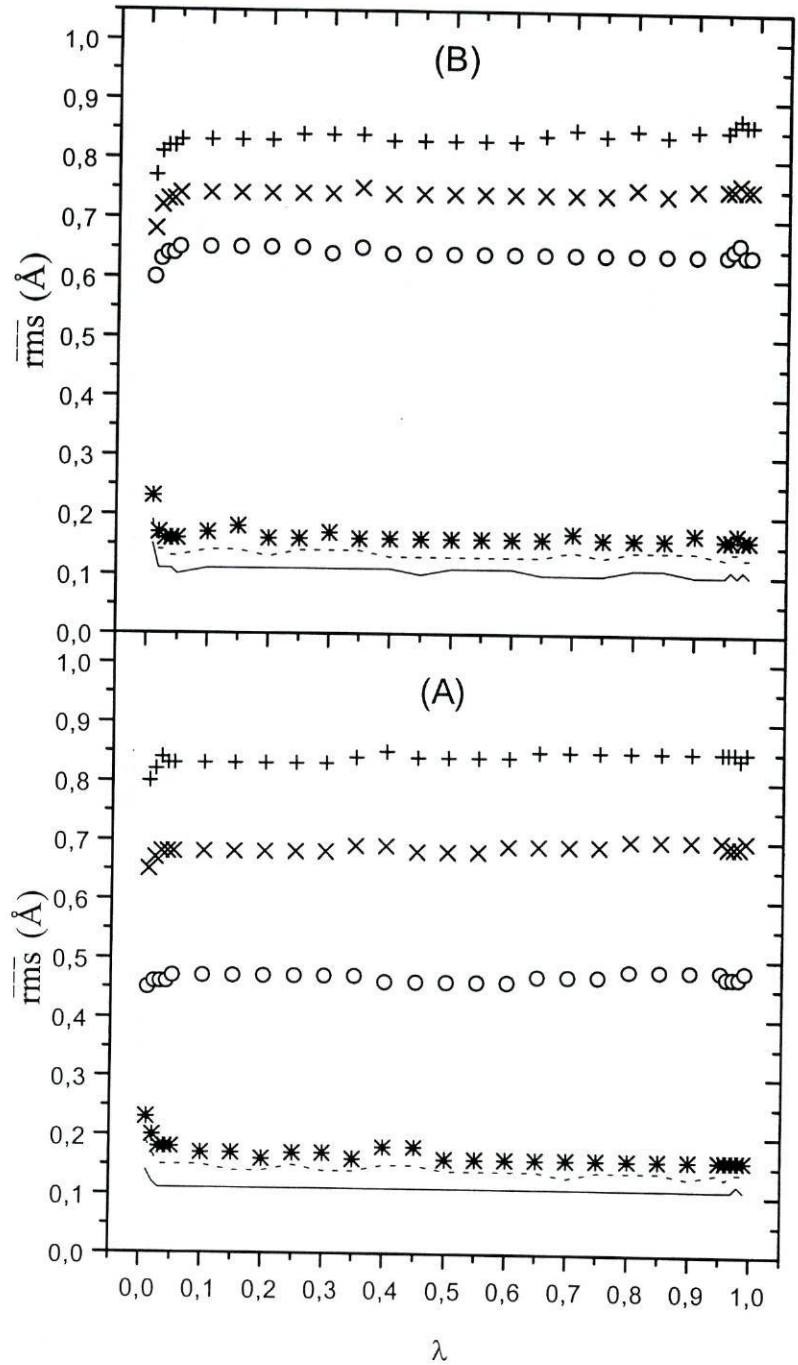


Figura II.8. Gráfico dos desvios quadráticos médios totais ($\overline{\text{rms}}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (PTI→MPTI) e (B) o processo 4 (Tripsina:PTI→Tripsina:MPTI) na mutação Arg^{17}→Arg_D, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

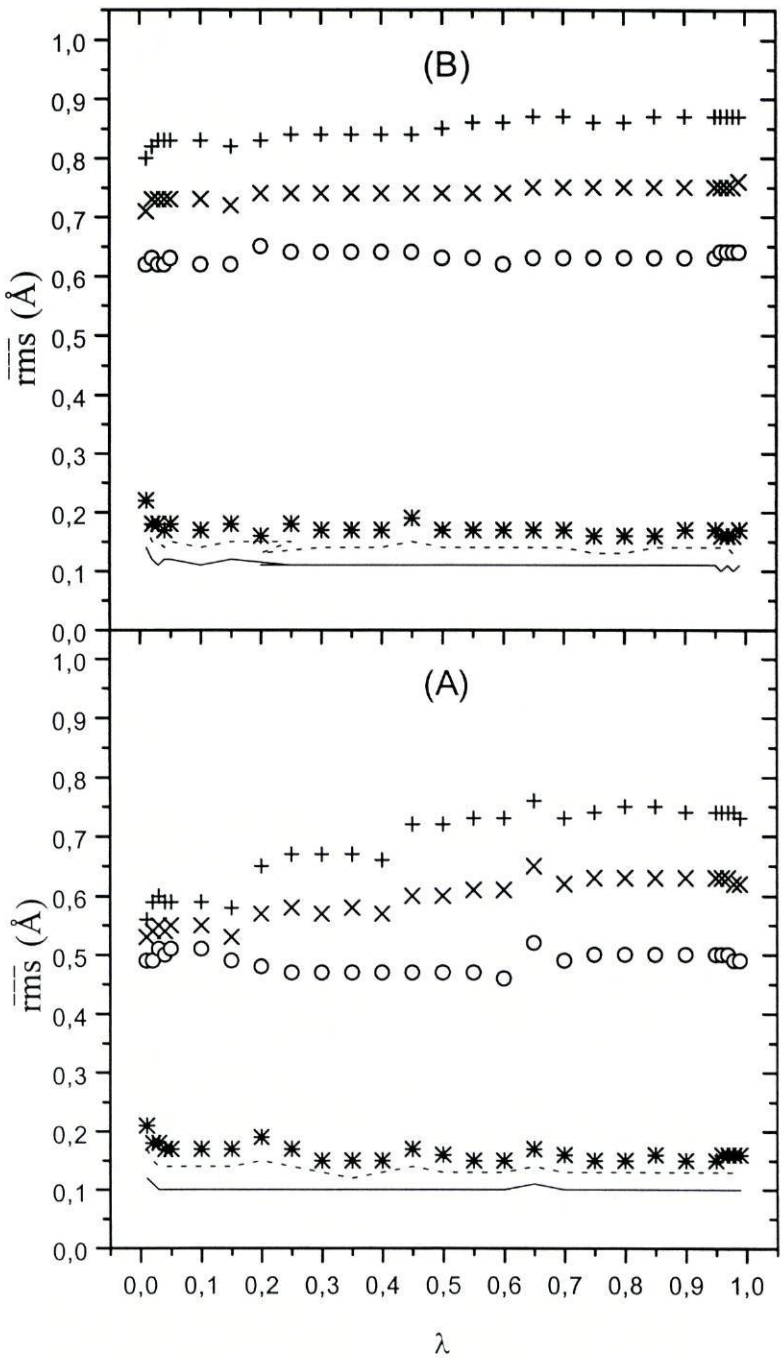


Figura II.9. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (PTI→MPTI) e (B) o processo 4 (Tripsina:PTI→Tripsina:MPTI) na mutação Lys^{15}→Lys_{D2}, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).

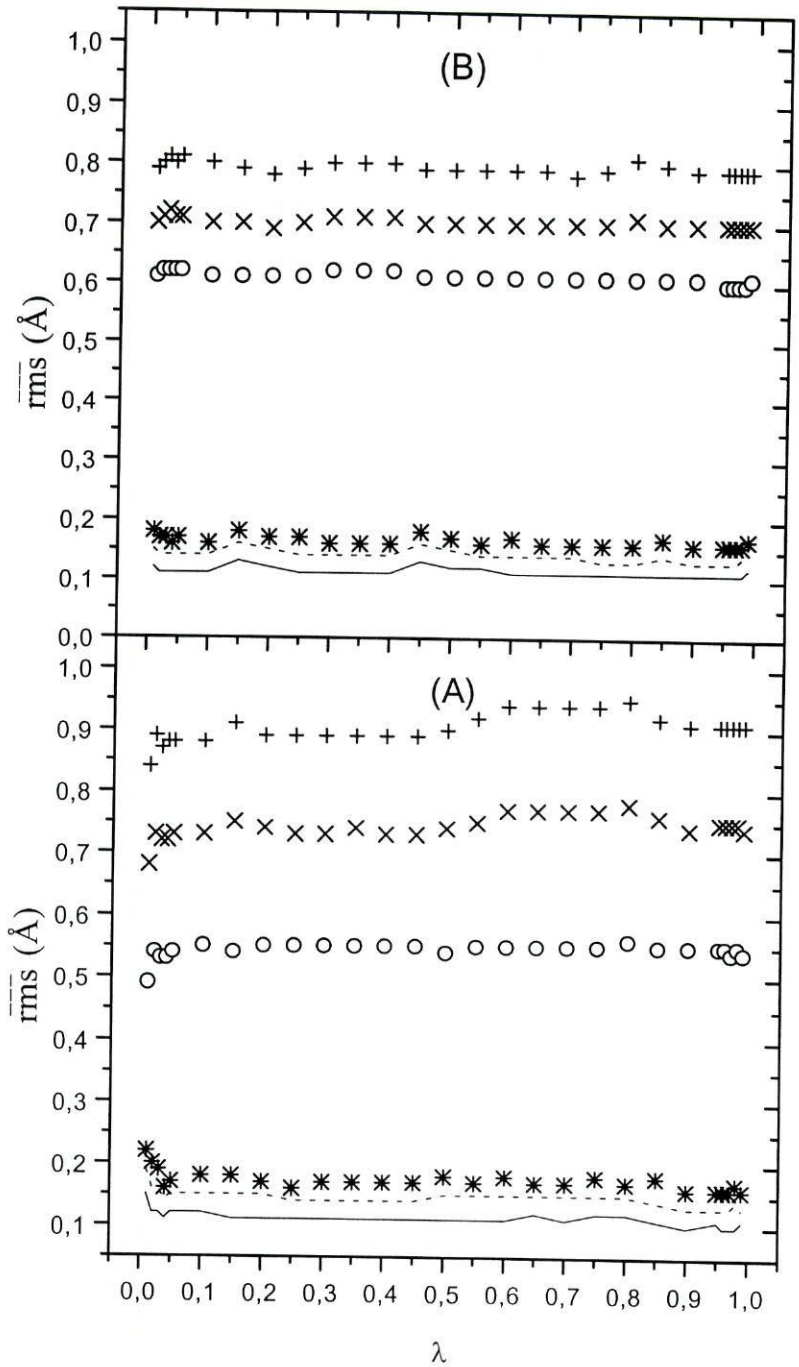


Figura II.10. Gráfico dos desvios quadráticos médios totais ($\overline{rms}$) em função de λ, calculados com base nas trajectórias obtidas por dinâmica molecular para (A) o processo 3 (PTI→MPTI) e (B) o processo 4 (Tripsina:PTI→Tripsina:MPTI) na mutação Pro^{13}→Pro_D, dos átomos pesados relativamente à estrutura média (--); dos átomos pesados da cadeia principal relativamente à estrutura média (—); dos átomos pesados das cadeias laterais relativamente à estrutura média (*); dos átomos pesados relativamente à estrutura de raios X (x); dos átomos pesados da cadeia principal relativamente à estrutura de raios X (o); dos átomos pesados das cadeias laterais relativamente à estrutura de raios X (+).