

MARÇAL DE OLIVEIRA NETO

**" ASPECTOS QUÂNTICOS DA REATIVIDADE DE BIOMOLÉCULAS:
POTENCIAIS ELETROSTÁTICOS E INTERAÇÕES SOLUTO-SOLVENTE "**

T e s e

de

D O U T O R A D O

Centro Brasileiro de Pesquisas Físicas

Rio de Janeiro

1 9 8 5

AGRADECIMENTOS

Aos Professores Mário e Míriam Giambiagi pelo valioso apoio e enriquecidas discussões durante a preparação deste trabalho.

Aos Professores Bernard e Alberte Pullman que muito me ensinaram. Agradeço-os igualmente pelas facilidades que me foram oferecidas no Institut de Biologie Physico-Chimique em Paris para a realização dos cálculos computacionais.

A todos os amigos do Centro Brasileiro de Pesquisas Físicas pela calorosa acolhida e a gentileza presente em todos os dias.

Aos meus pais pela constante dedicação.

Ao Carlos Lopes pelo cuidadoso trabalho de datilografia.

Ao Centro Brasileiro de Pesquisas Físicas, pelas oportunidades oferecidas para a realização deste trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo auxílio financeiro.

Ao povo brasileiro pela reconquista de nossa democracia.

RESUMO

Neste trabalho estudamos, através dos métodos da Fisi-co-Química Quântica, dois diferentes problemas implicando moléculas biológicas.

É apresentada uma técnica para o cálculo do potencial eletrostático macromolecular e aplicando-a à molécula do ácido ribonucleico de transferência tARN^{Phe}. Baseados nos resultados obtidos discutimos alguns dos aspectos das propriedades reativas desta molécula.

Foi também considerado o problema da interação de bio moléculas com o solvente. Em um primeiro estudo investigamos a hidratação do glioxal, metil-glioxal e outros aldeídos simples.

Posteriormente, elaboramos uma metodologia com o objetivo de estabelecer os sítios preferenciais de interação entre macromoléculas e pequenas moléculas polares. Este método foi aplicado para estudar a interação dos constituintes dos ácidos nucleicos com moléculas de água e, em seguida, no sentido de obter as feições gerais de hidratação de uma dupla-hélice modelo do ácido desoxiribonucleico ADN na sua conformação B. Comparamos estes resultados com estudos experimentais.

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	1
CAPÍTULO 2 - MÉTODOS QUÂNTICOS MOLECULARES UTILIZADOS ..	8
2.1 - Método do Campo Auto-Consistente (ou Método SCF)	8
2.2 - O Método LCAO	11
2.3 - A Decomposição da Energia da Interação Intermolecular. Método de Perturbação de London	15
2.4 - A Energia de Interação Eletrostática Molecular	17
2.5 - Uma Aproximação Multipolar para Cálculos de E_{cou} e V_{AB}	21
CAPÍTULO 3 - ASPECTOS REATIVOS DO ARN DE TRANSFERÊNCIA ESPECÍFICO DA FENILALANINA: tARN ^{Phe}	27
3.1 - O tARN ^{Phe} e sua Estrutura Molecular	27
3.2 - O Cálculo do Potencial Eletrostático de um Biopolímero	35
3.3 - Estudo dos Sítios Reativos das Bases Nucléicas do tARN ^{Phe}	39
3.3.1 - O Quetoxal	42
3.3.2 - O Carbodiimido	44
3.3.3 - O 7-bromometil-antraceno	44

3.3.4 - O N-2-Acetoxi-2-Acetilaminofluoreno (AAAF)	45
CAPÍTULO 4 - HIDRATAÇÃO DE SISTEMAS BIOMOLECULARES: Algumas Aplicações Teóricas	49
4.1 - Hidratação do Glioxal, Metil-Glioxal e Outros Aldeídos Simples	49
4.2 - Interações do ADN-B e seus Constituintes com Moléculas de Água: Uma Abordagem Eletrostática	54
4.2.1 - Caracterização da Flexibilidade de uma Molécula de Água em um Determinado Sítio de Interação	54
4.2.2 - Aplicações aos Constituintes dos Ácidos Nucléicos	62
4.2.3 - Sítios Preferenciais de Atração de uma Molécula de ADN-B por Moléculas de Água ...	66
CONCLUSÕES E PERSPECTIVAS	73
APÊNDICE A	77
APÊNDICE B	84
APÊNDICE C	101
APÊNDICE D	109
APÊNDICE E	138
BIBLIOGRAFIA	165

CAPÍTULO 1INTRODUÇÃO

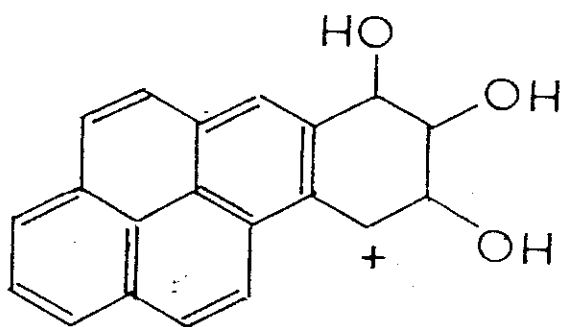
Dentro do vasto campo de pesquisas que envolvem as moléculas biológicas, o estudo da interação destes sistemas moleculares com agentes exteriores à célula assim como com aqueles que integram o meio celular, constitui parte importante deste domínio.

Os ácidos nucléicos, suportes químicos da matéria viva, quando atacados por uma determinada classe de compostos, agentes alquilantes ou hidrocarbonetos aromáticos policíclicos podem levar a uma desorganização do funcionamento do código genético caracterizando assim o câncer químico.

Um número importante de trabalhos experimentais tem mostrado que várias substâncias cancerígenas são agentes fortemente eletrofílicos; ions positivos ou moléculas neutras cujos átomos deficientes em elétrons tornam-se sítios reativos, procurando assim se fixarem nas regiões ricas em elétrons de macromoléculas biológicas, tais como as proteínas e os ácidos nucléicos^{1,2}.

Estes últimos apresentam numerosos sítios nucleofílicos (grupos fosfatos e certos átomos ricos em elétrons das bases púricas e pirimídicas, por exemplo) e constituem lugares importantes de fixação de um grande número de moléculas cancerígenas³. Uma das características destes compostos constitui a a

finidade seletiva para com sítios específicos dos ácidos nucleicos. Por exemplo, para o ADN (ácido desoxiribonucléico), o cancerígeno benzo(a)pirene FIGURA 1.1.a se fixa preferencialmente sobre os grupos amino das bases nucleicas na seguinte ordem relativa: guanina, após adenina e finalmente citosina); já o N-acetoxi-N₂-aminofluoreno FIGURA 1.1.b prefere se fixar no carbono 8 das guaninas^{3,4}.



a) Benzo(a)pirene

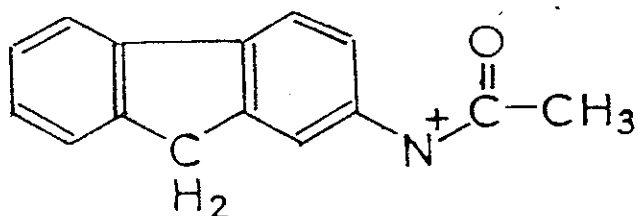
b) N-acetoxi-N₂-aminofluoreno

FIGURA 1.1

Uma das técnicas utilizadas para abordar teóricamente o problema da reatividade de moléculas biológicas com compostos que tem as características acima apresentadas consiste no cálculo do potencial eletrostático molecular gerado pelo conjunto dos elétrons e núcleos da molécula em estudo⁵⁻⁷. Esta técnica pode fornecer informações valiosas sobre as propriedades reativas de biomoléculas e um grande número de estudos foi devotado à molécula do ADN (ver por exemplo uma revisão nas Refs. 8 e 9).

No Capítulo 2 desta tese apresentaremos o procedimen

to teórico - fundamentado na mecânica quântica - que nos permitirá obter uma expressão para o potencial eletrostático molecular. No Capítulo 3 mostraremos uma aplicação para a molécula do ácido ribonucléico (ARN) de transferência específico da fenilalanina ($tARN^{Phe}$).

Este Capítulo será ainda complementado por dois artigos incorporados na forma de apêndices. O primeiro destes trabalhos se refere aos potenciais eletrostáticos gerados pelas diversas conformações dos grupos fosfatos encontradas no $tARN^{Phe}$ e o segundo focaliza os resultados do potencial global desta macromolécula calculado nas vizinhanças dos principais sítios nucleofílicos das bases nucleicas que a constituem.

O outro aspecto importante no estudo de moléculas biológicas, mencionado no início deste Capítulo, é o da influência do meio sobre o comportamento destas moléculas. Os elementos constituintes do meio no qual se encontra uma determinada biomolécula exercem um papel fundamental não somente sobre suas características conformacionais como também sobre suas propriedades químicas e biológicas.

Entre os numerosos fatores ambientais, a influência da água, solvente natural dos sistemas moleculares biológicos, ocupa um lugar preponderante.

Seguindo este eixo de pesquisa, apresentaremos no Capítulo 4 dois tipos de enfoque para tratar o problema da interação entre substratos biológicos com moléculas de água:

- 1 - um dentro do quadro do procedimento *ab initio* SCF supermoléculas no estudo comparativo da hidratação de moléculas relativamente pequenas, tais como o glioxal, o metilglioxal e outros aldeídos simples;
- 2 - o outro, objetivando determinar através de uma abordagem eletrostática os sítios preferenciais de atração de macromoléculas biológicas por moléculas de água, calculando as energias de interação (ver Capítulo II) entre a molécula do substrato com as moléculas do solvente. Uma aplicação será feita para a molécula do ADN.

Com respeito ao segundo enfoque, sua motivação decorre das numerosas pesquisas experimentais realizadas no sentido de estudar o efeito ambiental sobre as estruturas conformacionais da molécula do ADN logo após sua descoberta por Watson e Crick em 1953¹⁰.

Os trabalhos efetuados sobre a difração de raios X por fibras de ADN mostraram que esta molécula apresenta diferentes conformações em função da taxa de hidratação e da concentração iônica do meio. O primeiro destes trabalhos^{11,12} colocou em evidência a existência de dois tipos de diagramas de difração produzidos pelas fibras de ADN (designadas pelos autores de B e A respectivamente), segundo o estado mais ou menos hidratado destas fibras. Este fato levou a uma série de estudos experimentais, utilizando diferentes métodos, no sentido de obter informações sobre a interação do solvente com os ácidos nucléicos (uma revisão sobre este assunto está exposta na Ref. 13).

Outras conformações de dupla-hélices foram descobertas: ADN-C¹⁴, ADN-D¹⁵ e ADN-Z¹⁶. Podemos também citar os trabalhos de Falk e colaboradores¹⁷⁻¹⁹ na determinação dos prováveis sítios de hidratação dos ácidos nucléicos.

Em relação aos estados conformacionais da molécula de ADN, o modelo desta na conformação B parece ser a melhor representação global do ADN em solução ou na célula²⁰.

Os primeiros trabalhos teóricos realizados no sentido de estudar o problema da solvatação de macromoléculas biológicas consistiram em aplicar uma abordagem levando em conta unicamente os fatores estereoquímicos²¹. A molécula do solvente é representada por uma esfera rígida colocada em contato com um átomo ou grupos de átomos da macromolécula. A acessibilidade destes sítios para com a molécula do solvente pode ser então estimada via interações puramente estereoquímicas. Este método foi inicialmente aplicado aos átomos das proteínas²¹ e em seguida aos átomos dos ácidos nucléicos²².

Um estudo teórico ou experimental detalhado sobre a natureza da hidratação de macromoléculas biológicas continua ainda problemático tendo em vista as dimensões e a complexidade destes sistemas. Investigações teóricas têm sido feitas no sentido de obter a distribuição média de moléculas de água em torno de substratos tais como o ADN ou proteínas usando o método Monte Carlo com diferentes tipos de potenciais para interações entre os átomos do solvente e os átomos do substrato^{23,24}.

No Capítulo 4 apresentaremos uma técnica que permite considerar simultaneamente os fatores estereoquímicos e energéticos da interação entre macromoléculas biológicas e pequenas moléculas polares, em particular a molécula de água. No desenvolvimento desta técnica nós nos propusemos não somente estabelecer a posição de estabilização e a energia respectiva da molécula de água, como também caracterizar a sua flexibilidade, calculando as energias de um grande número de configurações desta molécula nas vizinhanças de um determinado sítio de interação. Embora um tal procedimento não leve a preencher as lacunas existentes neste domínio ele pode permitir, entretanto, obter informações relevantes sobre as características globais da atração de sistemas macromoleculares para com moléculas de água, fornecendo, por exemplo, algumas das feições sobre a estruturação destas últimas nas diversas regiões das biomoléculas estudadas. Esta sistemática será primeiramente aplicada aos constituintes fundamentais dos ácidos nucleicos: as bases púricas (guanina e adenina) e pirimídicas (citosina, timina e uracil), o grupo fosfato e açúcar. Em seguida apresentaremos os resultados obtidos quando o substrato estudado é uma dupla hélice modelo do ADN na sua conformação B FIGURA 1.2.

As discussões que seguirão no corpo do Capítulo 4 terão como complementos três trabalhos apresentados em forma de apêndices no final desta tese.

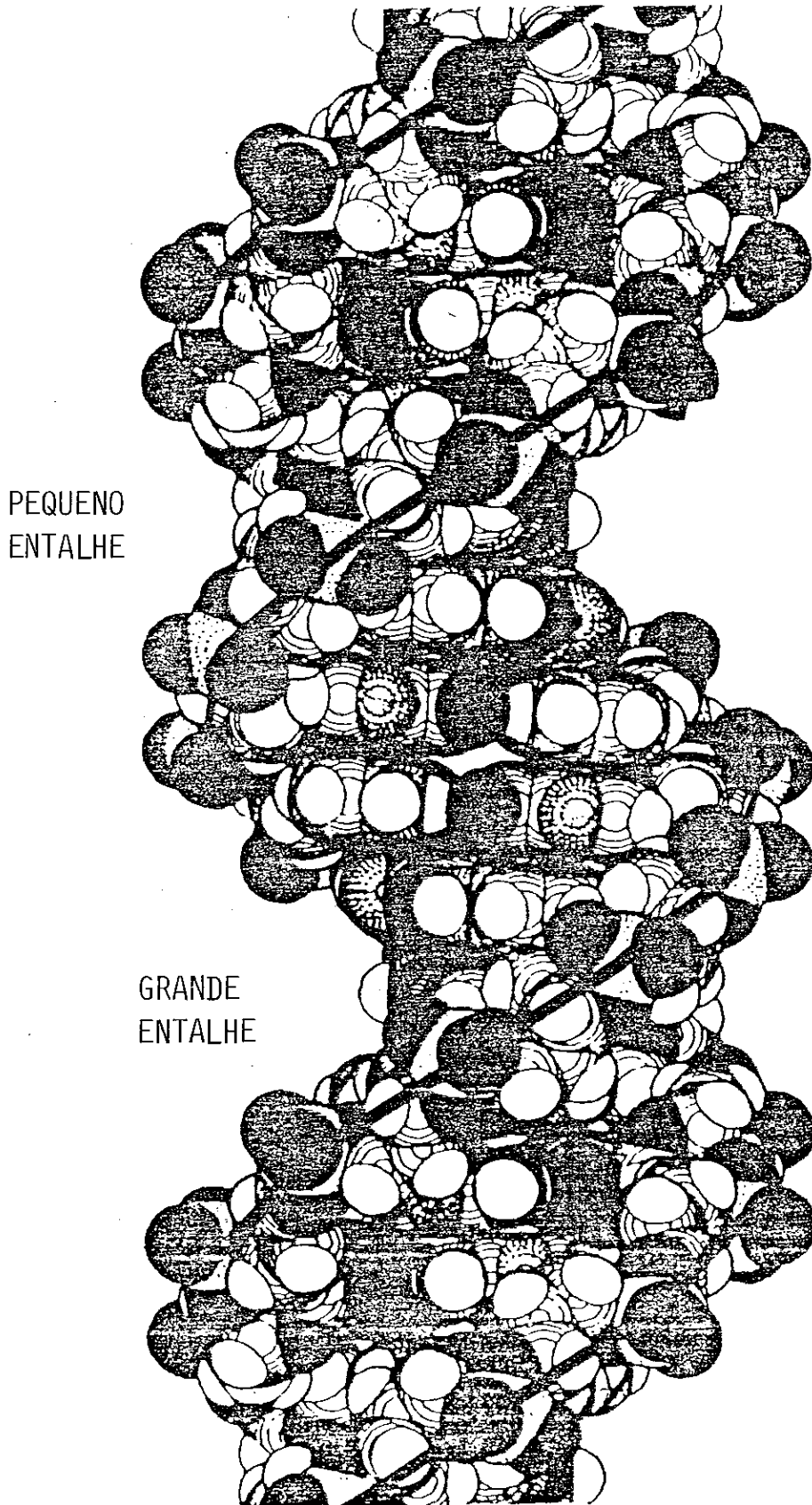


FIGURA 1.2 ADN - B

(Para a definição dos entalhes ver Capítulo 4)

CAPÍTULO 2

MÉTODOS QUÂNTICOS MOLECULARES UTILIZADOS

Neste Capítulo vamos lembrar inicialmente os princípios do método LCAO-SCF que serão utilizado para determinar as funções de onda moleculares. Em seguida, discutiremos a decomposição da energia de interação entre duas moléculas via o tratamento perturbacional, onde obteremos explicitamente uma expressão para a contribuição eletrostática, de particular interesse no desenvolvimento deste trabalho.

2.1 - Método do Campo Auto-consistente (ou Método SCF)

Este método está amplamente exposto na literatura (ver, por exemplo, Ref. 25); nós nos limitaremos a descrever os princípios sobre os quais ele se baseia, assim como os resultados fundamentais obtidos para o caso de moléculas a "camadas fechadas", as únicas que constituíram objeto de cálculos no presente estudo.

Consideremos a expressão para o hamiltoniano molecular na aproximação de Born-Oppenheimer:

$$H = \sum_{\mu} \sum_{\nu} \frac{Z_{\mu} Z_{\nu}}{r_{\mu\nu}} - \frac{1}{2} \sum_i \Delta_i - \sum_{\mu} \sum_i \frac{Z_{\mu}}{r_{\mu i}} + \sum_{i>j} \sum \frac{1}{r_{ij}} \quad (2.1)$$

Com as seguintes notações:

Z_{μ} : carga nuclear,

$r_{\mu\nu}$: distância entre dois núcleos,

$r_{\mu i}$: distância entre o núcleo μ e o elétron i ,

r_{ij} : distância entre dois elétrons i e j ,

Δ_i é o operador laplaciano definido em relação às coordenadas de i .

Supõe-se que a molécula estudada contém um número par n de elétrons. Podemos deixar de lado na eq. (2.1) o primeiro termo que representa a repulsão nuclear; ele é, na realidade, constante para a molécula se os núcleos são supostos imóveis.

A equação de Schroedinger para o sistema eletrônico torna-se então:

$$\left[-\frac{1}{2} \Delta_i + \sum_{\mu} \sum_i \frac{Z_{\mu}}{r_{i\mu}} + \sum_{i>j} \sum \frac{1}{r_{ij}} \right] \psi_{el} = E_{el} \psi_{el} \quad (2.2)$$

Vamos supor que podemos tomar como valor aproximado de ψ_{el} o determinante de Slater:

$$\psi = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(1)\alpha(1) & \phi_1(1)\beta(1) & \dots & \phi_{n/2}(1)\alpha(1) & \phi_{n/2}(1)\beta(1) \\ \phi_1(2)\alpha(2) & \phi_1(2)\beta(2) & \dots & & \\ \vdots & & & & \\ \vdots & & & & \end{vmatrix} \quad (2.3)$$

As $n/2$ funções ϕ_k são funções monoelétrônicas e representam os orbitais moleculares; α e β representam os dois estados de spin que cada elétron pode ocupar. As funções ϕ_k serão considerados reais e impõe-se a elas uma condição de ortonormalização:

$$\langle \phi_k | \phi_l \rangle = \delta_{kl} \quad (2.4)$$

Nestas condições, o valor médio do hamiltoniano eletrônico definido pela eq. (2.2), tomado em relação a ψ é dado por

$$E = 2 \sum_k^{n/2} I_k + \sum_k^{n/2} \sum_\ell^{n/2} (2 J_{k\ell} - K_{k\ell}) \quad (2.5)$$

com:

$$I_k = \int \phi_k(\vec{r}_1) \left[\frac{1}{2} \Delta_1 - \sum_\mu \frac{Z_\mu}{r_{1\mu}} \right] \phi_k(\vec{r}_1) d\tau_1 \quad (2.6)$$

$$J_{k\ell} = \int \phi_k(\vec{r}_1) \phi_k(\vec{r}_1) \frac{1}{r_{12}} \phi_\ell(\vec{r}_2) \phi_\ell(\vec{r}_2) d\tau_1 d\tau_2 \quad (2.7)$$

$$K_{k\ell} = \int \phi_k(\vec{r}_1) \phi_\ell(\vec{r}_1) \frac{1}{r_{12}} \phi_k(\vec{r}_2) \phi_\ell(\vec{r}_2) d\tau_1 d\tau_2 \quad (2.8)$$

Impondo-se que a energia média E seja estacionária em relação a uma variação de uma qualquer das funções ϕ_k , conservando a condição expressa pela eq. (2.4), encontramos que os ϕ_k verificam a seguinte equação,

$$F \phi_k(i) = \epsilon_k \phi_k(i) \quad (2.9)$$

que é também denominada equação de Hartree-Fock²⁶. F é designado o operador de Fock e tem a seguinte expressão:

$$F = -\frac{1}{2} \Delta_1 + \sum_\mu \frac{Z_\mu}{r_{1\mu}} + \sum_k^{n/2} (2J_k - K_k) \quad (2.10)$$

J_k e K_k são chamados operadores coulombiano e de troca respectivamente e são funções dos orbitais moleculares ϕ_k :

$$J_k \theta(2) = \int \phi_k(1) \phi_k(1) \frac{1}{r_{12}} d\tau_1 \theta(2) \quad (2.11)$$

$$K_k \theta(2) = \int \phi_k(1) \theta(1) \frac{1}{r_{12}} d\tau_1 \phi_k(2) \quad (2.12)$$

2.2 - O Método LCAO

Neste método os orbitais moleculares ϕ_k são expressos como combinações lineares dos orbitais atômicos

$$\phi_k = \sum_r^m C_r^k \chi_r \quad m \geq n/2 \quad (2.13)$$

Os χ_r são os orbitais de base, centralizados sobre os átomos, normalizados mas não ortogonais.

O problema de Hartree-Fock, formulado sobre a base dos orbitais atômicos (OA) é escrito da seguinte maneira em notação matricial:

$$\mathbf{F} \mathbf{C}_k = \epsilon \mathbf{S} \mathbf{C}_k \quad (2.14)$$

ou

$$(\mathbf{F} - \epsilon \mathbf{S}) \mathbf{C}_k = \mathbf{0} \quad (2.15)$$

A matriz \mathbb{F} é construída sobre os orbitais atômicos:

$$F_{pq} = \int \chi_p(i) \left[-\frac{1}{2} \Delta_i - \sum_{\mu} \frac{Z_{\mu}}{r_{i\mu}} \right] \chi_q(i) d\tau_i + \sum_{k=1}^{n/2} \int \chi_p(i) [2J_k - K_k] \chi_q(i) d\tau_i \quad (2.16)$$

A matriz \mathbb{C}_k é constituída pelos coeficientes do orbital molecular ϕ_k (eq. 2.13).

$$\mathbb{C}_k = \begin{vmatrix} C_1^k \\ C_2^k \\ \vdots \\ C_m^k \end{vmatrix}$$

e a matriz \mathbb{S} caracterizada pelos elementos

$$S_{pq} = \int \chi_p(i) \chi_q(i) d\tau_i \quad (2.17)$$

As equações 2.15 são as chamadas equações de Roothaan²⁷. Elas constituem um sistema linear e homogêneo onde as incógnitas são os C_r^k . Este sistema admite soluções não todas nulas se - e somente se - seu determinante é nulo:

$$| \mathbb{F} - \epsilon \mathbb{S} | = 0 \quad (2.18)$$

e a cada valor de ϵ que satisfaz esta equação corresponde uma matriz \mathbb{C}_i .

As equações de Roothaan podem ser resolvidas iterativamente. Constroem-se inicialmente um jogo de $n/2$ orbitais moleculares (funções tentativas) que são utilizadas para construir a matriz \mathbb{F} . A resolução da eq. 2.18 conduz a m soluções ϵ_k , as quais correspondem a m jogos de coeficientes C_r^k . Os $n/2$ coeficientes correspondendo aos $n/2$ valores mais baixos de ϵ_k , servem para reconstruir uma matriz \mathbb{F} a qual dará novos valores ϵ_k e novos C_r^k para as eqs. 2.15 e 2.18. Trata-se então de um cálculo iterativo que se repete até que as matrizes \mathbb{C}_k não variam mais de uma iteração à outra. Os OM $\phi_k = \sum_r C_r^k \chi_r$ obtidos não são as soluções as mais gerais da eq. 2.9, a menos que a base χ_r dos orbitais atômicos seja infinitamente grande. Estes orbitais são designados auto-consistentes e o método de resolução iterativa da eq. 2.15 é o método do campo auto-consistente (SCF).

Entretanto, um problema de ordem prática se coloca na determinação dos elementos de matriz:

$$J_{k,pq} = \sum_r \sum_s C_r^k C_s^k \int \chi_r(1) \chi_s(1) \frac{1}{r_{12}} \chi_p(2) \chi_q(2) d\tau_1 d\tau_2 \quad (2.19)$$

$$K_{k,pq} = \sum_r \sum_s C_r^k C_s^k \int \chi_r(1) \chi_p(1) \frac{1}{r_{12}} \chi_s(2) \chi_q(2) d\tau_1 d\tau_2 \quad (2.20)$$

da eq. 2.18, que exigem, para uma molécula poliatômica, o cálculo de integrais bi-eletrônicas do tipo

$$\int \chi_r(1) \chi_s(1) \frac{1}{r_{12}} \chi_p(2) \chi_q(2) d\tau_1 d\tau_2 .$$

O cálculo destas integrais é um problema de dificuldades que va

riam em função das bases atômicas que são empregadas.

Podemos escolher como base uma composta pelos orbitais de Slater centradas sobre cada átomo:

$$\chi_r(n, \ell, m) = N r^{n-1} e^{-\alpha r} y_\ell^m(\theta, \phi)$$

onde N é uma constante de normalização e y_ℓ^m um harmônico esférico. Podemos obter uma excelente aproximação dos orbitais de Hartree-Fock para o caso de átomos, utilizando uma combinação linear de um pequeno número destes orbitais. Entretanto, o cálculo das integrais multicêntricas, quando tratamos de moléculas, é extremamente difícil nesta base.

A fim de contornar esta dificuldade, Boys²⁸ propôs utilizar os orbitais atômicos de base como sendo uma combinação linear de funções gaussianas do tipo:

$$\chi(\alpha, n, \ell, m) = N x^n y^\ell z^m e^{-\alpha r^2};$$

onde α é um parâmetro positivo e x, y, z definidos em relação aos núcleos dos átomos que comportam as gaussianas. Desta maneira torna-se viável o cálculo das integrais multicêntricas intervindo nas eqs. 2.19 e 2.20 o que permite resolver a eq. 2.18 sem aproximações e constitui a razão pela qual, recentemente, uma tal base de orbitais atômicos tem sido amplamente empregadas no domínio da Química Teórica. Neste trabalho nós utilizamos para os átomos da segunda fila da classificação periódica uma base de gaussianas de dimensão (7s, 3p/3s) contraídas em base mínima

(5, 2, 3/3) empregadas por Clementi et al. em estudos de moléculas de interesse biológico²⁹.

2.3 - A decomposição da energia de interação intermolecular. Método de perturbação de London³⁰

Sejam A e B duas moléculas em interação e H_A e H_B seus hamiltonianos respectivos. Nós vamos supor conhecidas as funções de onda das moléculas isoladas e designaremos ψ_A^O e ψ_B^O os estados fundamentais de A e B, ψ_A^P e ψ_B^Q dois estados excitados de A e B. O hamiltoniano total do sistema supermolecular é escrito como:

$$H = H_A + H_B + V_{AB} \quad (2.21)$$

V_{AB} sendo o operador potencial de interação. Se i e j são respectivamente dois elétrons de A e B, k e ℓ os núcleos de A e B, o operador de interação V_{AB} terá a seguinte expressão:

$$V_{AB} = \sum_{k \in A} \sum_{\ell \in B} \frac{Z_k Z_\ell}{|r_k - r_\ell|} + \sum_{i \in A} \sum_{j \in B} \frac{1}{|r_i - r_j|} \\ - \sum_{i \in A} \sum_{\ell \in B} \frac{Z_\ell}{|r_i - r_\ell|} - \sum_{j \in B} \sum_{k \in A} \frac{Z_k}{|r_j - r_k|} \quad (2.22)$$

Os dois primeiros termos da soma representam a repulsão nuclear e a repulsão eletrônica entre A e B, os dois termos seguintes representam a atração entre um elétron de A e um núcleo de B e reciprocamente.

A energia de interação do sistema A-B até os termos

de perturbação de segunda ordem em V_{AB} será dada por:

$$\begin{aligned}
 E_{AB} = E_A^O + E_B^O + \langle \psi_A^O \psi_B^O | V_{AB} | \psi_A^O \psi_B^O \rangle - \\
 \sum_P \frac{\langle \psi_A^P \psi_B^O | V_{AB} | \psi_A^O \psi_B^O \rangle^2}{E_A^P - E_A^O} - \\
 \sum_Q \frac{\langle \psi_A^O \psi_B^Q | V_{AB} | \psi_A^O \psi_B^Q \rangle^2}{E_B^Q - E_B^O} - \sum_P \sum_Q \frac{\langle \psi_A^O \psi_B^O | V_{AB} | \psi_A^P \psi_B^Q \rangle^2}{E_B^Q + E_A^P - E_B^O - E_A^O}
 \end{aligned} \quad (2.23)$$

onde, sobre o exemplo da molécula A

$$E_A^O = \int \psi_A^O H_A \psi_A^O d\tau_A \quad (2.24)$$

$$E_A^P = \int \psi_A^P H_A \psi_A^P d\tau_A \quad (2.25)$$

O terceiro termo da eq. 2.22 constitue a energia de interação eletrostática entre as moléculas A e B e a chamaremos de E_{cou} . Neste desenvolvimento este termo é então a contribuição de primeira ordem em V_{AB} para a energia de interação intermolecular.

Os termos seguintes do desenvolvimento de perturbação constituem a contribuição de segunda ordem em V_{AB} . O quarto e quinto termo indicam a passagem para um estado excitado de cada uma das moléculas no campo criado pela outra; eles constituem a energia de polarização. O último termo indica uma excitação simultânea de A e B e constitue a energia de dispersão.

onde foi considerado a antissimetria da função de onda de ordem zero do sistema A-B em relação a troca de dois elétrons de A e B, para explicar porque, a curtas distâncias, duas moléculas se associam para formar ligações químicas ou então porque elas podem se repelir fortemente. Desta maneira duas novas contribuições aparecem; uma indicando a repulsão das nuvens eletrônicas a curtas distâncias (termo de primeira ordem) e a outra indicando a passagem de um elétron de uma das moléculas para um orbital molecular vazio da outra (energia de transferência de carga - termo de segunda ordem).

2.4 - A Energia de Interação Eletrostática Molecular

Podemos definir para cada molécula A e B, cuja interação estamos estudando, um operador de densidade de carga local⁶:

$$\rho^A(\vec{r}) = \rho_{\text{nuc}}^A(\vec{r}) - \rho_{\text{el}}^A(\vec{r}) \quad (2.26)$$

com

$$\rho_{\text{nuc}}^A(\vec{r}) = \sum_k Z_k \delta(\vec{r} - \vec{r}_k) \quad (2.27)$$

$$\rho_{\text{el}}^A(\vec{r}) = \sum_i \delta(\vec{r} - \vec{r}_i) \quad (2.28)$$

Z_k representa a carga de cada núcleo k de A, \vec{r}_k e \vec{r}_i os vetores posição de cada núcleo e de cada elétron de A.

Resulta destas definições que o operador V_{AB} dado pela eq. 2.22 se identifica ã integral:

$$V_{AB} = \int \frac{\rho^A(\vec{r}_A) \rho^B(\vec{r}_B)}{|\vec{r}_B - \vec{r}_A|} d\tau_A d\tau_B \quad (2.29)$$

E de maneira explícita esta integral tem a seguinte expressão:

$$I = \int \left[\sum_k z_k \delta(\vec{r}_A - \vec{r}_k) - \sum_i \delta(\vec{r}_A - \vec{r}_i) \right] \left[\sum_\ell z_\ell \delta(\vec{r}_B - \vec{r}_\ell) - \sum_j \delta(\vec{r}_B - \vec{r}_j) \right] \frac{d\tau_A d\tau_B}{|\vec{r}_A - \vec{r}_B|} \quad (2.30)$$

O desenvolvimento de I e a aplicação das propriedades da função δ nos leva a obter exatamente o operador V_{AB} expresso pela eq. 2.22.

Desta maneira é possível representar o potencial V_{AB} como o resultado da interação de distribuições de carga, ρ^A e ρ^B , definidas independentemente para cada uma das moléculas A e B.

A energia E_{cou} pode também ser expressa em função de ρ^A e ρ^B :

$$\begin{aligned} E_{\text{cou}} &= \langle \psi_A^0 \psi_B^0 | V_{AB} | \psi_A^0 \psi_B^0 \rangle \\ &= \int \frac{\langle \psi_A^0 \psi_B^0 | \rho^A(\vec{r}_A) \rho^B(\vec{r}_B) | \psi_A^0 \psi_B^0 \rangle}{|\vec{r}_A - \vec{r}_B|} d\tau_A d\tau_B \\ &= \int \frac{\langle \psi_A^0 | \rho^A(\vec{r}_A) | \psi_A^0 \rangle \langle \psi_B^0 | \rho^B(\vec{r}_B) | \psi_B^0 \rangle}{|\vec{r}_A - \vec{r}_B|} d\tau_A d\tau_B \quad (2.31) \end{aligned}$$

Somos então levados a calcular elementos de matriz do tipo

$$\langle \psi_A^0 | \rho^A(\vec{r}_A) | \psi_A^0 \rangle$$

Na prática, as funções ψ_A^0 dependem somente das coordenadas eletrônicas (aproximação de Born-Oppenheimer). Supondo-se que a molécula A contém n elétrons, teremos:

$$\begin{aligned} \langle \psi_A^0 | \rho^A(\vec{r}_A) | \psi_A^0 \rangle &= \langle \psi_A^0 | \sum_k Z_k \delta(\vec{r}_A - \vec{r}_k) - \sum_{i=1}^n \delta(\vec{r}_A - \vec{r}_i) | \psi_A^0 \rangle \\ &= \sum_k Z_k \delta(\vec{r}_A - \vec{r}_k) \langle \psi_A^0 | \psi_A^0 \rangle - \langle \psi_A^0 | \sum_{i=1}^n \delta(\vec{r}_A - \vec{r}_i) | \psi_A^0 \rangle \end{aligned} \quad (2.32)$$

Todos os elétrons de A sendo indistinguíveis:

$$\begin{aligned} \langle \psi_A^0 | \rho^A(\vec{r}) | \psi_A^0 \rangle &= \sum_k Z_k \delta(\vec{r}_A - \vec{r}_k) - n \langle \psi_A^0 | \delta(\vec{r}_A - \vec{r}_i) | \psi_A^0 \rangle \\ &= \sum_k Z_k \delta(\vec{r}_A - \vec{r}_k) - \\ &\quad n \int \psi_A^0(\vec{r}_A, \vec{r}_2, \dots, \vec{r}_n)^* \psi_A^0(\vec{r}_A, \vec{r}_2, \dots, \vec{r}_n) d\tau_2 \dots d\tau_n \end{aligned} \quad (2.33)$$

O segundo termo desta expressão pode ser definido como a *densidade total de carga eletrônica no ponto r_A* e será denotado por $\rho_A(\vec{r}_A)$

Se definirmos⁶:

$$\gamma_A(\vec{r}_A) = -\rho_A(\vec{r}_A) + \sum_{k \in A} Z_k \delta(\vec{r}_A - \vec{r}_k) \quad (2.34)$$

A expressão de E_{cou} será dada por:

$$E_{\text{cou}} = \int \frac{\gamma_A(\vec{r}_A) \gamma_B(\vec{r}_B)}{|\vec{r}_A - \vec{r}_B|} d\tau_A d\tau_B \quad (2.35)$$

Podemos escrever E_{cou} igualmente como:

$$E_{\text{cou}} = \int V_A(\vec{r}_B) \gamma_B(\vec{r}_B) d\tau_B$$

onde $V_A(\vec{r}_B) = \int \frac{\gamma_A(\vec{r}_A) d\tau_A}{|\vec{r}_A - \vec{r}_B|}$ (2.36)

O termo $V_A(\vec{r}_B)$ depende somente do ponto \vec{r}_B e da densidade de carga da molécula A. Designaremos $V_A(\vec{r}_B)$ o potencial eletrostático molecular gerado por A no ponto \vec{r}_B ; podemos também considerá-lo formalmente como a energia eletrostática de interação entre dois sub-sistemas constituídos pela molécula A e uma carga igual à de um proton situada no ponto \vec{r}_B .

Utilizando as eqs. (2.34) e (2.36) obtém-se:

$$V_A(\vec{r}_B) = \sum_{k \in A} \frac{z_k}{|\vec{r}_k - \vec{r}_B|} - \int \frac{\rho_A(\vec{r}_A)}{|\vec{r}_B - \vec{r}_A|} d\tau_A \quad (2.37)$$

Lembrando que: $\rho_A(\vec{r}_A) = \langle \psi_A^0 | \sum_{i=1}^n \delta(\vec{r}_A - \vec{r}_i) | \psi_A^0 \rangle$

Se ψ_A^0 é constituído por um determinante de Slater construído a partir de N orbitais moleculares (N=n/2) duplamente ocupados $a_\mu(r_i)$ e lembrando que $\sum_i^n \delta(\vec{r}_A - \vec{r}_i)$ é uma soma de operadores monoelétrônicos:

$$\rho_A(\vec{r}_A) = 2 \sum_{\mu=1}^N \int a_{\mu}^*(\vec{r}_i) \delta(\vec{r}_A - \vec{r}_i) a_{\mu}(\vec{r}_i) d\tau_i$$

$$\rho_A(\vec{r}_A) = 2 \sum_{\mu=1}^N |a_{\mu}(\vec{r}_A)|^2 \quad (2.38)$$

Expressando os orbitais moleculares a_{μ} como combinações lineares dos orbitais atômicos:

$$a_{\mu} = \sum_{\alpha} C_{\alpha}^{\mu} \chi_{\alpha}$$

$$\rho_A(\vec{r}_A) = 2 \sum_{\mu=1}^N \sum_{\alpha} \sum_{\alpha'} C_{\alpha}^{*\mu} C_{\alpha'}^{\mu} \chi_{\alpha}^*(\vec{r}_A) \chi_{\alpha'}(\vec{r}_A) \quad (2.39)$$

Desta maneira a expressão do potencial eletrostático molecular criado pela molécula A no ponto \vec{r}_B , no quadro da teoria LCAO é escrita como:

$$V_A(\vec{r}_B) = \sum_{k \in A} \frac{Z_k}{|\vec{r}_k - \vec{r}_B|} - 2 \sum_{\mu=1}^N \sum_{\alpha} \sum_{\alpha'} C_{\alpha}^{*\mu} C_{\alpha'}^{\mu} \int \frac{\chi_{\alpha}^*(\vec{r}_A) \chi_{\alpha'}(\vec{r}_A) d\tau_A}{|\vec{r}_B - \vec{r}_A|} \quad (2.40)$$

2.5 - Uma Aproximação Multipolar para os Cálculos de E_{cou} e V_{AB}

A expressão de $V_A(\vec{r}_B)$ como dada pela eq. (2.40) mostra que o cálculo prático tanto de $V_A(\vec{r}_B)$ assim como de E_{cou} necessita a determinação de um grande número de integrais do tipo

$$\int \frac{\chi_{\alpha}^*(\vec{r}_A) \chi_{\alpha'}(\vec{r}_A)}{|\vec{r}_B - \vec{r}_A|} d\tau_A \quad ;$$

cálculo este que se revela extremamente longo e dispendioso .

Após os trabalhos de London, foram realizados numerosas tentativas para determinar a energia eletrostática, utilizando no lugar da distribuição de carga molecular um conjunto de múltiplos convenientemente escolhido e aplicando as fórmulas clássicas da eletrostática (ver por exemplo Ref. 32 e referências aí mencionadas). Diferentes autores recorreram a desenvolvimentos multipolares a partir de funções de onda para o cálculo de interações moleculares: Amos e Crispin³³ utilizam orbitais localizados SCF para calcular via um tal desenvolvimento , a interação de duas moléculas de metano ou de duas moléculas de água. Mulder e Huiszoon³⁴ empregam este método para o cálculo de interação de moléculas orgânicas e comparam os resultados assim obtidos com resultados baseados em potenciais átomo-átomo. Rein et al.³⁵⁻³⁷ lançam mão de expansões multipolares a partir de funções de onda moleculares IEHT ou CNDO, para o cálculo da contribuição eletrostática das energias de interação envolvendo as bases púricas e pirimídicas. Pullman e Perahia³⁸ mostraram que os esquemas de hidratação das bases pirimídicas obtidos a partir de uma soma de termos de interação multipolo-multipolo , são aptos para reproduzir corretamente as características principais dos esquemas de hidratação destes sistemas, obtidas no quadro do método *ab initio* supermolecular. A expansão utilizada consiste em desenvolver em multipolos atômicos e de recobrimento (até os termos quadrupolares) as distribuições contínuas *ab initio* SCF de densidade eletrônica das moléculas individuais.

A abordagem que nós adotamos neste trabalho para calcular as grandezas teóricas que serão analisadas nos nossos estudos macromoleculares (Capítulos 3 e 4) é a mesma empregada por estes últimos autores.

A fim de indicar como o desenvolvimento multipolar é obtido a partir da função de onda molecular nós vamos expor em seguida os aspectos principais deste procedimento.

Reescrevendo a expressão que nós encontramos para a densidade total de carga eletrônica para a molécula A:

$$\rho_A(\vec{r}) = 2 \sum_{\mu=1}^N \sum_{\alpha} \sum_{\alpha'} C_{\alpha}^{*\mu} C_{\alpha'}^{\mu} \chi_{\alpha}^*(\vec{r}) \chi_{\alpha'}(\vec{r})$$

podemos notar que ela nos leva a considerar dois tipos de distribuições:

- distribuições de um centro: os orbitais atômicos χ_{α} e χ_{α}^* , centrados sobre o mesmo átomo.
- distribuições de dois centros: são as distribuições do tipo $\chi_{\alpha}^* \chi_{\alpha'}$, onde χ_{α}^* e $\chi_{\alpha'}$ são centrados em dois núcleos diferentes.

Se definirmos $P_{\alpha\alpha'} = 2 \sum_{\mu=1}^N C_{\alpha}^{*\mu} C_{\alpha'}^{\mu}$, então

$$\rho_A(\vec{r}) = \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'} \chi_{\alpha}^*(\vec{r}) \chi_{\alpha'}(\vec{r}) \quad (2.41)$$

Sejam k e l os índices dos átomos sobre os quais estão centrados χ_{α} e $\chi_{\alpha'}$, considerando igualmente que, na prática, os orbi-

tais atômicos e seus coeficientes são reais. Podemos então reescrever a eq. (2.41)

$$\rho_A = \sum_k \sum_{\alpha \in k} \sum_{\alpha' \in k} P_{\alpha\alpha'}^{k k} \chi_{\alpha} \chi_{\alpha'} + 2 \sum_{k>l} \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k l} \chi_{\alpha} \chi_{\alpha'} \quad (2.42)$$

explicitando assim os dois tipos de distribuições: atômicas e de recobrimento. Desta maneira podemos expandir estas distribuições de carga em diferentes momentos multipolares:

distribuição atômica

monopolos: $q_k = Z_k - \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k k} \int \chi_{\alpha}^* \chi_{\alpha'} d\tau$

dipolos: μ_k com três componentes do tipo:

$$\mu_{k,x} = - \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k k} \int \chi_{\alpha}^* x \chi_{\alpha'} d\tau$$

quadrupolos: Q_k com nove componentes:

$$Q_{k,xy} = - \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k k} \int \chi_{\alpha}^* xy \chi_{\alpha'} d\tau$$

distribuição de recobrimento

$$q_{kl} = -2 \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k l} \int \chi_{\alpha}^* \chi_{\alpha'} d\tau$$

μ_{kl} de componente:

$$\mu_{kl,x} = -2 \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k l} \int \chi_{\alpha}^* x \chi_{\alpha'} d\tau$$

Q_{kl} de componente

$$Q_{kl,xy} = -2 \sum_{\alpha} \sum_{\alpha'} P_{\alpha\alpha'}^{k l} \int \chi_{\alpha}^* xy \chi_{\alpha'} d\tau$$

Para as distribuições de recobrimento, a origem da expansão multipolar é escolhida arbitrariamente; no nosso caso ela é tomada no centro do segmento que une os dois átomos.

Pode-se mostrar, utilizando as leis da eletrostática

clássica³⁹, que o potencial criado por um quadrupolo Q depende somente de dois parâmetros se este é levado à sua forma diagonal. Isto significa que a diagonalização da matriz de elementos Q_{xy} nos levará a somente dois auto-valores não nulos e desta maneira podemos tratar o quadrupolo considerando simplesmente dois quadrupolos axiais, o que simplifica o cálculo de propriedades eletrostáticas de moléculas a partir de procedimentos utilizando expansões multipolares.

Denotando por \vec{r} o vetor posição que une os dois multipolos i e j , e por \vec{u} o vetor unitário correspondente, as fórmulas utilizadas para as diferentes contribuições multipolares da energia de interação entre dois sistemas moleculares são as seguintes⁴⁰:

- monopolo - monopolo

$$E = \frac{q_i q_j}{r} \quad (2.43)$$

- monopolo - dipolo

$$E = - q_i \vec{\mu}_j \cdot \frac{\vec{u}}{r^2} \quad (2.44)$$

- dipolo - dipolo

$$E = \frac{1}{r^3} [\vec{\mu}_i \cdot \vec{\mu}_j - 3(\vec{\mu}_i \cdot \vec{u})(\vec{\mu}_j \cdot \vec{u})] \quad (2.45)$$

- monopolo - quadrupolo (quadrupolo axial \vec{a} , Q_a)

$$E = q \frac{Q_a}{2r^3} [3(\vec{a} \cdot \vec{u})^2 - 1] \quad (2.46)$$

- dipolo - quadrupolo

$$E = \frac{3Q_a}{2r^4} \left\{ 2(\vec{a} \cdot \vec{u})(\vec{a} \cdot \vec{u}) + [1-5(\vec{a} \cdot \vec{u})^2](\vec{u} \cdot \vec{u}) \right\} ; \quad (2.47)$$

nesta expressão, os vetores \vec{r} e \vec{u} devem ser orientados do quadrupolo para o dipolo.

- quadrupolo (\vec{a}_1, Q_1) - quadrupolo (\vec{a}_2, Q_2)

$$E = \frac{3Q_1 Q_2}{4r^5} \left\{ 1-5[(\vec{a}_1 \cdot \vec{u})^2 + (\vec{a}_2 \cdot \vec{u})^2] + 2(\vec{a}_1 \cdot \vec{a}_2) - \right. \\ \left. 20(\vec{a}_1 \cdot \vec{u})(\vec{a}_2 \cdot \vec{u})(\vec{a}_1 \cdot \vec{a}_2) + 35(\vec{a}_1 \cdot \vec{u})^2(\vec{a}_2 \cdot \vec{u})^2 \right\} \quad (2.48)$$

O potencial eletrostático de um sistema molecular pode ser então determinado a partir da soma de termos de interações, expressos pela eqs. (2.43), (2.44) e (2.46) quando substituímos nestas expressões o monopolo q por uma carga unitária. Cálculos teóricos foram realizados⁴¹ no sentido de comparar o potencial *ab initio* eq. (2.40) criado por uma molécula de água ao longo da bissetriz do ângulo $\widehat{H\hat{O}H}$, com o potencial criado pelo conjunto de multipolos descritos acima. Os resultados mostraram que a partir de uma distância de 2\AA do átomo de oxigênio, os potenciais *ab initio* e multipolares são praticamente idênticos, sendo do tipo 7s, 3p/3s a base de gaussianas utilizadas para a molécula de água. O tempo de cálculo para potenciais multipolares é aproximadamente 30 vezes mais rápido (considerando um computador IBM 370-168) que o cálculo do potencial *ab initio* para a mesma molécula⁴¹.

CAPÍTULO 3ASPECTOS REATIVOS DO ARN DE TRANSFERÊNCIAESPECÍFICO DA FENILALANINA: tARN^{Phe}3.1 - O tARN^{Phe} e sua Estrutura Molecular

A importância dos ácidos ribonucléicos de transferência (tARN) no metabolismo celular é indiscutível. Eles desempenham vários papéis nos processos bioquímicos e ocupam um lugar fundamental na síntese de proteínas. A sequência nucleotídica do ADN é inicialmente transcrita na sequência complementar do ARN mensageiro (mARN). Em seguida este dirige o conjunto dos amino-ácidos em uma sequência linear específica que caracteriza uma dada proteína, processo este denominado translação. As moléculas do ARN de transferência agem como veículos para transferir os amino-ácidos do sítio livre, no interior dos ribossomos, para a cadeia de proteínas.

A fim de realizar estudos teóricos neste domínio, uma etapa essencial consiste em obter a estrutura detalhada da macromolécula. Entretanto, inúmeras dificuldades são encontradas na realização destes experimentos e somente foram superadas no estudo da estrutura tridimensional do ARN de transferência da fenilalanina do levedo "yeast tRNA^{Phe}". Para esta molécula, um cristal altamente ordenado foi obtido na presença de sais de magnésio, sódio e de poliaminas⁴²⁻⁴⁴.

As moléculas de tARN formam um grupo de polinucleotí-

deos pequenos e globulares (a serem distinguidos das cadeias fi br os encontradas no ADN e no mRNA) tendo aproximadamente 75 a 90 nucleotídeos distribuídos ao longo de uma estrutura muito com plexa.

Cada nucleotídeo FIGURA 3.1.1 é formado a partir da associação dos seguintes sistemas moleculares: uma ribose (açúcar), um ácido fosfórico (grupo fosfato), uma base nucléica *usual* (guanina, adenina, citosina, uracil) ou uma base nucléica modificada (também designadas *não usuais*). É importante realçar que a presença destas últimas constitui uma feição de todos os tARN. A maior parte destas bases não usuais se diferenciam das bases usuais pela adição de um ou mais grupos metilas sendo co mum denotar as regiões dos tARN em função de suas posições nas diversas partes da molécula FIGURA 3.1.2. Por exemplo, para o tARN^{Phe} o laço T ψ C é assim denominado visto que ele contém a timina*; o laço D contém a dihidrouracila FIGURA 3.1.3. Por ou tro lado, o laço do anticodon é assim designado porque ele con tém as três bases nucléicas que reconhecem o codon.

* Lembremos que a timina é característica dos ADNs aparecendo, entre toda a família dos ARNs, somente no ARN de transferência.

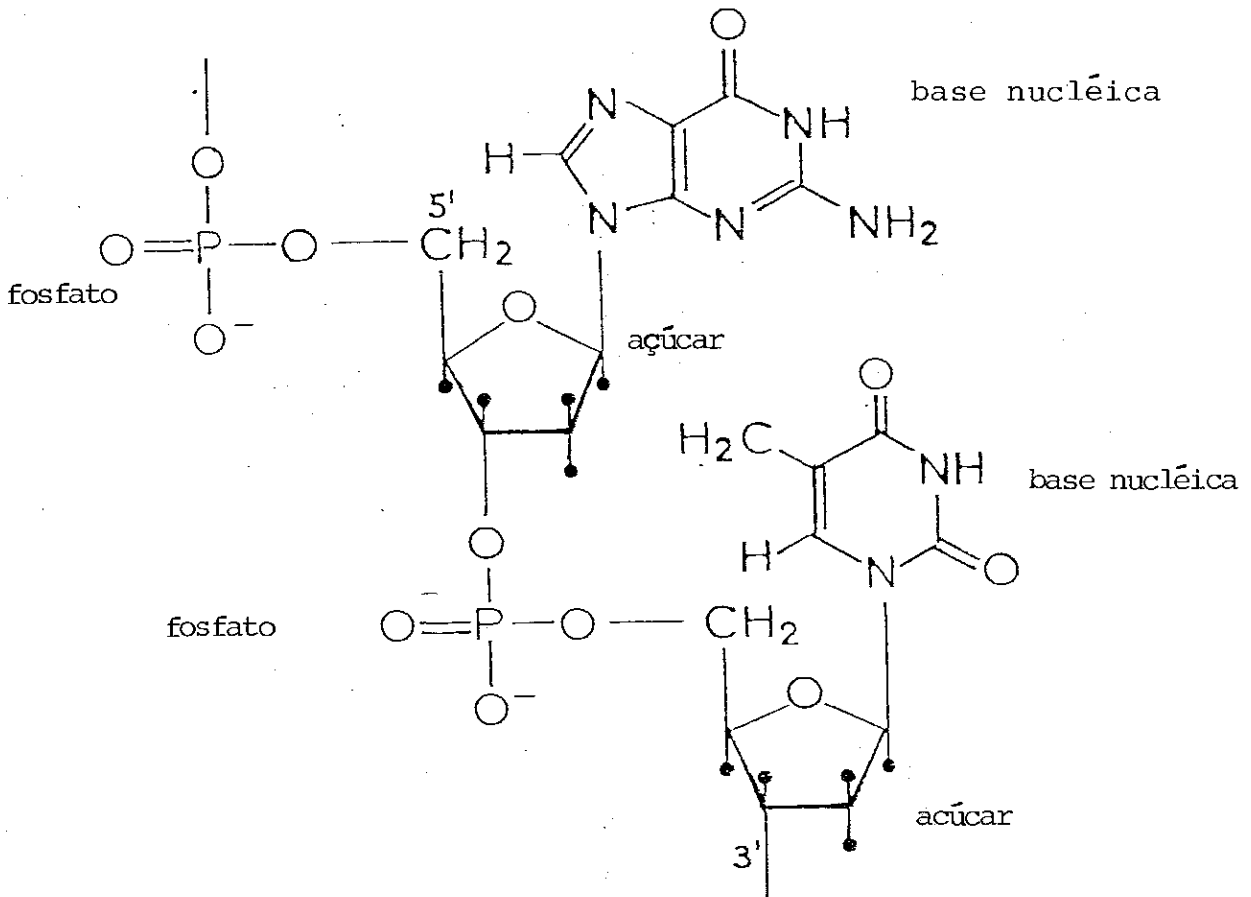


FIGURA 3.1.1 - Encadeamento dos nucleotídeos nos ácidos nucléicos

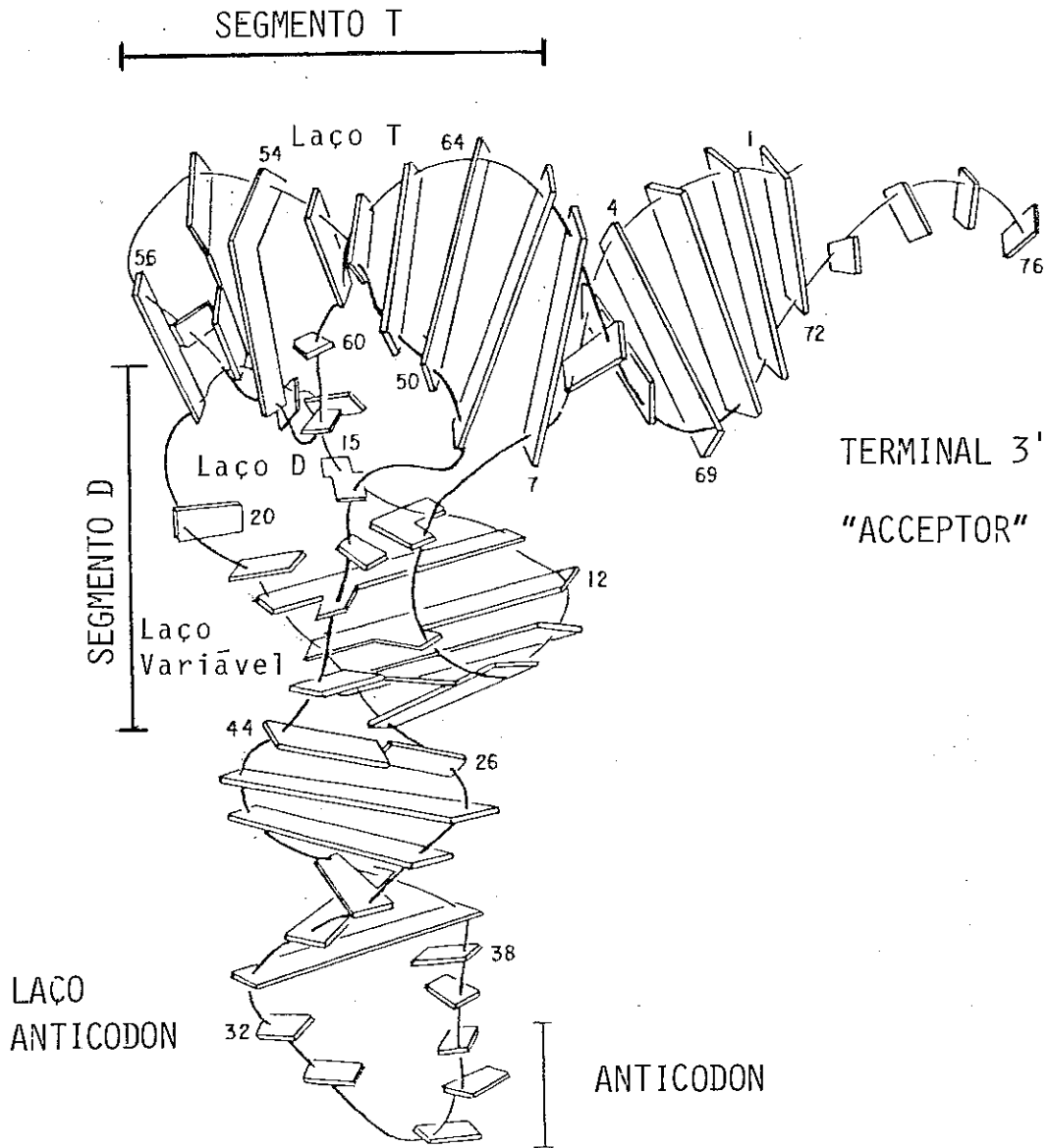


FIGURA 3.1.2 - Laços e segmentos em dupla-hélice do tARN^{Phe}. A figura indica igualmente a orientação de suas bases nucleicas.

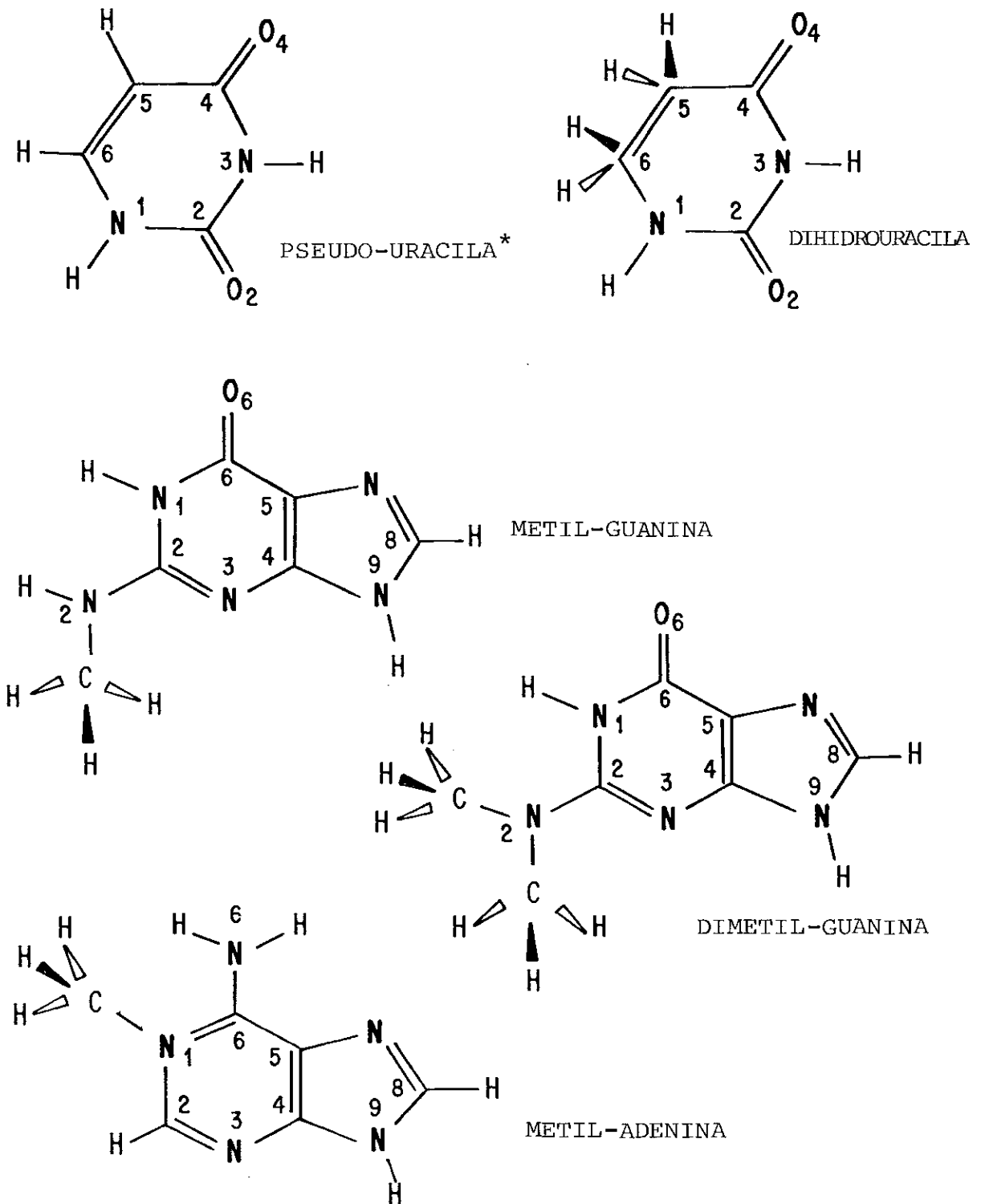


FIGURA 3.1.3 - Exemplo de Bases Nucléicas não usuais encontradas no tARN^{Phe}

* A PSEUDO-URACILA tem a mesma forma da URACILA mas interage com as adeninas do tARN^{Phe} através de pontes de hidrogênio que não são do tipo Watson-Crick (ver FIGURA 3.1.5)

A estrutura cristalina do tARN^{Phe} possui grosseiramente uma forma em L, os dois segmentos "acceptor" e TψC formam uma dupla-hélice contínua e os dois segmentos "D" e anti-codon formam outra dupla-hélice. As duas dupla-hélices constituem os dois braços do "L", cada um tendo um comprimento de aproximadamente 60 Å e um diâmetro tendo cerca de 20 Å. O terminal 3' "acceptor", onde se faz a elongação do peptídeo na cadeia proteica, encontra-se em uma das extremidades da molécula enquanto que o anti-codon que reconhece o codon sobre o ARN mensageiro encontra-se na outra extremidade.

É frequente considerar por razões de comodidade o diagrama chamado "folha de trevo" mostrado na FIGURA 3.1.4. Este diagrama apresenta a vantagem de mostrar mais claramente o encaimento e a natureza das bases, as partes helicoidais e não helicoidais da molécula e a natureza das interações entre as bases. Todos os pares de bases, encontrados nas *dupla hélices* são do tipo Watson Crick: adenina - uracil e guanina - citosina FIGURA 3.1.5 com exceção do par guanina 4 G(4) - uracil 69 U(69). Mas como é mostrado na FIGURA 3.1.4, as interações pontes de hidrogênio "padrões" não são as únicas a contribuírem para a estabilidade da macromolécula. Nas partes *não helicoidais*, algumas bases usuais e modificada participam em ligações hidrogênio não convencionais e são chamadas interações terciárias. Tais ligações podem ocorrer entre duas ou três bases como também entre uma base e o filamento fosfodiéster da cadeia macromolecular ou mesmo entre regiões diferentes do tARN^{Phe}^{42,43}.

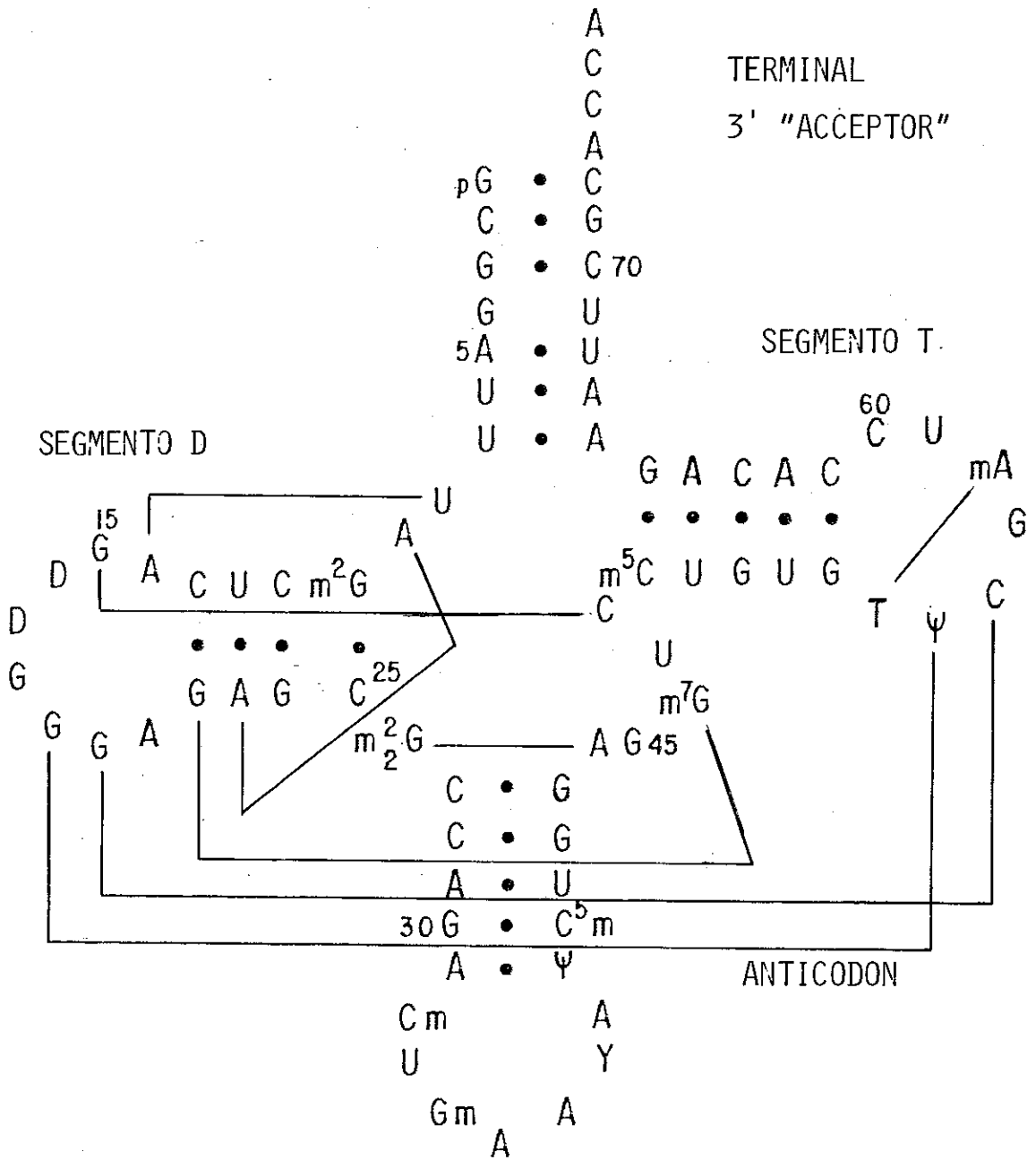


FIGURA 3.1.4 - Diagrama "Folha de Trevo" do tARN^{Phe}. Os traços contínuos indicam as interações terciárias, via pontes de hidrogênio, entre as bases nucleícas.

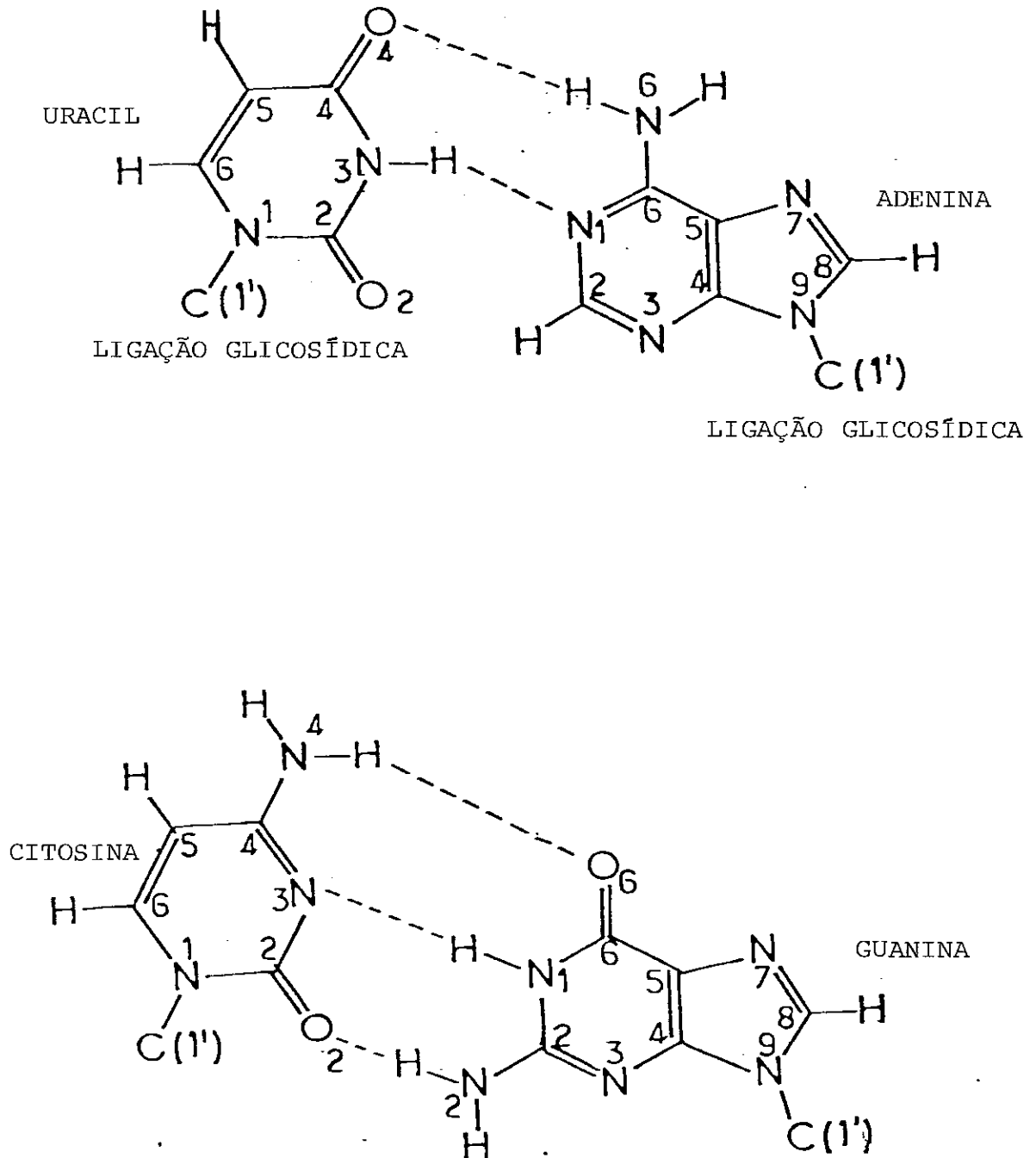


FIGURA 3.1.5 - Pares complementares Watson-Crick de bases nucleicas do tARN^{Phe}

Um grande número de pesquisas teóricas tem sido devotado ao estudo das interações envolvidas na estrutura tridimensional do tARN^{Phe}⁴⁵⁻⁴⁸.

3.2 - O Cálculo do Potencial Eletrostático de um Biopolímero

Considerando-se sistemas macromoleculares um problema deve ainda ser resolvido quanto ao cálculo de seus potenciais eletrostáticos; obter diretamente as funções de onda de tais sistemas é inviável em nossos dias mesmo com os computadores mais modernos. A técnica adotada para realizar nossos cálculos consiste em fragmentar a macromolécula em sub-unidades escolhidas de maneira adequada⁴⁹⁻⁵¹. No caso dos ácidos nucléicos estas sub-unidades são as bases, os grupos fosfatos e os açúcares, e as ligações inter-unidades que são seccionadas pela divisão são "saturadas" pela adição de um átomo de hidrogênio a cada um destes componentes. Desta maneira, torna-se possível calcular a função de onda *ab initio* SCF de cada sub-unidade e a partir das distribuições eletrônicas correspondentes obter o *desenvolvimento multipolar* como mostramos no Capítulo 2. O potencial eletrostático da macromolécula é então calculado superpondo os potenciais gerados pelos multipólos das sub-unidades, convenientemente orientadas segundo a geometria macromolecular.

Se a divisão da macromolécula é corretamente escolhida, a perturbação eletrônica introduzida nas sub-unidades pela divisão pode ser minimizada e o potencial obtido pela superposição pode reproduzir o potencial eletrostático macromolecular com

precisão em regiões não muito próximas das ligações inter-unidades⁴⁹⁻⁵¹.

Esta técnica foi amplamente aplicada no estudo das propriedades eletrostáticas de diferentes conformações de dupla-hélices de ácidos desoxiribonucléicos, com modelos comportando aproximadamente 500 átomos cada um^{7-9,52}.

A existência de uma estrutura cristalográfica para o tARN^{Phe} permite, então, estudos teóricos sobre suas propriedades reativas, mas a tarefa é considerável. Embora o tARN^{Phe} pertença a classe I que agrupa os menores tARN⁵³, ele é no entanto constituído de 76 nucleotídeos, ou seja aproximadamente 2500 átomos. É esta a razão pela qual o estudo desta macromolécula progrediu passo a passo, de modelos muito simplificados a modelos mais completos.

Em primeiro lugar foi calculado o potencial eletrostático molecular devido somente aos grupos fosfatos da macromolécula⁵⁴, sendo justificada esta simplificação pela carga negativa associada a cada um destes grupos. Como consequência, pode-se prever que os potenciais individuais gerados por cada grupo fosfato contribuem de maneira significativa para o potencial molecular do tARN^{Phe}.

Apesar de todos os fosfatos do tARN^{Phe} terem conformações diferentes - em termos dos ângulos de torção \underline{W} , C_5 , O_5 , $P - O_3$, e \underline{W}' , O_5 , $P - O_3$, C_3 , - FIGURA 3.2.1 - , esses estudos mostraram que estas conformações podiam ser classificadas em

10 grupos de conformações semelhantes. Em seguida, e com uma boa precisão, o potencial de cada um dos fosfatos em um mesmo grupo foi obtido a partir de somente uma função de onda, sendo esta calculada para um fosfato escolhido como representativo do grupo. A base atômica utilizada para o átomo de fósforo foi construída a partir da base de gaussianas (10s, 6p) proposta por Roos e Siegbahn⁵⁵ à qual foram adicionados seis orbitais d com o objetivo de fornecer uma representação mais apropriada deste átomo na cadeia fosfodiester⁵⁶.

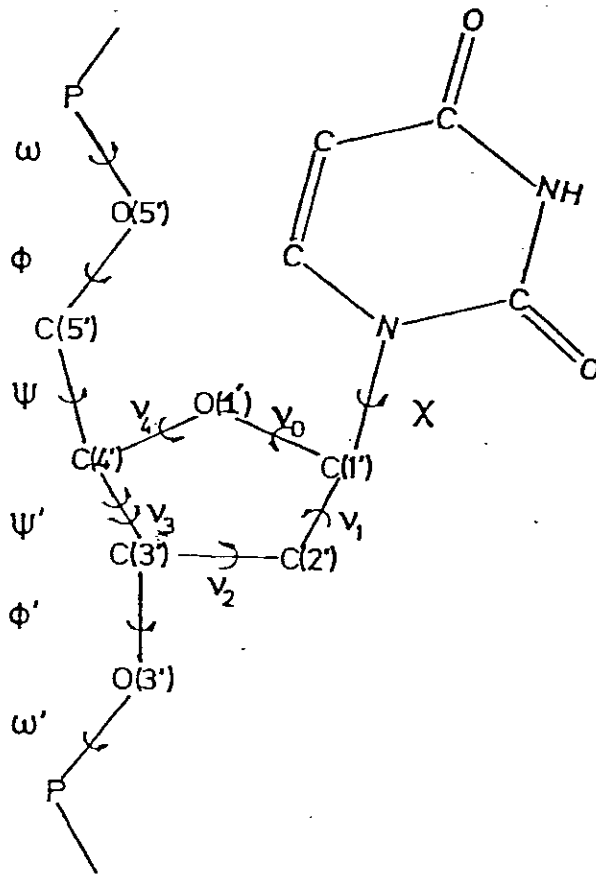


FIGURA 3.2.1 - Definição dos ângulos de torção da cadeia fosfodiéster

Os potenciais das diferentes conformações dos fosfatos isolados estão descritos em detalhe no Apêndice A desta te se sendo os seguintes os resultados relevantes desse trabalho:

- 1) Os cálculos *ab initio* SCF para as variações energéticas existentes entre os dez fosfatos representativos indicam que as conformações mais estáveis em energia (grupo 8 e grupo 10) correspondem aos grupos mais populosos. Os grupos 1 (fosfato 25) e 5 (fosfato 17 e 19) são particularmente desfavoráveis energeticamente e suas presenças no tARN^{Phe} são provavelmente devidas às *interações terciárias locais* que exigem tais conformações, ou seja as interações segmentos D (C25) - laço variável (G45) envolvendo o grupo 1 e laço D (G18,G19) - laço TψC (ψ55, C56) envolvendo o grupo 5.
- 2) As variações dos ângulos de torção fosfodiester tem um efeito desprezível sobre os *mínimos principais de potencial* localizados em uma posição equidistante de 2.15 Å ("bridge position") dos dois oxigênios aniônicos do grupo fosfato.
- 3) Estes efeitos são realçados em outras regiões em torno do grupo fosfato e de maneira notável ao longo da bissetriz do ângulo $O_3' \hat{P} O_5$, que em uma cadeia de polinucleotídeos é gros seiramente dirigida para as bases nucléicas circunvizinhas.

Os estudos realizados sobre o potencial global, devido à superposição dos potenciais de todos os fosfatos do tARN^{Phe}, mostraram que de maneira geral, os potenciais associados aos fosfatos nas regiões em dupla-hélice do tARN^{Phe} são mais atrativos de que aqueles localizados nas regiões dos laços, estes últimos estando mais afastados do centro da estrutura⁵⁴.

3.3 - Estudo dos Sítios Reativos das Bases Nucleicas do tARN^{Phe}

Para que uma reação possa ocorrer entre uma macromolécula e um reativo, é necessário que este último possa atingir estereoquímicamente um determinado sítio ou alvo da macromolécula. Portanto, que a um tal alvo esteja associado um potencial favorável à interação é uma condição necessária mas insuficiente. Em consequência, paralelamente ao estudo do potencial eletrostático, foram consideradas as *acessibilidades estereoquímicas* associadas aos átomos das bases estudadas. Elas serão apresentadas como *medidas quantitativas das áreas acessíveis* sobre os átomos em consideração (ou átomos receptores) em relação a uma esfera teste em contato com eles⁵⁷.

A primeira etapa para a obtenção destas áreas consiste em construir uma *superfície envelope* para a macromolécula: cada um de seus átomos é envolvido por uma esfera cujo raio é o raio de van der Waals do átomo. Os valores tomados para estes raios são os seguintes⁵⁸: H(1,2Å); C(1,6Å); N(1,5Å); O(1,4Å) e P(1,9Å). Em seguida, uma rede de N pontos uniformemente distribuídos é gerada sobre a superfície da esfera de van der Waals envolvendo o átomo receptor. A esfera atacante é colocada em contato com cada um destes pontos (da esfera do receptor) e para cada configuração é estudada a interseção entre esta esfera atacante com as esferas envolvendo os outros átomos da macromolécula. Verifica-se assim se a configuração é estereoquímicamente acessível dentro do quadro do modelo de esferas rígidas utilizado. Após o exame das N configurações e supondo ter encontrado entre elas N' acessíveis, a área acessível A (em Å²) do átomo

receptor é definida como:

$$A = \frac{N'}{N} 4\pi r^2$$

A fim de obter uma boa precisão no cálculo destas áreas nós tomamos 600 pontos para o valor de N; sendo o erro deste cálculo inferior a $0,1\text{\AA}$. O valor considerado para o raio da esfera atacante é $1,2\text{\AA}$ representando um átomo de hidrogênio.

No decorrer destes estudos, foi levado em conta, além de todas as sub-unidades da macromolécula, a influência dos quatro cátions de magnésio localizados no cristal do tARN^{Phe} sobre os potenciais e as acessibilidades dos sítios reativos das bases nucleicas.

Os resultados relativos a estes cálculos estão apresentados no Apêndice B desta tese e mostram essencialmente a individualidade dos potenciais e acessibilidades associadas aos sítios reativos de cada uma das bases do tARN^{Phe}.

Em um polinucleotídeo de *estrutura regular*, como, por exemplo o ADN-B, os potenciais associados a uma base nucleica qualquer são influenciados pela sequência de bases circunvizinhas, mas este efeito é relativamente fraco (geralmente da ordem de algumas Kcal/mole)⁵⁹. O efeito da sequência de bases sobre a acessibilidade é também pouco pronunciado.

No tARN^{Phe} a situação se apresenta de maneira diferente devido à *forma complexa da macromolécula*. Cada base é caracterizada pelo seu potencial médio, pela ordem dos potenciais e

acessibilidades associadas aos sítios das bases. As similaridades entre duas bases não são mais função do tipo de base, mas principalmente função da proximidade entre elas na estrutura tri dimensional da macromolécula.

A seguir apresentaremos uma discussão geral sobre os potenciais e acessibilidades associados aos sítios das bases e sua utilização na interpretação da reatividade destes componentes no interior da macromolécula.

Vários reativos, atacantes de sítios específicos das bases dos ácidos nucléicos, são conhecidos e foram realizados um número importante de trabalhos experimentais envolvendo o tARN^{Phe}. No presente estudo consideraremos quatro reativos ele trofílicos: o quetoxal, o carbodiimido, o 7-bromometil-antraceno e o AAAF.

Os potenciais e as acessibilidades, extraídos de nossos cálculos, associados aos sítios de ataque destes reativos estão apresentados graficamente nas FIGURAS 3.3.1, 3.3.2 e 3.3.3. Para cada uma destas FIGURAS, os potenciais são dispostos so bre o eixo vertical e as acessibilidades sobre o eixo horizontal.

As bases são indicadas pelo número de seus nucleolídeos (1-76) na cadeia polinucleotídica do tARN^{Phe}; os números no interior de um círculo correspondem às bases que não se emparelham com outras no interior desta macromolécula. Os sítios para os quais a acessibilidade calculada é nula não estão

indicados.

O estudo destas FIGURAS permite fazer as seguintes observações:

3.3.1 - O Quetoxal

Estudos experimentais⁶⁰⁻⁶¹ indicaram que este composto (β -etoxi- α -quetobutiraldeído) reage simultaneamente com os átomos N_1 e N_2 , sejam da guanina G20 ou bem da guanina G34 (ver FIGURA 3.1.5 para a localização destes sítios na guanina). Podemos notar pela FIGURA 3.3.1 que nossos cálculos indicam que as guaninas sujeitas a reações no tARN^{Phe} tem os sítios N_1 e N_2 acessíveis. Estes sítios são também acessíveis para somente uma das outras guaninas, G19, mas esta é emparelhada (com C56), ao contrário de G20 e G34, e podemos supor que isto diminua sua reatividade. É interessante constatar que G57, também não emparelhada, não reage com o quetoxal, o que correlaciona-se bem com a fraca acessibilidade desta base como indicado na FIGURA 3.3.1. Sendo o quetoxal um reativo eletrofílico, o potencial macromolecular deveria favorecer sua reação com G20 em relação a G34, mas isto só poderia ser verificado por observações experimentais mais detalhadas.

Como consequência da clara correlação entre a acessibilidade e a reatividade podemos prever que N_2 , sendo mais acessível que N_1 , deve ser provavelmente o *sítio de ataque inicial* (esta também foi a conclusão dos estudos quânticos de uma reação similar entre uma guanina e o glioxal⁶²).

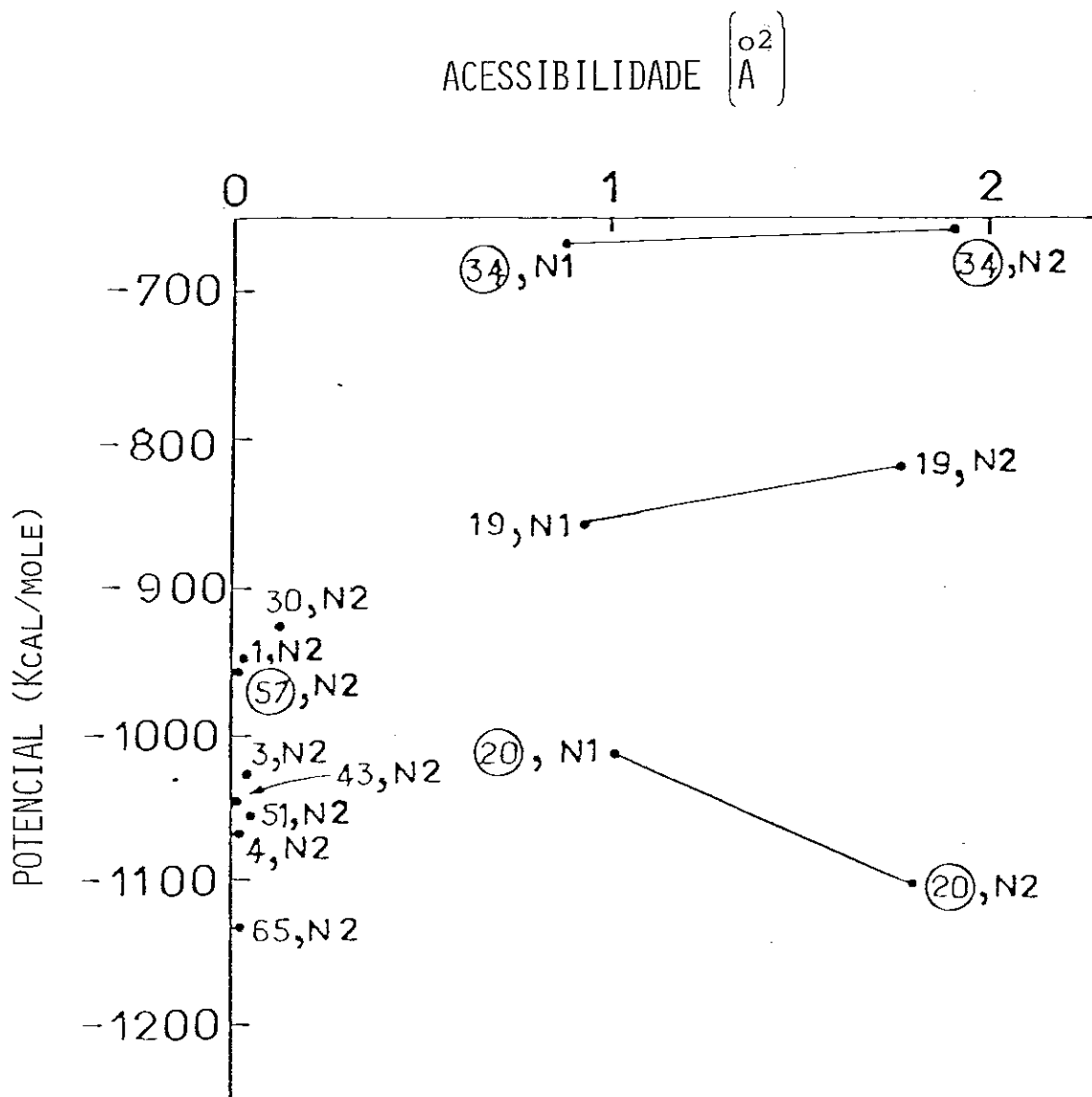


FIGURA 3.3.1 - Os potenciais e as acessibilidades dos sítios $N_1(G)$ e $N_2(G)$ do $tARN^{Phe}$

3.3.2 - O Carbodiimido

O carbodiimido (mais precisamente, 1-ciclohexil 3-[2-morfolino (4) - etil]carbodiimido) pode reagir com a guanina, uracila ou as uracilas modificadas - pseudouracila PU e dihidrouracila DU (ver FIGURA 3.1.3) - atacando N_1 no caso da guanina e N_3 no caso destas últimas⁶³. Com a guanina, tais reações foram observadas para G20 e G34, as mesmas bases potencialmente reativas com o quetoxal, reação esta que como já indicamos implica também o sítio N_1 . Podemos assim, aplicar a mesma interpretação para a reatividade destas guaninas calcada nos resultados da FIGURA 3.3.1.

Com respeito à uracila e as uracilas modificadas, três bases foram observadas experimentalmente reagindo com o carbodiimido: U47, DU16 e U33 (moderadamente). Os resultados da FIGURA 3.3.2 mostram que estas três bases, com DU17, são as únicas que apresentam o átomo N_3 acessível. A razão pela qual DU17 não reaja, pode ser atribuída ao seu potencial em N_3 , o menos atrativo das quatro bases acessíveis. É interessante notar que U33, reagindo moderadamente, tem também um potencial menos atrativo que aqueles associados às duas bases que reagem fortemente.

3.3.3 - O 7-bromometil-antraceno

Este composto reage com as guaninas do tARN^{Phe}⁶⁴. Os resultados dos estudos experimentais realizados indicam que uma das guaninas 18,19,20,22,24 (ver FIGURA 3.1.4) foi alterada, mas sua localização exata na estrutura molecular não pode ainda ser

determinada. Este reativo ataca principalmente o átomo N₂ da guanina⁶⁵ e de maneira análoga à discussão sobre o quetoxal, podemos notar através da FIGURA 3.3.1 as acessibilidades e os potenciais associados aos átomos N₂ de G19, G20 e G34. Tratando-se de um reativo eletrofílico, a base associada ao potencial mais atrativo deveria ser a mais sensível ao ataque, o que nos leva a predizer que a reação se desenvolverá com G20.

3.3.4 - O N-2-Acetoxi-2-Acetilaminofluoreno (AAAF)

A reação do AAAF com a guanina⁶⁶⁻⁶⁷ implica principalmente o átomo C₈ desta base nucléica⁶⁸. A ou as posições das guaninas sensíveis do tARN^{Phe} não foram ainda determinadas, mas o estudo de dois outros tARN, o tARN^{met}⁶⁶ o tARN^{tyr}⁶⁷ permitiram observar reações deste composto com G20 e com G18 e G19 respectivamente. Entretanto, como o número de nucleotídeos nestes ácidos ribonucléicos, e, particularmente, nas regiões dos laços D, não se correspondem com aquele do tARN^{Phe} uma previsão dos sítios reativos neste último torna-se difícil.

Os resultados teóricos, apresentados na FIGURA 3.3.3, indicam que os sítios C₈ mais acessíveis são aqueles de G1, G20 e G34. A reação com G1 é provavelmente desfavorecida visto a sua participação na dupla-hélice que constitui o segmento "acceptor". Entre as duas outras guaninas, G20 deve ser a mais reativa em razão do potencial mais favorável associado a esta base. Ela está situada no segmento "D" e ocupa uma posição similar àquela das guaninas reativas nos dois outros tARN estudados.

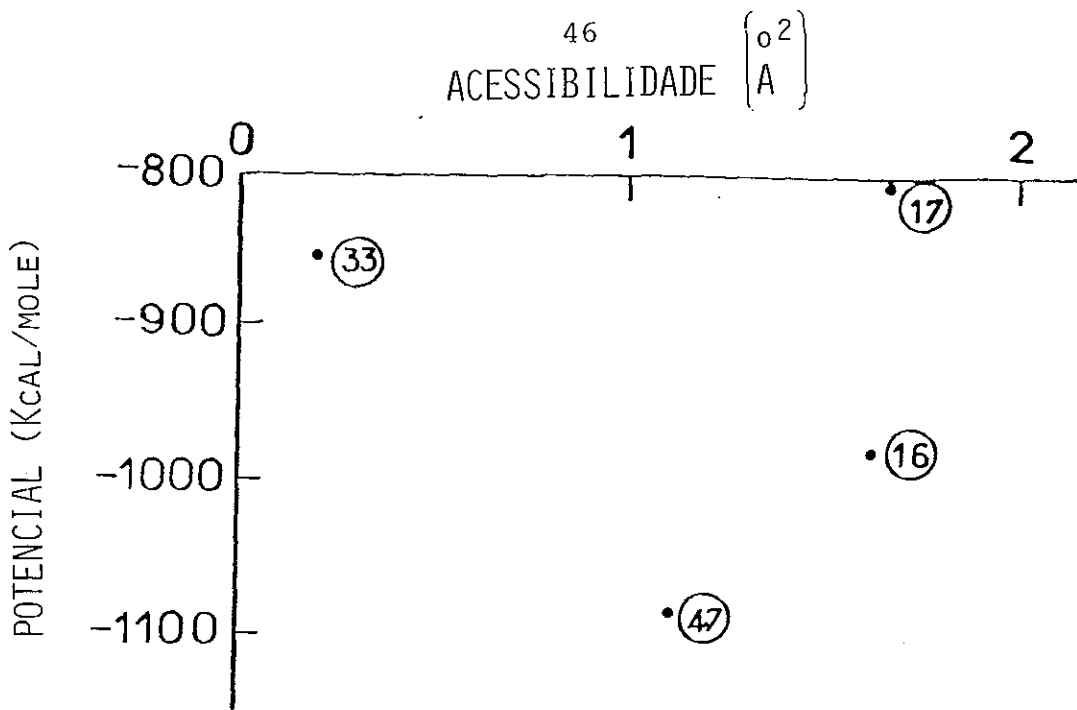


FIGURA 3.3.2 - Os potenciais e acessibilidades do sítios $N_3(U)$, $N_3(DU)$ e $N_3(PU)$ do $tARN^{Phe}$

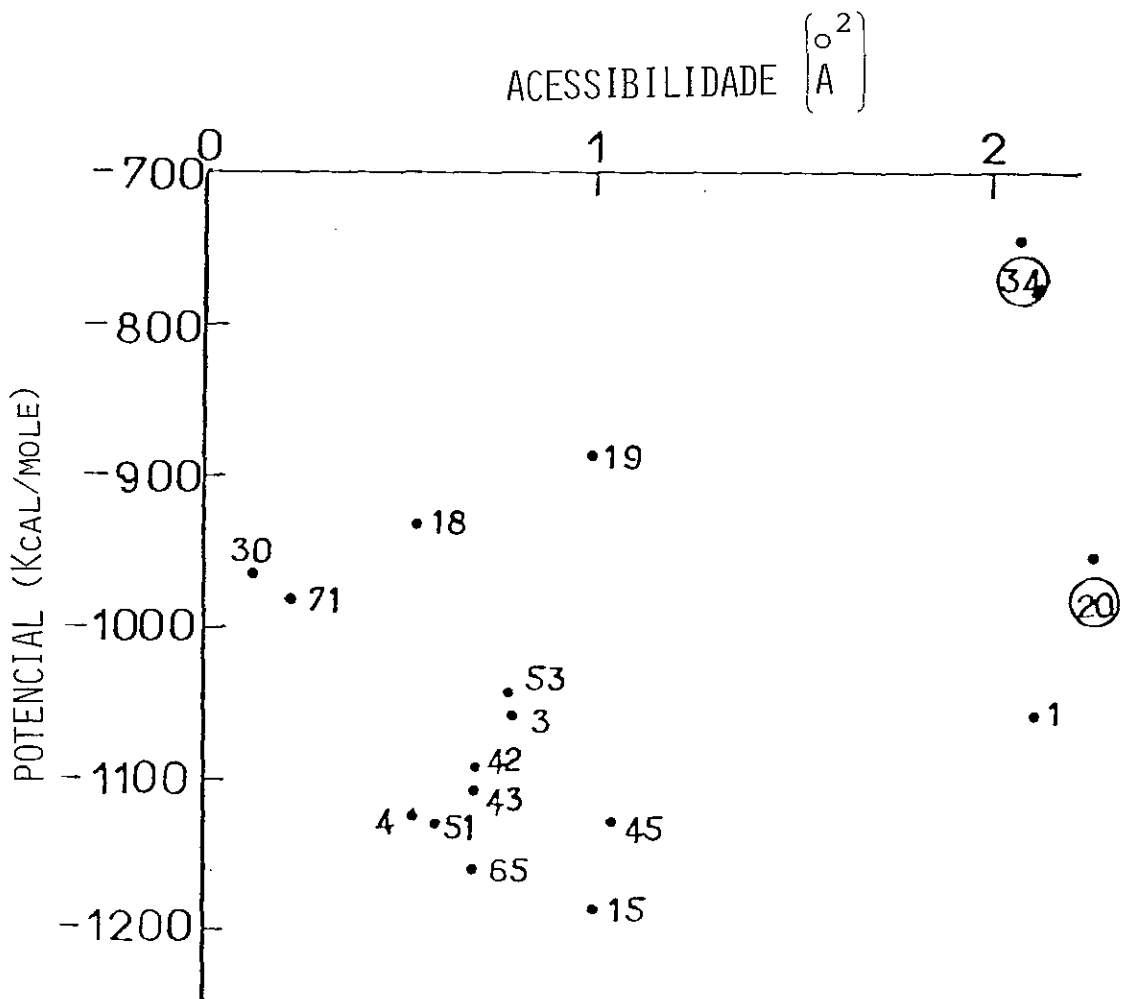


FIGURA 3.3.3 - Os potenciais e as acessibilidades do sítios $C_8(G)$ do $tARN^{Phe}$

Os cálculos teóricos apresentados aqui poderão ser utilizados para interpretar outras observações experimentais com respeito às reatividades específicas do tARN^{Phe}. Não obstante, alguns fatores experimentais podem influir na probabilidade de obter uma boa correlação:

- a) para que uma correlação seja obtida, é preciso que a *estrutura cristalográfica* utilizada para os cálculos teóricos e a *estrutura em solução* do tARN^{Phe} sejam praticamente as mesmas; os estudos experimentais foram realizados sob condições diferentes o que pode influenciar a estrutura da macromolécula. Além disso, não deve ser excluído o fato de que as reações elas mesmas possam engendrar modificações estruturais importantes. Tendo em vista que as acessibilidades são calculadas levando em conta a geometria molecular detalhada na vizinhança do sítio estudado, é provável que esta grandeza seja mais afetada que os potenciais em função de pequenas variações estruturais.
- b) para confirmar as reatividades relativas dos sítios de tARN^{Phe}, seria importante que cada tipo de sítio pudesse ser estudado levando em consideração uma variedade de reativos diferentes e que os resultados fossem coerentes entre si. Este estudo poderia também dissipar as dúvidas quanto às possíveis mudanças estruturais da macromolécula devido aos reativos. Neste sentido, até o momento, tais informações não estão disponíveis na literatura deste domínio.

Tendo em mente estes fatores, é no entanto estimulante observar que para os reativos considerados, os resultados te

óricos apresentados para a reatividade das bases nucleicas est
dados indicam uma correlação satisfatória com aqueles obtidos
experimentalmente.

CAPÍTULO 4HIDRATAÇÃO DE SISTEMAS BIOMOLECULARES:ALGUMAS APLICAÇÕES TEÓRICAS4.1 - Hidratação do Glioxal, Metil-Glioxal e Outros AldeídosSimples

Está bem estabelecido que o glioxal e alguns de seus derivados reagem com diferentes tipos de ácidos nucleicos, dando origem a consequências biológicas importantes. Estes compostos são capazes de tornar inativo o ARN do vírus do mosaico do tabaco⁶⁹ ou o ARN^{Phe}⁷⁰ e de desnaturar o ADN^{71,72}. E como assinalou Albert Szent-Györgyi eles podem estar igualmente implicados nas atividades carcinogênicas devido ao seu envolvimento no controle da divisão celular (o metil-glioxal, por exemplo)^{73,74}.

Diversas investigações teóricas tem sido realizadas no sentido de estudar as interações destas moléculas com os constituintes dos ácidos nucleicos^{62,75-76}.

Tendo em vista a importância da presença do glioxal e do metil-glioxal nas concentrações celulares, nossos estudos tiveram como propósito mostrar alguns aspectos da reação destes com moléculas de água para formar gem-diols (FIGURA 4.1.1). Estudos experimentais indicam que o glioxal e o metil-glioxal, bem como a molécula do formaldeído, formam hidratos estáveis; o mesmo fenômeno entretanto não é observado para a molécula do ace -

aldeído e aldeídos de ordem superior⁷⁷⁻⁷⁸

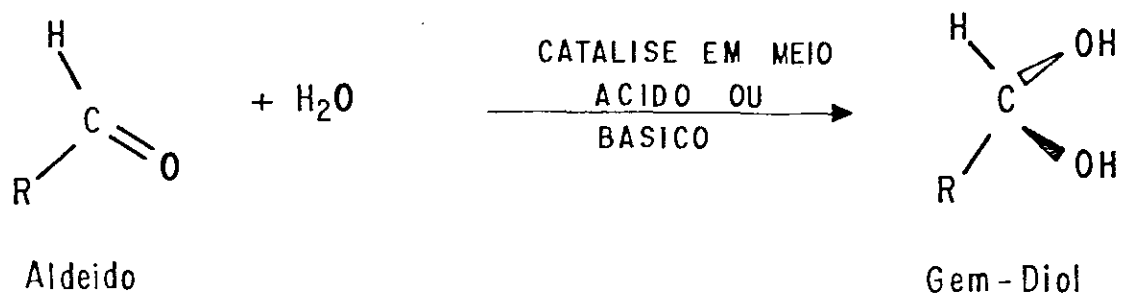
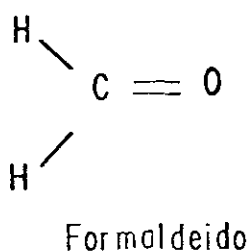


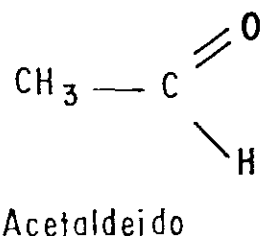
FIGURA 4.1.1 - Reação de um aldeído com uma molécula de água

Objetivando compreender as características da reatividade desta série de moléculas consideramos para o radical R os seguintes grupos:

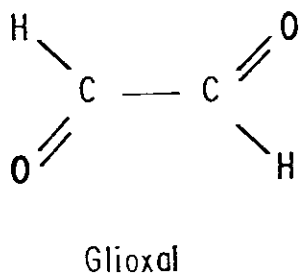
(a) - H



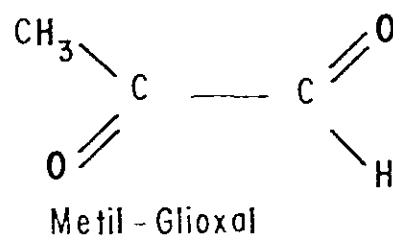
(b) - CH₃



(c) - CHO



(d) - COCH₃



O mecanismo de reação que estudamos se baseia em conclusões experimentais sobre a hidratação do formaldeído catalizada em um meio ácido. Esta reação envolve uma concorrência entre um próton e uma molécula de água no sentido de reagirem com seu grupo carbonila⁷⁹.

Nossos cálculos foram realizados através do método *ab initio* SCF, considerando o conjunto *substrato-solvente* como uma supermolécula, e resolvendo assim as equações de Roothaan (ver expressão 15 do Capítulo 2) para este sistema. As bases atômicas utilizadas são gaussianas do tipo STO-3G⁸⁰ para os cálculos de otimizações de geometrias e do tipo 7s 3p/3s para os cálculos das energias resultantes dos complexos formados.

Os resultados estão apresentados no Apêndice C; salientaremos aqui a sistemática que adotamos para abordar o problema estudado assim como discutiremos de forma abreviada os resultados essenciais encontrados.

Primeiramente investigamos as estruturas dos complexos de transição entre a molécula do formaldeído e uma molécula de água. Como foi sugerido pelos estudos experimentais sobre a cinética desta reação⁸¹, consideramos o átomo de oxigênio da molécula de água se ligando diretamente ao átomo de carbono do grupo carbonila do formaldeído. Esses estudos mostraram uma distorção do formaldeído, levando a uma estrutura tetraedral, a medida que a molécula de água se aproxima desse composto (FIGURA 4.1.2).

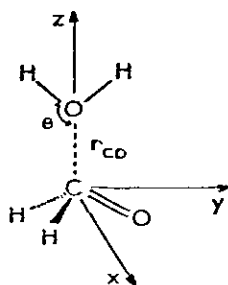


FIGURA 4.1.2 - Geometria do complexo água-formaldeído

Na prática, os complexos de transições são distorcidos progressivamente até a forma final de equilíbrio. A fim de simplificar o estudo destas distorções consideramos somente quatro conformações da molécula do formaldeído, que foram designadas por I, II, III e IV. A conformação I corresponde à forma natural da molécula (plana), a conformação IV apresenta esta molécula com a geometria típica de uma estrutura tetraedral e as conformações II e III tem geometrias de distorção intermediárias entre as conformações I e IV. Assim, os ângulos formados entre as ligações do átomo de carbono com os átomos restantes da molécula do formaldeído (o eixo z é definido seguindo a FIGURA 4.1.2), para as conformações I, II, III e IV, são respectivamente: 90° , 96.5° , 103° e 109.5° .

Permitimos à molécula de água se aproximar ao longo do eixo z e procuremos para cada conformação estudada a distância r_{CO} que correspondesse ao mínimo energético destes complexos. Entretanto, estes resultados indicaram que a energia do complexo torna-se instável quando se aproximam as duas moléculas

uma da outra.

Prosseguimos então estes estudos no sentido de investigar uma protonação inicial do átomo de oxigênio do formaldeído e, em seguida, uma complexação destes compostos protonados com uma molécula de água.

Os resultados mostraram que a protonação do formaldeído é favorável energeticamente e adotamos o seguinte esquema para investigar a complexação deste sistema protonado com uma molécula de água.

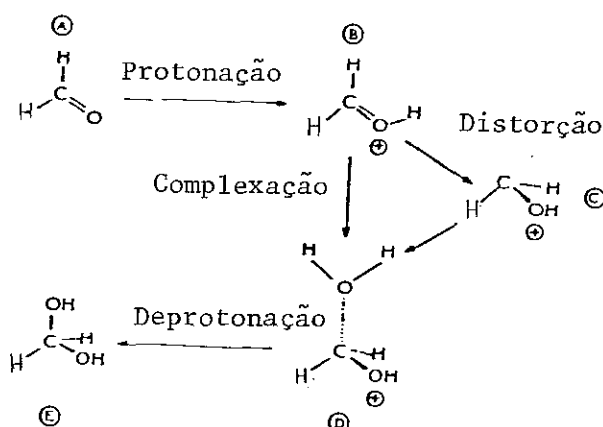


FIGURA 4.1.3 - Etapas utilizadas para estudar a hidratação do formaldeído catalizada em um meio ácido

A etapa deste esquema, envolvendo uma distorção, implica as quatro conformações que apresentamos previamente e as geometrias dos complexos encontrados, em cada uma destas etapas, foram otimizadas com o objetivo de determinar seus mínimos energéticos.

Este estudo indica que, através deste mecanismo, podem ser formados complexos estáveis do formaldeído com uma molécula de água. Desta maneira, parece plausível concluir que uma protonação inicial do grupo carbonila do formaldeído atua no sentido de permitir sua complexação com a molécula do solvente.

Aplicamos a mesma sistemática para o estudo da hidratação do acetaldeído, glioxal e metil-glioxal. Procedemos a uma análise detalhada das energias resultantes de cada uma das etapas envolvidas nas reações - como ilustradas para o formaldeído FIGURA 4.1.3 - assim como da energia total de complexação. Os resultados indicam que hidratos estáveis podem ser formados a partir da complexação do glioxal e do metil-glioxal com moléculas de água. Entretanto, para o caso do acetaldeído, os aspectos energéticos envolvidos neste mecanismo de reação desfavorecem a formação de complexos estáveis desta molécula em solução aquosa.

4.2 - Interações do ADN-B e seus Constituintes com Moléculas de Água: Uma Abordagem Eletrostática

4.2.1 - Caracterização da Flexibilidade de uma Molécula de Água em um Determinado Sítio de Interação

Como foi apresentado no Capítulo 2, expressaremos a energia de interação intermolecular como uma soma de termos de interação entre os multipolos do substrato e da molécula de água. Para cada uma destas moléculas, suas distribuições contí

nuas *ab initio* SCF de densidade eletrônica, são então desenvolvidas em multipolos atômicos e de recobrimento.

No desenvolver deste procedimento nos propusemos, não somente estabelecer a posição de mínimo de energia da molécula de água, mas igualmente caracterizar sua flexibilidade, investigando rigorosamente todas as orientações possíveis desta molécula nos seus sítios de interação com o substrato.

A técnica adotada para colocar a molécula de água em contato com a molécula receptora, é similar àquela apresentada no Capítulo 3, para o estudo das acessibilidades estereoquímicas dos sítios reativos de sistemas macromoleculares. O primeiro passo consiste em envolver a molécula de água e o substrato com envelopes - formados pela superposição de esferas de van der Waals - centrados sobre todos os átomos que constituem as duas moléculas. Dois átomos são colocados em contacto e nós chamamos átomo receptor, o átomo da molécula do substrato onde liga-se uma molécula de água. Esta última pode estar em contato via seu átomo de oxigênio ou um de seus átomos de hidrogênio, que serão designados átomos atacantes.

Impusemos um raio de 1Å^{O} para os átomos atacante e receptor o que significa um limite de aproximação entre eles de 2Å^{O} . Esta mesma distância foi imposta aos átomos vizinhos do receptor, susceptíveis de formar pontes de hidrogênio com a molécula de água. A escolha deste limite mínimo foi calcada nos resultados de trabalhos anteriores sobre os esquemas de hidratação de sistemas biomoleculares³⁸. Estes esquemas foram obtidos

utilizando a aproximação eletrostática para a energia de interação e comparados com aqueles resultando da aplicação do método *ab initio* SCF para a supermolécula substrato-molécula de água. Para os outros átomos o limite mínimo foi considerado como a soma de seus raios de van der Waals.

Podemos variar a orientação da molécula de água girando a esfera envolvendo o átomo atacante sobre a esfera do átomo receptor; as posições onde ocorrem interseções entre os envelopes das duas moléculas são consideradas inacessíveis.

As posições exatas, estudadas para um par de átomos em contato, são determinadas a partir da construção de uma rede de pontos uniformemente distribuídos sobre as *esferas dos átomos atacantes e receptores*. Esta rede de pontos é obtida pela técnica de Korobov (ver por exemplo Ref. 82), que pode gerar sobre a superfície das esferas um número de pontos correspondendo aos números da sequência de Fibonacci: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610 ... obtidos pela seguinte relação de recorrência:

$$a_{n+1} = a_n + a_{n-1} \quad ; \quad a_0 = 1 \quad , \quad a_1 = 1$$

Impõe-se às duas esferas contatos nestes pontos. Visando estudar o máximo de configurações possíveis, foram também consideradas rotações da molécula de água em torno de um eixo ligando os centros dos dois átomos em contato para cada par desses pontos. Os nossos testes indicaram, para diferentes números de configurações estudados, que a flexibilidade da molécula

de água torna-se estável a partir de 144 pontos sobre a esfera do receptor, combinados com 89 pontos sobre a esfera do atacante e 24 passos de rotações. Mostraremos os resultados desses testes, através de um exemplo, no fim desta seção.

Podemos notar que os números de pontos sobre as esferas envelopes e os passos de rotações adotados, nos conduz ao estudo de aproximadamente 310.000 configurações de uma molécula de água em contato com um só átomo receptor. Embora este número seja reduzido de maneira considerável proibindo a intersecção dos dois envelopes, sobram no entanto investigar entre 50.000 a 100.000 configurações para um átomo receptor relativamente acessível. Tendo em vista o tempo de cálculo computacional, proibitivo para as energias de interação eletrostática deste número de configurações, mesmo utilizando as expansões multipolares, tornamos este problema através dos seguintes critérios:

I) Primeiramente, a molécula de água é representada simplesmente por um dipolo μ e um quadrupolo Q , obtido a partir da sua função de onda *ab initio* SCF, colocados no baricentro desta molécula. Desta maneira, a energia de interação de uma configuração é estimada inicialmente pela seguinte expressão simplificada.

$$E_{el} = E_{\mu} + E_Q = -\vec{\mu} \cdot \vec{\epsilon} - \frac{1}{2} Q(\vec{a} \otimes \vec{a}) : \text{grad } \vec{\epsilon}$$

Os termos E_{μ} e E_Q indicam respectivamente a energia do dipolo $\vec{\mu}$ e do quadrupolo Q no campo eletrostático local $\vec{\epsilon}$ gerado pelo substrato; o símbolo $:$ indica um produto tensorial contraí

do duas vezes. Se expressarmos por \vec{r} a distância entre os multipolos do substrato e o ponto onde o campo é calculado e por \vec{u} o vetor unitário correspondente, os campos eletrostáticos e seus gradientes são escritos como:

- campo e gradiente do campo gerado por um monopolo q :

$$\vec{\varepsilon}_q = q \frac{\vec{u}}{r^2} ;$$

$$\overline{\text{grad}} \vec{\varepsilon} = \frac{q}{r^3} [1 - 3(\vec{u} \otimes \vec{u})]$$

- campo e gradiente do campo gerado por um dipolo :

$$\vec{\varepsilon}_\mu = \frac{1}{r^3} [\vec{\mu} - 3(\vec{\mu} \cdot \vec{u})\vec{u}] ;$$

$$\overline{\text{grad}} \vec{\varepsilon}_\mu = \frac{3}{r^4} [(\vec{u} \otimes \vec{\mu}) + (\vec{\mu} \otimes \vec{u}) + (\vec{u} \cdot \vec{\mu})(1 - 5\vec{u} \otimes \vec{u})]$$

- campo e gradiente do campo gerado por um quadrupolo axial (\vec{a}, Q_a) :

$$\vec{\varepsilon}_Q = - \frac{3}{2} \frac{Q_a}{r} \left\{ 2(\vec{a} \cdot \vec{u})\vec{a} + [1 - 5(\vec{a} \cdot \vec{u})^2] \right\}$$

$$\overline{\text{grad}} \vec{\varepsilon}_Q = - \frac{Q}{r^5} \left\{ 3 \vec{a} \otimes \vec{a} - 15(\vec{u} \cdot \vec{u}) [(\vec{a} \otimes \vec{u}) + (\vec{u} \otimes \vec{a})] \right. \\ \left. + \frac{3}{2} (1 - 5(\vec{a} \cdot \vec{u})^2) \mathbb{1} - \frac{15}{2} [1 - 7(\vec{a} \cdot \vec{u})^2] \vec{u} \otimes \vec{u} \right\}$$

Em vez de calcular o campo eletrostático e seu gradiente nas posições de cada um dos baricentros das configurações,

adotamos o seguinte procedimento: a partir de uma rede de pontos distribuídos em torno do átomo receptor, determinamos o ponto desta rede o mais próximo do baricentro da molécula de água para calcular essas energias de interação. Os pontos são tomados ao longo de vetores radiais - tendo como origem o centro do átomo receptor - e passam pelos pontos de Korobov correspondendo às configurações acessíveis da molécula de água ligada a este átomo. Este procedimento nos permite assegurar que existe no menos um ponto distante, alguns décimos de \AA no máximo, do baricentro de cada uma das configurações da molécula de água estudada. Desta maneira, consideramos todas as configurações acessíveis e selecionamos as 500 configurações energeticamente mais estáveis.

II) Levamos em conta novamente as configurações selecionadas acima, dentro da representação dipolar e quadrupolar da molécula de água. Neste estágio, determinamos a energia eletrostática entre as duas moléculas, considerando *todos os multipolos do substrato*. Foram escolhidas as 200 configurações de energia mais estáveis.

III) As energias de interação das configurações obtidas pelo segundo critério são recalculadas levando em conta todos os multipolos, atômicos e de recobrimento, da molécula de água e do substrato. A reorganização final das energias destas configurações permite então determinar a posição do mínimo de energia da molécula de água, ligada a um determinado sítio de interação.

Estes cálculos podem nos fornecer, além do mínimo energético da molécula de água, toda uma gama de configurações tendo energias muito próximas uma da outra. Desta maneira, nos foi possível caracterizar a flexibilidade da molécula de água representando, por meio de gráficos simples - para cada sítio estudado - todas as configurações que tem energias compreendidas em um intervalo de 1 Kcal/mole, a partir do mínimo de energia.

A FIGURA 4.2.1.1 ilustra um exemplo destes gráficos assim como os resultados obtidos quando foram utilizados três diferentes números de configurações da molécula de água. O sítio estudado é o átomo N₇ da guanina. Os números de configurações estão indicados no seguinte quadro:

FIGURA 4.2.2.1	número de pontos sobre o átomo receptor	número de pontos sobre o átomo atacante	passos de rotações	número de configurações
(a)	377	144	36	2×10^6
(b)	233	89	24	5×10^5
(c)	144	89	24	3×10^5

No lado esquerdo da FIGURA 4.2.1.1 indicamos as configurações da molécula de água no plano da guanina. Uma visualização em um plano perpendicular ao plano da guanina é mostrado no lado direito desta figura. Os seis diagramas e os outros que mostraremos depois foram obtidos através de um "plotter" associado a um micro-computador. A molécula de água é representada na sua posição de mínimo de energia, para cada um dos sítios característicos de interação, as ou

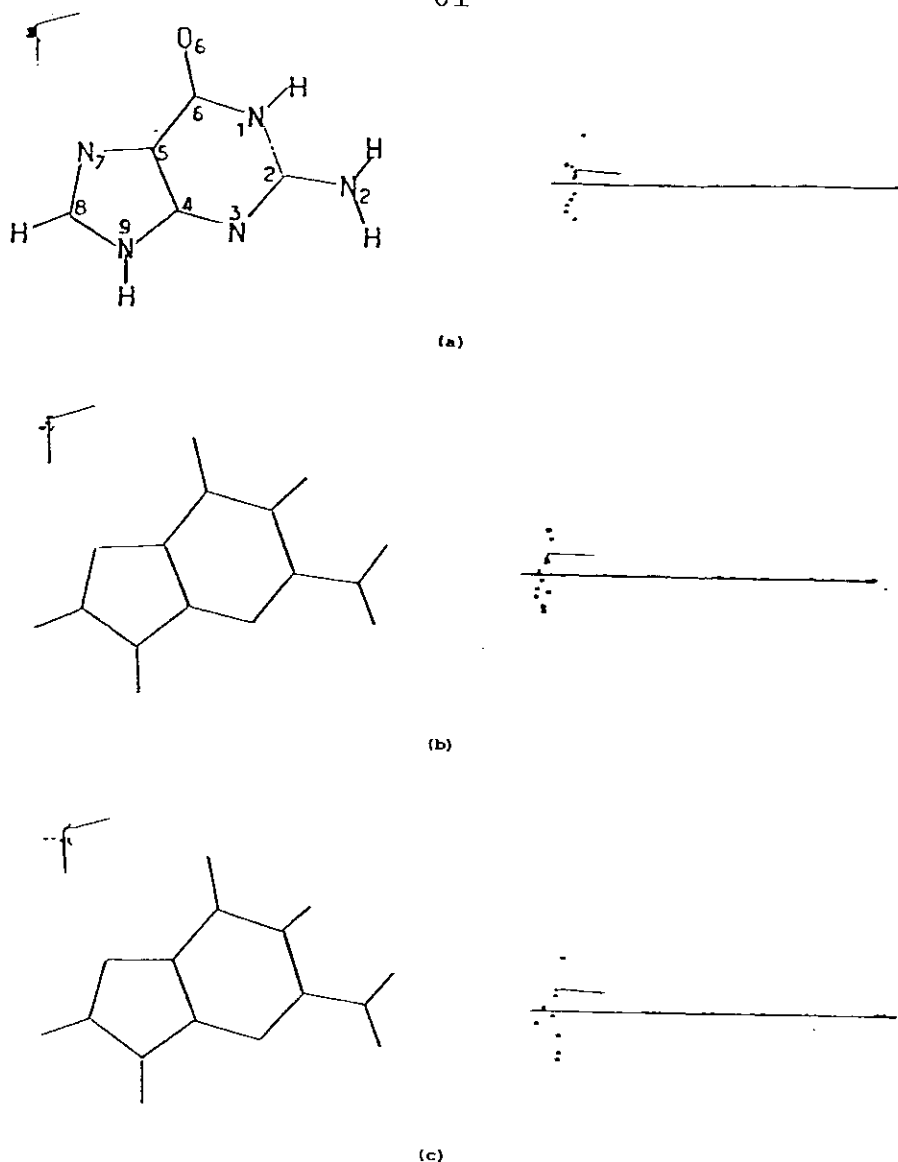


FIGURA 4.2.1.1 - Resultado dos testes realizados relativo ao número de configurações da molécula de água

tras configurações que pertencem à zona de interação observada foram indicadas por um ponto representando as posições dos átomos de oxigênios. Os diagramas da FIGURA 4.2.1.1 mostram que $\sim 3 \times 10^5$ configurações, são suficientes para reproduzir os resultados obtidos no caso de um número maior de configurações.

4.2.2 - Aplicações aos Constituintes dos Ácidos Nucléicos

A FIGURA 4.2.2.1 ilustra as zonas de hidratação da guanina. No diagrama central desta figura estão indicadas as zonas de interação visualizadas no plano da base. Estas regiões são numeradas na ordem decrescente - em valor absoluto - dos $m\bar{i}$

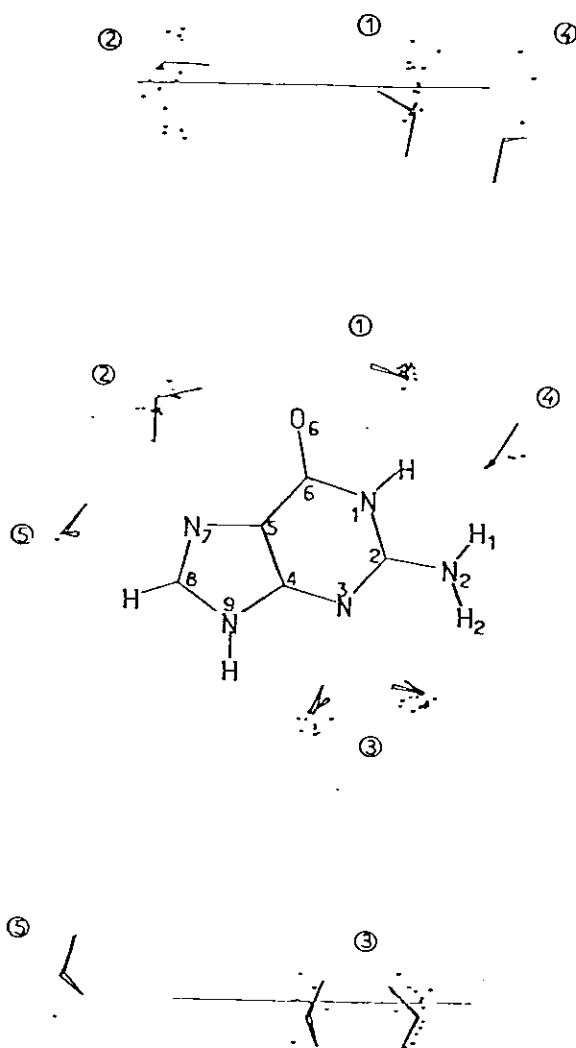


FIGURA 4.2.2.1 - Zona de hidratação da guaninas visualizadas no plano da base (diagrama central) e em um plano perpendicular ao plano de base (diagrama superior e inferior)

níveis de energia das moléculas de água nas suas zonas preferenciais de interação (ver TABELA 4.2.2.1).

Zona	Mínimo de energia associado à zona	Átomos do substrato implicados	NC
1	-11,9	O ₆ , H(N ₁)	24
2	-11,3	N ₆ , O ₆	13
3	-11,3	H(N ₉), N ₃ , H ₂ (N ₂)	22
4	-10,1	H(N ₁) , H, (N ₂)	94
5	- 7,0	H(C ₈)	28

TABELA 4.2.2.1 - Energias (Kcal/mole) e sítios de interação da guanina com moléculas de água. NC indica o número encontrado de configurações desta última molécula que pertence à zona de interação

Os dois outros diagramas mostram a distribuição de configurações de cada molécula de água visualizada em um plano perpendicular ao plano da base. Um destes diagramas indica as zonas detectadas na metade superior da guanina. O outro (abaixo da figura central) indica as zonas detectadas na metade inferior da base. Nos três diagramas expostos na FIGURA 4.2.2.1 é conservada a mesma numeração para as zonas de hidratação da guanina.

Esta figura mostra 5 zonas de hidratação da guanina, das quais a mais forte em interação está associada ao hidrogênio do N₁ e ao átomo O₆. A distribuição de posições da molécula

de água constituindo a zona 1 aparece estar concentrada nesta região quando ela é visualizada no plano da base. Entretanto, como mostra o diagrama superior, ela cobre uma zona relativamente grande do espaço de ambos os lados do plano da base. Desta maneira, esta molécula de água, fortemente ligada (-11,9 Kcal/mole), pode se deslocar facilmente na região delimitada pelo conjunto das configurações que constituem a zona 1. Podemos observar a mesma situação para o segundo sítio localizado entre N_7 e O_6 (-11,3 Kcal/mole). O terceiro sítio de interação está associado aos seguintes átomos: N_3 , o hidrogênio de N_9 e um dos hidrogênios do grupo amino N_2 (-11,3 Kcal/mole). É interessante observar que esta zona é dividida em duas regiões espaciais distintas, de um lado e de outro do átomo N_3 no plano da base. A quarta zona de interação se situa entre o hidrogênio de N_1 e o outro hidrogênio de N_2 (-10,2 Kcal/mole). As configurações associadas à zona 4 se distribuem radialmente em torno do hidrogênio de N_2 . Como a TABELA 4.2.2.1 indica, esta zona compreende 94 configurações. Tais como estão dispostas nesta região, elas se apresentam repartidas em quatro grupos distintos; para cada um deles os átomos de oxigênio estão na mesma posição e os de hidrogênio em posições diferentes. Esta divisão em grupos distintos é naturalmente uma consequência do fato de utilizarmos um número discreto de pontos sobre as esferas dos átomos atacantes e receptor. Assim, podemos esperar que estas configurações se distribuirão continuamente nesta região quando o número de pontos tende ao infinito.

Configurações radiais foram observadas também para outras bases estudadas neste trabalho e, de maneira geral, não cor

respondem a moléculas de água fortemente ligadas a estes subtratos.

A quinta zona de hidratação da guanina implica o hidrogênio do átomo C₈ e a energia de interação determinada para esta associação é de -7,0 Kcal/mole.

Os resultados envolvendo os outros constituintes dos ácidos nucleicos são mostrados no Apêndice D desta tese e seguem, de maneira geral, o mesmo eixo de discussão feito para a molécula da guanina. Destes estudos podemos extrair as seguintes observações relevantes:

- quando se comparam os esquemas de hidratação obtidos para as bases púricas, guanina e adenina, as particularidades de hidratação de um determinado grupo funcional não se transferem de uma base para outra. Os átomos vizinhos desempenham um papel fundamental modificando a natureza das configurações resultantes. Por exemplo, a molécula de água que se situa na zona 2 de interação da guanina tende fixar-se no plano da base. Se observarmos a molécula de água interagindo nesta mesma região da adenina (zona 2) podemos notar que, no caso desta última base, a molécula de água efetua uma rotação em torno de seu eixo de simetria de maneira que seus dois átomos de hidrogênio beneficiem de uma interação com o átomo N₇ (ver FIGURA 3 do Apêndice D)
- os esquemas de hidratação das bases pirimidicas, uracila e timina (FIGURAS 5 e 6 do Apêndice D), se apresentam de maneira similar:

- a) a associação mais estável em energia implica uma molécula de água se situando entre o hidrogênio do N_1 e o átomo O_2 .
- b) o segundo sítio de interação se localiza na região compreendida entre os átomos O_2 e O_4 ; neste sítio uma molécula de água pode se deslocar facilmente em torno do hidrogênio do N_3 .
- c) a última associação implica o átomo O_4 e a molécula de água se situa entre este átomo e os grupos ligados ao C_5 .
- d) a única diferença observada nos esquemas de hidratação destas bases envolve o grupo metila da timina. As configurações de interação de uma molécula de água com este grupo se distribuem radialmente, em torno dos átomos de hidrogênio, caracterizando a maior zona de interação determinada neste estudo. Os resultados indicaram também uma energia de interação fraca associada a este sítio (-3,0 Kcal/mole).
- por outro lado, os oxigênios do grupo fosfato atraem *fortemente* moléculas de água (-19,0 Kcal/mole) apresentando flexibilidades relativamente grandes nas zonas de interação respectivas.

4.2.3 - Sítios Preferenciais de Atração de uma Molécula de ADN-B por Moléculas de Água

Discutiremos nesta seção os resultados obtidos aplicando o procedimento desenvolvido na seção anterior a um modelo da molécula do ADN na sua conformação B.

Este polímero comporta 11 grupos fosfatos e 10 açúcares sobre cada hélice fosfodiéster e 10 pares de bases complementares, constituindo assim um passo da dupla hélice do ADN.

A macromolécula é construída a partir de seus fragmentos constituintes segundo a mesma técnica utilizada no estudo do potencial eletrostático do tARN^{Phe} (Capítulo 3). Desta maneira, a energia total de interação, entre a dupla hélice de ADN e a molécula de água, é expressa através de uma soma de interações individuais entre os multipolos de cada sub-unidade e os multipolos da molécula de água.

Os resultados, apresentados no Apêndice E, estão divididos em três seções. Cada uma delas se refere às interações de moléculas de água com uma dada região da dupla hélice. Essas regiões são: *o grande entalhe* ("major groove"), *o pequeno entalhe* ("minor groove") e *os grupos fosfatos da cadeia fosfodiéster*.

A molécula do ADN compreende dois entalhes de profundidades diferentes. As definições dos entalhes estão associadas à geometria dos pares de bases: o "grande entalhe" se situa do lado do ângulo externo formado pelas ligações glicosídicas respectivas das duas bases; "o pequeno entalhe", do lado interno deste ângulo. Assim, os átomos das bases podem ser classificados em duas categorias: aqueles que pertencem ao grande entalhe e aqueles que pertencem ao pequeno entalhe (FIGURA 4.2.3.1).

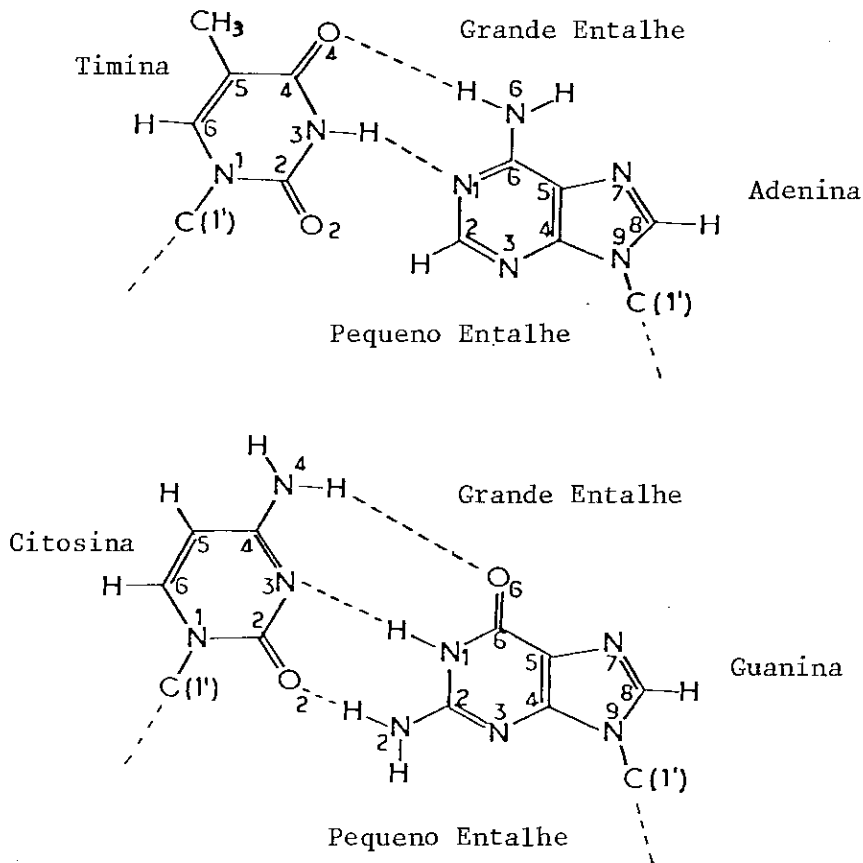


FIGURA 4.2.3.1 - Definição do pequeno e grande entalhe

Foram utilizadas duas sequências diferentes de pares de bases: uma constituída de pares adenina-timina (AT) e a outra de pares guanina - citosina (GC). Assim, para examinar a influência das diferentes sequências de bases nucleicas sobre a estruturação de moléculas de água nos seus entalhes, nosso modelo para o ADN compreende duas dupla-hélices distintas, poly(dA) . poly(dT) e poly(dG) . poly(dC).

Focalizamos somente as interações de moléculas de água com os *nucleotídeos centrais* de nosso modelo; não foram levados

em consideração os efeitos gerados pelas extremidades das dupla-hélices sobre suas hidratações. Para as bases centrais investigamos estas interações com os heteroátomos eletronegativos e com os hidrogênios dos grupos aminos *que não participam* das pontes de hidrogênio formadas pelos pares de bases dos ácidos nucléicos. Nos denotamos, por exemplo, este átomo de hidrogênio do grupo amino N_4 no caso da citosina por $H_4(C)$. O número 4 se refere ao átomo N_4 no qual o hidrogênio é ligado e a letra C designa a citosina. Para os grandes entalhes de poly(dG).poly(dC) e poly(dA).poly(dT) os átomos estudados são: $N_7(G)$, $O_6(G)$, $H_4(C)$, $N_7(A)$, $H_6(A)$ e $O_4(T)$ e para os pequenos entalhes os átomos seguintes; $N_3(A)$, $O_2(T)$, $N_3(G)$ e $H_2(G)$. Na cadeia fosfodiéster foram estudados os oxigênios do açúcar ($O_{1,}$), os aniônicos (O_1 e O_2) e os estéricos ($O_{3,}$ e $O_{5,}$) dos grupos fosfato.

Os sítios de interação das moléculas de água da primeira camada de hidratação nos entalhes das duas dupla-hélice estão ilustrados nas FIGURAS 1-6 do Apêndice E. A fim de simplificar a visualização destes resultados, mostramos somente 3 nucleotídeos de cada hélice dos polímeros; as moléculas de água estão indicadas nas suas posições de mínimo energético e o conjunto das posições dos átomos de oxigênio - das outras configurações - é representado por uma zona hachuriada.

Os resultados indicam que *diferenças importantes* são estabelecidas quando se compara a maneira pela qual se apresentam as moléculas de água no *grande* e no *pequeno* entalhe de poly(dA).poly(dT) e poly(dG).poly(dC).

Os grandes entalhes dos dois polímeros apresentam *ca*
racterísticas similares de hidratação; as moléculas de água in-
teragem com uma única base nucléica e algumas delas podem for-
mar pontes de hidrogênio com dois átomos de *uma mesma base*. No
interior do grande entalhe foram igualmente observadas molécu-
las de água apresentando grande flexibilidade, estando associa-
das a um único átomo da base.

Nos pequenos entalhes, as moléculas de água são pouco
flexíveis e formam pontes de hidrogênio entre átomos de *diferen*
tes constituintes das dupla-hélices, implicando uma base e uma
desoxiribose (poly(dA).poly(dT)) ou mesmo duas bases e uma deso-
xiribose (poly(dG).poly(dC)).

Prosseguimos este trabalho no sentido de estabelecer
uma correlação com resultados experimentais recentemente obti-
dos⁸³⁻⁸⁷. Esses se referem aos estudos de cristais de fragmen-
tos de uma dupla-hélice sintética do ADN-B contendo a sequência
de bases CGCGAATTCGCG. Pela difração de raio X sobre os cris-
tais estudados puderam ser localizadas com precisão moléculas de
água presentes nesta estrutura.

Uma das características de hidratação do dodecâmero
acima apresentado refere-se ao modo particular pelo qual se es-
truturam as moléculas de água nas camadas de hidratação no inte-
rior dos pequenos entalhes das diferentes regiões - constitui-
das ou de pares de bases GC ou AT - da dupla-hélice⁸⁵⁻⁸⁷.

Fizemos tal estudo comparativo, através de um modelo

simples para a primeira camada de hidratação nos pequenos entalhes de poly(dA).poly(dT) e poly(dG).poly(dC), e desta maneira obtivemos as feições de uma segunda camada de hidratação.

Nossos resultados mostraram que no pequeno entalhe de poly(dA).poly(dT), as moléculas de água da segunda camada unem as primeiras camadas de hidratação de cada filamento helicoidal da dupla-hélice ("cross-linking configuration") através de pontes de hidrogênio (FIGURAS 7 e 9, Apêndice E). No caso de poly(dG).poly(dC) não é observado esse "cross-linking"; a presença do grupo amino da guanina no interior do pequeno entalhe desempenha um papel fundamental no sentido de proibir - via impedimentos estereoquímicos - tal tipo de configuração (FIGURA 8 e 10, Apêndice E).

Os esquemas de hidratação, nos entalhes do ADN-B, de terminados teóricamente se mostraram em boa correlação com os resultados obtidos pelas experiências. Uma configuração aproximadamente tetraedral de ligações entre moléculas de água é observada nos pequenos entalhes das regiões centrais (compostas de pares AT) do dodecâmero CGCGAATTCGCG, enquanto que nas extremidades deste polímero (compostos de pares GC) uma tal ordenação é interrompida. Nos grandes entalhes desta dupla-hélice a maior parte das moléculas de água observadas se associam com uma única base nucléica.

Os primeiros trabalhos experimentais sobre a hidratação deste dodecâmero foram realizados em cristais a temperatura ambiente. Esta temperatura originou vibrações térmicas conside-

ráveis no interior do cristal tornando assim difícil a observação de moléculas de água⁸⁵. Estudos posteriores a uma temperatura de 16°K⁸⁶⁻⁸⁷ permitiram refinar as conclusões iniciais. A maior diferença observada entre os dois estudos experimentais está relacionada com a hidratação dos grupos fosfato da dupla-hélice; o primeiro indica uma fraca concentração de moléculas de água em torno desses grupos e o segundo coloca em evidência uma alta capacidade de hidratação dos mesmos.

Nossos estudos mostraram que as moléculas de água associadas aos grupos fosfato do ADN-B, são as que apresentam as mais fortes energias de interação com a dupla-hélice (ver TABELA 1 do Apêndice E). Como para o caso da hidratação do grupo fosfato isolado (Apêndice D), essas moléculas de água são caracterizadas - igualmente - por suas altas flexibilidades nesta região. Este pode ser um dos fatores preponderantes que dificultou encontrar moléculas de água em torno dos grupos fosfato, no caso de estudos experimentais em altas temperaturas.

CONCLUSÕES E PERSPECTIVAS

Os trabalhos expostos nesta tese seguiram dois eixos de pesquisa de interesse da Biofísica-Bioquímica Quântica. Um dentro do quadro da aplicação do cálculo de potenciais eletrostáticos moleculares para investigar alguns dos aspectos das propriedades reativas de biomoléculas. O outro se orienta no sentido de estudar o problema da hidratação destes sistemas.

O estudo paralelo do potencial molecular e das acessibilidades estereoquímicas associadas aos sítios reativos das bases nucleicas do tARN^{Phe} permitiu indicar como a estrutura terciária complexa desta macromolécula distingue bases nucleicas de um mesmo tipo em funções dessas grandezas. O "meio" macromolecular eletrostático ou estereoquímico de uma dada base se superpõe, assim, a suas propriedades intrínsecas. Comparamos nossos estudos com diversos resultados experimentais. Foi possível mostrar que, as grandezas teóricas utilizadas, são capazes de fornecer informações importantes na interpretação da reatividade de do tARN^{Phe} frente a compostos eletrofílicos tais como o que toxal⁶⁰⁻⁶¹, carbodiimido⁶³ e o AAAF⁶⁶⁻⁶⁷. Desta maneira, podemos esperar que estes cálculos teóricos poderão ser úteis na formulação de novas experiências envolvendo esta biomolécula.

Em uma etapa posterior ao estudo macromolecular realizado acima, desenvolvemos uma metodologia que permitiu considerar - simultaneamente - os aspectos energéticos e estereoquímicos da interação de pequenas moléculas polares com sistemas bio

lógicos, comportando um número relativamente grande (~ 3000) de átomos. Aplicamos este método para investigar o problema da interação de um outro ácido nucléico, o ADN-B, com moléculas de água. As distribuições eletrônicas de carga das moléculas de água foram consideradas explicitamente para os cálculos das energias de interação entre os dois sistemas. Também caracterizamos a flexibilidade da molécula de água estudando um grande número de configurações (~ 300 mil) desta molécula, quando ela se encontra em um dado sítio de interação do substrato. Utilizamos como modelo para este último duas dupla-hélices do ADN tendo respectivamente as sequências homopoliméricas poly(dA).poly(dT) e poly(dG).poly(dC). Avalia-se assim como as moléculas de água se estruturam nas diferentes regiões deste ácido nucléico em função das diferentes sequências de pares de bases implicadas. Os cálculos efetuados indicaram as seguintes características:

- 1) Os oxigênios aniônicos dos grupos fosfato da dupla-hélice constituem os sítios que apresentam as mais fortes afinidades para com as moléculas de água. Nesta região as moléculas de água apresentam também flexibilidade considerável.
- 2) As moléculas de água observadas no pequeno entalhe se associam com átomos pertencendo a diferentes constituintes da dupla-hélice. Mostramos igualmente que notáveis diferenças se estabelecem no que se refere à disposição de moléculas de água no pequeno entalhe de um ou de outro dos polímeros estudados. No caso de poly(dA).poly(dT), as moléculas de água se estruturam numa configuração geométrica ordenada de ligações, esta situação não sendo observada no caso da sequência poly(dG).poly(dC).

3) As moléculas de água no grande entalhe das duas sequências são dispostas de maneira similar. Nos dois casos elas se ligam com uma única base nucléica e aquelas que são associadas a um único átomo da base apresentam grande flexibilidade.

Estes resultados se mostraram em correlação satisfatória com resultados obtidos através de experiências recentes sobre a localização de moléculas de água em dupla-hélices do ADN-B⁸⁷.

Naturalmente as técnicas empregadas no desenvolver deste trabalho podem ser extendidas para estudar outros sistemas biomoleculares.

Podemos ainda dedicar esforços no sentido de refina-las para levar em consideração modelos possivelmente mais realistas para os sistemas biomacromoleculares e o meio que o circunda; por exemplo, incorporando nestes estudos as mudanças de estrutura que podem sofrer moléculas tais como os ácidos nucléicos e proteínas sob a influência dos diversos fatores externos (meio catiônico, efeito do solvente, reativos, etc). Seria igualmente importante aprimorar a metodologia apresentada no Capítulo IV com o objetivo de tratar de maneira simples a interação simultânea de um grande número de moléculas de água com macromoléculas biológicas.

No que diz respeito às propriedades reativas destes sistemas um passo posterior consistiria no estudo detalhado de uma certa reação levando em conta, não somente a contribuição e

letrostática, como também todos os termos que contribuem para a energia total de interação intermolecular (polarização, transferência de carga, repulsão e dispersão)..

APÊNDICES*

* Os apêndices D e E constituem trabalhos que foram extraídos de minha tese de "Doctorat de 3^{ème} Cycle", defendida no "Institut de Biologie Physico-Chimique", laboratório associado ao "Centre National de Recherches Scientifiques" (Paris - França), em 20 de dezembro de 1983.

Título da tese:

"Une méthodologie pour l'étude des sites préférentiels d'interaction de macromolécules biologiques avec des petites molécules polaires: Application à l'étude de l'interaction d'une double hélice modèle de l'ADN-B avec des molécules d'eau".

Membros da Banca:

Professor Bernard Pullman (Orientador)

Professor Raymond Daudel

Mme. Christine Bonnelle

The Electrostatic Potential of Yeast *tRNA*^{Phe}. II. The Potentials of the Phosphate Groups in Their Various Conformational States

Richard Lavery, Marçal de Oliveira, and Bernard Pullman

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique, associé au C.N.R.S., 13, rue P. et M. Curie, 75005 Paris, France

Received February 25, 1980; accepted May 5, 1980

The range of conformational states of the phosphate groups observed in the published crystal structure of *tRNA*^{Phe} is used as the basis for theoretical studies on the effect of conformation on the electrostatic potentials of these moieties. Deductions concerning the influence of these effects on the potential of a complete *tRNA*^{Phe} macromolecule are presented.

INTRODUCTION

The importance and the variety of the roles of *tRNAs* in cellular function have prompted our laboratory to undertake a theoretical study of this family of nucleic acids. As a first stage of this work we are evaluating the electrostatic molecular potential of the macromolecule. The concept and technique of calculating electrostatic molecular potentials¹⁻³ have become an important tool for the determination of fundamental features of the macromolecular structure and of the effect of this structure on the biochemical reactivity of the nucleic acids and their constituents.⁴⁻¹⁰

In the first part of this series on *tRNAs*¹¹ we have presented potentials produced by the 76 phosphates of the *tRNA*^{Phe}, which, because of their anionic nature, may be expected to contribute the dominant component to the potential of the complete macromolecule. In this study we have taken into account the variability of the conformational states of the phosphate groups in the backbone of the macromolecule, as observed in the x-ray structure, but the effects of these various conformations on the potentials of the individual phosphates were not considered. We now present a discussion of these effects.

In a forthcoming article of this series¹² the phosphate potentials presently discussed will be combined with the corresponding potentials of the ribose sugars and the nucleic acid bases to present the electrostatic potential of a complete *tRNA*^{Phe} macromolecule.

METHOD

The strategy for calculating the electrostatic potential of a molecule begins with the production of a molecular wavefunction by an *ab initio* self-consistent field (SCF) calculation. The electrostatic molecular potential associated with the corresponding electron distribution $\rho(i)$ is then given by the classical formula

$$V(P) = \sum_{\alpha} \frac{Z_{\alpha}}{r_{\alpha P}} - \int \frac{\rho(i)}{r_{pi}} d\tau_i$$

where Z_{α} is the nuclear charge of nucleus α .

It was, of course, evident from the beginning of this work that in treating macromolecules it is impossible to calculate a molecular wavefunction for the complete system. We have therefore investigated the possibility of constructing the potential of such a system as a superposition of the potentials of its individual constituent units. For nucleic acids, the most convenient division is into phosphates, sugars, and bases. More precisely, the macromolecule is divided between consecutive phosphates and sugars at the C_{3'}—O_{3'} and C_{5'}—O_{5'} bonds, and between the sugars and bases at the glycosidic bond. H termini are subsequently added to the phosphate O_{3'} and O_{5'} atoms to yield a subunit, H₂PO₄⁻ (or HPO₄⁻ in the case of the doubly ionized phosphate 1 at the 5' end of the backbone; in Part I of this series the double ionization of this phosphate was not taken into account. This has now been corrected. The effect on the summed phosphate potentials is limited to a

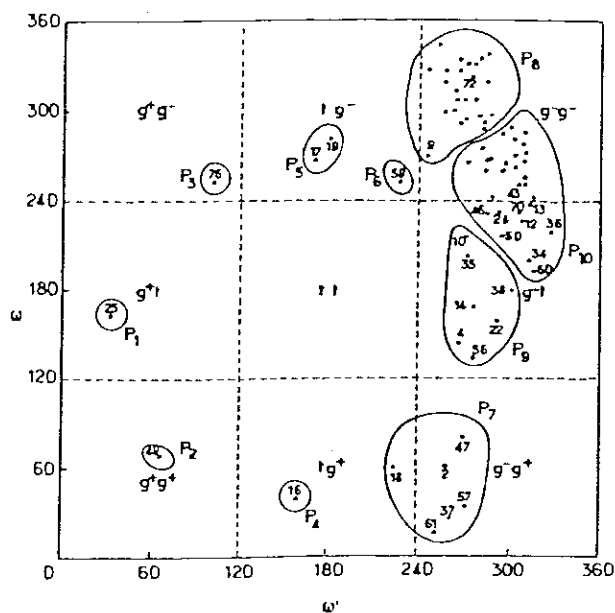


Figure 1. Conformational states of the phosphates in $tRNA^{Phe}$.

small region around phosphate 1 where the resulting potentials are found to be approximately 100 kcal/mole more negative); to the sugar C_3 , C_5 , and C_1' atoms to yield a subunit, 3,5 dideoxyribose; and to the bases at the site of the glycosidic bond to yield the free nucleic acid bases. The effect of replacing the exact bonds by the superposed CH, NH, and OH bonds in the corresponding wavefunctions has been examined and the appropriate corrections made when necessary. Wavefunctions were calculated for these subunits using a contracted Gaussian atomic orbital basis¹³ and all geometries were based on the x-ray crystallographic structure of the orthorhombic crystal of $tRNA^{Phe}$ published by Sussman et al.¹⁴

Simultaneously we have also shown that it was possible to simplify significantly the computations of the potentials associated with the macromolecules by utilizing an overlap multipole expansion (OMTP) of the electron density of the constituent units.¹⁵ The main characteristic of the procedure is the replacement of the exact electronic density distribution by a polycentric multipole expansion made of the superposition of individual multipole expansions of every overlap distribution, each overlap multipole expansion being centered either on the appropriate atom or in the middle of the distance separating the atoms carrying the atomic

orbitals involved, yielding an $n(n + 1)/2$ polycentric representation of the electron distribution for a molecule with n atoms. This expansion was shown to reproduce the exact values of the potential of the constituent units satisfactorily until a distance of 2 Å from any of their atoms, and the procedure moreover has the advantage of being computationally inexpensive.¹⁵

RESULTS AND DISCUSSION

The conformational states of the phosphates of $tRNA^{Phe}$ are shown in Figure 1 by a plot of the two phosphodiester torsional angles $C_3'-O_3'-P-O_5'$ and $O_3'-P-O_5'-C_5'$, ω' and ω , respectively, in the terminology of Sundaralingam.¹⁶ Also shown in this figure are the nine conformational regions corresponding to Sundaralingam's notation of t (180°), g^+ (60°), and g^- (300°) for these two bond rotations.

It can readily be seen that the phosphates may be classified into a limited number of groups with similar conformational states (which, moreover, generally fall into one of Sundaralingam's conformational regions).

Because small variations in ω' or ω are unlikely to have other than small effects on the resulting molecular electrostatic potentials, we decided to study only one phosphate, roughly centrally positioned in the ω' , ω plot for each group of phosphates of similar conformation.

Ten such phosphates have been chosen, namely 2, 14, 16, 19, 25, 43, 49, 59, 72, and 76, with the geometries of these residues in the crystal structure of $tRNA^{Phe}$ (ref. 14) being taken as representative for all the phosphates in each of the ten resulting conformational groups. This grouping is shown in Figure 1, where the groups are denoted P_1 – P_{10} in order of increasing ω' values for the representative phosphates. Phosphate 1 must clearly be treated separately as it is doubly ionized and thus distinct from the remaining 75, singly ionized phosphates. This same grouping of the phosphates was used in Part I¹¹ where the multipole expansions of the representative phosphates were used to calculate the contribution of all the phosphates in the corresponding conformational group to the summed phosphate electrostatic potential.

Figure 1 shows the variety of conformational states of the phosphate groups in the highly con-

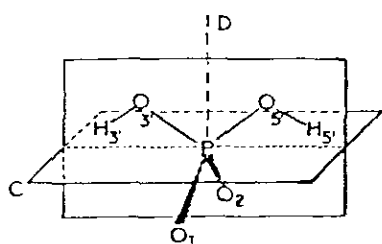


Figure 2. Representation of the phosphate plane C and line D.

volved *tRNA*^{Phe} backbone. Although the majority of the phosphates have the common g^-g^- conformation, there are phosphates in all regions of the ω' , ω space, with the exception of the central *tt* region. The instability of this latter conformational region, as theoretically predicted,¹⁷ may thus be deduced.

In order to compare the molecular electrostatic potentials of these phosphate groups we first study

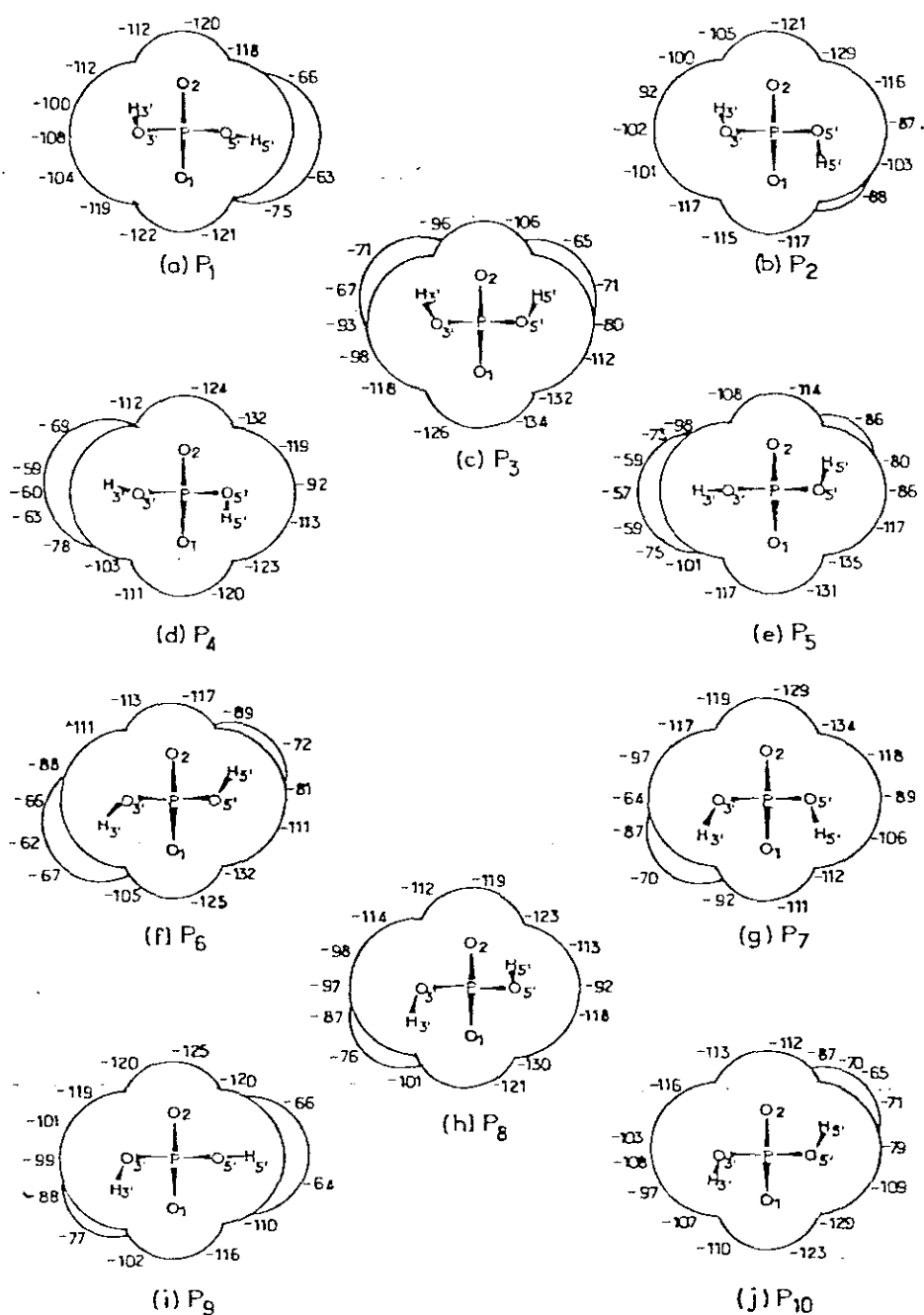


Figure 3. (a)-(j) Potentials along the 2-Å surface in plane C for the phosphate groups P_1 - P_{10} .

Table I. The properties of the molecular electrostatic potentials and the relative energies of the phosphate conformational groups of *tRNA*^{Phe}.

Phosphate group	Conformation	Minima in O ₁ PO ₂ plane (kcal/mole)	Minima in plane C (kcal/mole)	Position of plane C minima	Relative energy of representative phosphate ^a (kcal/mole)
P ₁	<i>g</i> ⁺ <i>t</i>	-155.8	-122	O ₁ —O ₃ '	+5.2
P ₂	<i>g</i> ⁺ <i>g</i> ⁺	-155.4	-129	O ₂ —O ₅ '	+1.7
P ₃	<i>g</i> ⁺ <i>g</i> ⁻	-153.2	-134	O ₁	+1.7
P ₄	<i>tg</i> ⁺	-152.8	-132	O ₂ —O ₅ '	+2.3
P ₅	<i>tg</i> ⁻	-155.9	-135	O ₁ —O ₅ '	+11.4
P ₆	<i>tg</i> ⁻	-153.7	-132	O ₁ —O ₅ '	+1.7
P ₇	<i>g</i> ⁻ <i>g</i> ⁺ and <i>tg</i> ⁺	-155.4	-134	O ₂	+2.8
P ₈	<i>g</i> ⁻ <i>g</i> ⁻	-152.7	-130	O ₁ —O ₅ '	0
P ₉	<i>g</i> ⁻ <i>t</i>	-153.3	-125	O ₂	+1.8
P ₁₀	<i>g</i> ⁻ <i>t</i> and <i>g</i> ⁻ <i>g</i> ⁻	-153.9	-129	O ₁ —O ₅ '	+1.0

^a With respect to the most stable conformer, the representative phosphate of group P₆.

the global minima (taking account of the 2-Å limit of approach for the multiple expansions) which, for all ten groups, are found to lie in a symmetric bridged position between the two anionic oxygens O₁ and O₂ (see Fig. 2). These minima are all approximately in the O₁PO₂ plane with displacements of no more than 0.2 Å toward O₃' or O₅' depending on the phosphate group considered. The values of these minima are given in column 3 of Table I: they may be seen to be stable to within ±2 kcal/mole from a mean value of -154 kcal/mole. Consequently, variation of the ω and ω' torsional angles have a negligible effect on this principal minimum.

Investigations of the potential in other regions around the phosphate moiety reveal, however, more marked effects of conformational change. We present two such regions. The first region considered is the plane C, passing through the phosphorus atom and from which the four phosphate oxygen atoms are roughly equidistant (Fig. 2). For each phosphate conformational group, potentials along the 2-Å contact surface of the phosphate were calculated for this plane with the results shown in Figures 3(a)–3(j). It may be noted that the values around the contact surfaces and the position of the corresponding minimal potentials (Table I) vary significantly with conformational changes, although in a manner qualitatively understandable on the basis of the positions of H₃ and H₅' shown in Figure 3. The minimal potentials along the 2-Å contact surface, given in Table I, show a maximal variation of 13 kcal/mole from -122 kcal/mole for P₁ to -135 kcal/mole for P₅,

their positions being localized near the anionic oxygen O₁ or O₂ or between these oxygens and O₅'. Only one phosphate conformational group from *tRNA*^{Phe}, P₁, yields a minimum displaced toward O₃'.

The second region of potential considered is described by a line D (Fig. 2) passing from the phosphorus atom toward O₃' and O₅' along the bisector of the O₃'PO₅' angle. This direction is interesting because although it points away from the strong anionic oxygen bridge potentials, it is oriented toward the bases neighboring the phosphate in a polynucleotide backbone. Potentials along this line for each of the phosphate groups are given in Figure 4.

These potentials are seen to be notably affected by conformational changes, with a maximal variation of 28 kcal/mole between groups P₆ and P₂ at a distance of 4.5 Å from the phosphorus atom. At 6 Å a maximal variation of 10 kcal/mole remains. Four groups show notably deeper potentials at all distances, P₃, P₅, P₆, and P₉, which refer to conformations of type *g*⁺*g*⁻, *tg*⁻, and *g*⁻*t*.

For the doubly ionized phosphate 1, we note that it is associated, as would be expected, with much stronger electrostatic potentials between its three pairs of anionic oxygens. Three local minima are located: O₃O₁, -263 kcal/mole, O₃O₂, -263 kcal/mole, and O₁O₂, -254 kcal/mole. The global minimum is, however, a point equidistant from all three oxygens on the 2-Å envelope, where a potential of -274 kcal/mole is calculated. Due to its different ionization state this phosphate was not included in the preceding discussion of the po-

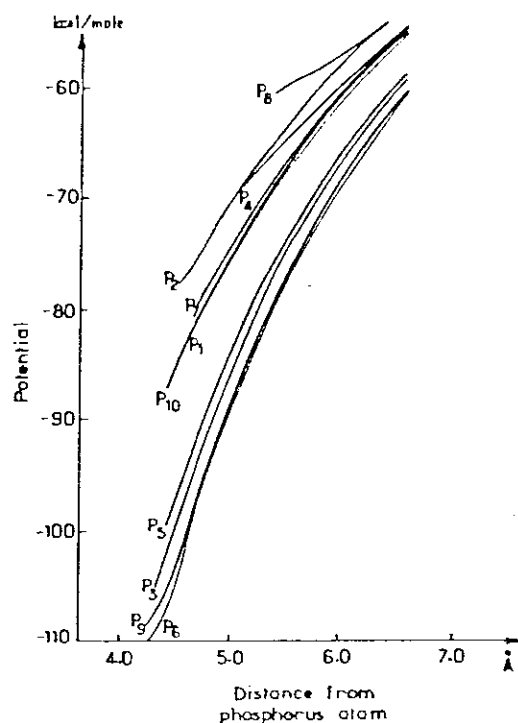


Figure 4. Potentials along a line D bisecting O_3PO_5 for the phosphate groups P_1 – P_{10} .

tentials in plane C or along line D, with the remaining phosphates, where we discuss only *conformational* effects on the phosphate potential.

In comparing the phosphate conformational groups it is interesting to mention the variation of the energy between the ten representative phosphates, revealed by our *ab initio* SCF calculations. In the final column of Table I these energies are shown with respect to the most stable conformer (the representative phosphate of group P_8 with a conformation g^-g^- close to that found in B-DNA) which is taken as energy zero. (It should be stressed that these conformational energies refer to the geometries of the representative phosphates in the *tRNA*^{Phe} molecule and are in no way the result of a conformation energy minimization procedure). It is interesting to note that the most stable conformations P_8 and P_{10} correspond to the most populated groups (see Fig. 2). P_1 (phosphate 25) and P_5 (phosphates 17 and 19) are particularly disfavored energetically and their presence in *tRNA*^{Phe} is presumably explained by important local tertiary interactions which necessitate these conformations, conceivably the D arm (C25)-variable loop (G45) interaction for group P_1 and the D loop (G18, G19)-T Ψ C loop (Ψ 55, C56) interaction for P_5 .

CONCLUSIONS

The results of the present study show that variations of the phosphodiester torsional angles have a negligible effect on the principal deepest potential, found in a bridge position between the anionic oxygens, of an isolated phosphate group. In other regions around the group the effects of conformational changes are more marked, notably along the bisector of the O_3PO_5 angle, which in a polynucleotide backbone is roughly directed toward the neighboring nucleic acid bases. The conformational states showing deeper negative potentials along this line are g^+g^- , tg^- , and g^-t .

Although such changes in potential with conformational state can amount to almost 30 kcal/mole in certain regions it is to be noted that, in the context of the potentials of a complete *tRNA*^{Phe} macromolecule, the topological positioning of the phosphate groups plays a much more important role, as will be shown in a subsequent part of this series.

We may add that we have completed a similar study of the sugar subunits of *tRNA*^{Phe} developing a grouping, as for the phosphates, based on conformational state and studying changes in electrostatic potential with their conformation. However, since the sugar subunits in our study are not complete ribose sugars (see Method section) and consequently not of general interest as isolated species, we do not present these results. We may state simply that the potentials associated with these units are several times weaker than those of the phosphates and that changes of the potential with their conformational states do not amount to more than 2–3 kcal/mole. As such, these effects are unlikely to be an important factor in determining the potential of a complete *tRNA*^{Phe}, within and around which very strong potentials exist.

This work was carried out with the support of the National Foundation for Cancer Research to which the authors wish to express their thanks. They also thank Melle Sylvie Corbin for her help in some of the computations. M.O.N. wishes to acknowledge the support of a doctoral fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

References

1. E. Scrocco and J. Tomasi, *Top. Curr. Chem.*, **42**, 95 (1973).
2. E. Scrocco and J. Tomasi, *Adv. Quantum Chem.*, **11**, 115 (1978).
3. P. Politzer and D. G. Truhlar, Eds., *Molecular Elec-*

- trostatic Potentials in Chemistry and Biochemistry*, Plenum, New York, in press.
4. R. Bonaccorsi, A. Pullman, E. Scrocco, and J. Tomasi, *Theor. Chim. Acta*, **24**, 51 (1972).
 5. R. Bonaccorsi, E. Scrocco, J. Tomasi, and A. Pullman, *Theor. Chim. Acta*, **36**, 339 (1975).
 6. R. Bonaccorsi, E. Scrocco, J. Tomasi, and A. Pullman, *Theor. Chim. Acta*, **36**, 339 (1975).
 7. A. Pullman, K. Zakrzewska, and D. Perahia, *Int. J. Quantum Chem.*, **16**, 395 (1979).
 8. D. Perahia and A. Pullman, *Theor. Chim. Acta*, **50**, 351 (1979).
 9. D. Perahia, A. Pullman, and B. Pullman, *Theor. Chim. Acta*, **51**, 349 (1979).
 10. B. Pullman, D. Perahia, and D. Cauchy, *Nucleic Acid Res.*, **6**, 3821 (1979).
 11. R. Lavery, A. Pullman, and B. Pullman, *Nucleic Acid Res.*, **8**, 1061 (1980).
 12. R. Lavery, A. Pullman, and B. Pullman, *Theor. Chim. Acta*, in press.
 13. B. Pullman, N. Gresh, H. Berthod, and A. Pullman, *Theor. Chim. Acta*, **44**, 151 (1977).
 14. J. L. Sussman, J. B. Holbrook, R. W. Warrant, G. M. Church, and S-H. Kin, *J. Mol. Biol.*, **123**, 607 (1978). In this work we have used a further refined coordinate system obtained as a personal communication from Dr. S-H. Kim.
 15. A. Goldblum, D. Perahia, and A. Pullman, *Int. J. Quantum Chem.*, **15**, 121 (1979).
 16. M. Sundaralingam, in "Conformations in Biological Molecules and Polymers," *Proc. 5th Jerusalem Symp.*, E. D. Bergmann and B. Pullman, Eds., Academic, New York, 1973.
 17. D. Perahia, B. Pullman, and A. Saran, *Biochem. Biophys. Acta*, **340**, 299 (1974).

The electrostatic molecular potential of tRNA^{Phe}. IV. The potentials and steric accessibilities of sites associated with the bases

Richard Lavery, Alberte Pullman, Bernard Pullman and Marçal de Oliveira

Institut de Biologie Physico-Chimique, Laboratoire de Chimie Théorique, associé au C.N.R.S., 13, rue Pierre et Marie Curie, 75005 Paris, France

Received 13 August 1980

ABSTRACT

The sites of the 76 nucleic acid bases of tRNA^{Phe} potentially reactive towards electrophiles are studied by calculations on the associated molecular electrostatic potentials and the static steric accessibilities. Each of these sites is treated in its environment within the macromolecule. The influence of various schemes of screening by counteractions of the backbone phosphates on the electrostatic potentials is investigated. The possible significance of the potentials and accessibilities in connection with observed chemical reactivities is discussed.

INTRODUCTION

Our laboratory is presently engaged in a theoretical exploration of certain aspects of the fine structure of the transfer ribonucleic acid, tRNA^{Phe}. In the previous publications of this series we have presented the electrostatic potential around the macromolecule produced by the assembly of its 76 anionic phosphate groups (1), the potentials of its individual phosphates as a function of their conformations (2) and the potentials and accessibilities associated with each of these phosphates in their macromolecular environment (3). In this publication we extend our studies to the evaluation of the electrostatic potentials and the steric accessibilities of potentially reactive sites associated with the nucleic acid bases of tRNA^{Phe}. These potentials and accessibilities are calculated taking into account the contribution of all the subunits, phosphates, sugars and bases of the macromolecule and, in addition, the effects of counteraction screening are also considered.

METHOD

The present calculations are based on the geometry of yeast tRNA^{Phe} in its orthorhombic crystal form given by Sussman *et al.* (4). The technique of evaluating the molecular electrostatic potential of a macromolecule has al-

ready been described (e.g. see the first article of this series (1) or our papers on similar calculation on B-DNA (5) and Z-DNA (6)). The potential sites considered in this paper are the negative potential minima which occur around the nucleic acid bases and are generally associated with the atoms of these moieties susceptible to electrophilic attack. These sites are of two types : 1) those which occur in the plane of the bases, and are associated essentially with the pyrimidine-like ring nitrogens or the carbonyl oxygens, and 2) those which occur out of the plane of the bases, and are associated primarily with the pyrrole-type ring nitrogens, C8 or the amino groups of the purines and C5 of the pyrimidines (6). For the sites which occur out of the plane of the bases there are always two related potentials, one on each side of the base. In the tables of results which follow we present only the more negative of these two values, for brevity.

The technique of calculating static steric accessibility has also been described previously in its application to the phosphates of tRNA^{Phe} (3). (See also ref. 7 and B for related calculations). As the attacking species we consider in the present calculations a sphere of 1.2 Å radius (the Van der Waals radius of the hydrogen atom) which has been shown in our previous studies (3) to reproduce well the accessibility of a macromolecule toward a water molecule, considered explicitly, via one of its hydrogen atoms. Although the results obtained for this particular simple species are, of course, rigorously valid only for this model, they certainly give an indication on the overall accessibilities to the different sites and allow to place an upper limit on the accessibility of these sites towards larger and more complex attacking molecules.

RESULTS AND DISCUSSION

A) The potentials and accessibilities of the base sites.

The results of the computations are presented in tables 1-8. The potentials reproduced in these tables have been calculated taking into account not only the macromolecule, tRNA^{Phe}, itself, but also four magnesium cations, Mg²⁺, in the positions localised by x-ray studies on the orthorhombic crystal of tRNA^{Phe} (4). These potentials thus refer to tRNA^{Phe} in its crystal form.

It may be remarked that in these tables the entries of potential and/or accessibilities are absent for certain in-plane base sites e.g. cytosine N3 (notation C(N3)) and adenine N1. The reason for these absences is the occlusion of the corresponding sites by base-pair hydrogen bonding. This occlusion strongly (although not completely (9)) reduces the magnitude of the site potentials and

the accessibility of the associated base atoms. Nevertheless, reactions at these sites cannot be ruled out, because of now well-established transient fluctuational base unpairings (see e.g. 10).

There are, in addition, certain missing entries for the modified bases, which imply the absence of a corresponding site in these residues.

We shall consider the potentials and accessibilities successively for the different types of base in tRNA^{Phe}.

TABLE 1. Guanine site potentials (kcal/mole)

Residue	N1	N2	N3	N7	C8	O6	Order
1	-978	-949	-979	-1116	-1060	-1055	N7 > C8 > O6 > N3 > N1 > N2
3	-1068	-1028	-1044	-1189	-1109	-1170	N7 > O6 > C8 > N1 > N3 > N2
4	-1091	-1068	-1086	-1206	-1125	-1196	N7 > O6 > C8 > N1 > N3 > N2
15	-1195	-1138	-1156	-1302	-1189	-1302	N7 = O6 > N1 > C8 > N3 > N2
18	-973	-1033	-1043	-986	-931	-983	N3 > N2 > N7 > O6 > N1 > C8
19	-859	-870	-928	-908	-887	-861	N3 > N7 > C8 > N2 > O6 > N1
20	-1016	-1104	-1031	-952	-955	-1005	N2 > N3 > N1 > O6 > C8 > N7
22	-1147	-1121	-1149	-	-1126	-1126	N3 > N1 > C8 = O6 > N2
24	-1098	-1071	-1119	-1223	-1141	-1183	N7 > O6 > C8 > N3 > N1 > N2
30	-965	-930	-968	-1047	-966	-1021	N7 > O6 > N3 > C8 > N1 > N2
34	-668	-660	-713	-767	-744	-717	N7 > C8 > O6 > N3 > N1 > N2
42	-1059	-1026	-1041	-1169	-1093	-1145	N7 > O6 > C8 > N1 > N3 > N2
43	-1077	-1040	-1062	-1201	-1108	-1183	N7 > O6 > C8 > N1 > N3 > N2
45	-1130	-1146	-1181	-1225	-1129	-1229	O6 > N7 > N3 > N2 > N1 > C8
51	-1096	-1056	-1079	-1257	-1132	-1230	N7 > O6 > C8 > N1 > N3 > N2
53	-1073	-1034	-1055	-1172	-1047	-1176	O6 > N7 > N1 > N3 > C8 > N2
57	-931	-957	-987	-926	-892	-949	N3 > N2 > O6 > N1 > N7 > C8
65	-1160	-1131	-1144	-1248	-1161	-1241	N7 > O6 > C8 > N1 > N3 > N2
71	-1010	-991	-987	-1089	-983	-1085	N7 > O6 > N1 > N2 > N3 > C8
10	-1078	-1069	-1067	-1182	-1133	-1105	N7 > C8 > O6 > N1 > N3 > N2
26	-1093	-1112	-1100	-1187	-1075	-1186	N7 > O6 > N2 > N3 > N1 > C8
46	-1133	-1166	-1237	-	-1072	-1152	N3 > N2 > O6 > N1 > C8

1) Guanine

The potentials (in kcal/mole) at the different sites associated with that base are shown in table I. In the first column is given the number of the base residue, ringed numbers denoting bases not involved in interbase hydrogen bonding ; the next six columns indicate the values of the potentials, each column being headed with the name of the atom of the base with which a particular site is associated ; in the final column the order of the sites, with respect to decreasingly negative potential, is given for each residue.

Several general points may be noted immediately from Table I. Firstly, the potentials are strongly negative, of the order of -800 to -1300 kcal/mole. These intensities are principally due to the strong, combined effect of potentials of the anionic phosphate groups, which superpose on the potentials inherent to the bases. Secondly, the values of potential differ appreciably for a given site from one guanine to another. This is due to the irregular and complex folded structure of tRNA^{Phe}. Thirdly, the ordering of the potentials of the sites varies from one guanine to another. This situation, again due to the folded structure of tRNA^{Phe}, is significantly different from that in a nucleic acid of regular structure, such as B-DNA, where each base in a given segment of the helix has very similar site potentials and site orderings, apart from relatively weak perturbations due to the neighbouring base sequence. tRNA^{Phe} thus presents a much more complicated picture of base site potentials, which depend significantly not only on the nature of the base concerned, but also on its location in the nucleic acid.

Considering the guanine potentials in more details we may note that the most negative potential for each base is commonly at N7, although N3 (residues 18, 19, 22, 46 and 57) and O6 (residues 45 and 53) may also occupy this position. It is interesting to note that N2, the guanine amino group site, which is generally associated with a relatively weak negative potential and is often the last site in the ordering also occurs as first, most negative site for one guanine : residue 20. The ordering for the remaining sites is very variable between the different guanines.

The average potentials associated with each base differ considerably and, from table 1, we may deduce that e.g. guanines 19 and 34 have weak average potentials (<-1000 kcal/mole) while guanines 15, 45, 51 and 65 have strong average negative potentials (>-1100 kcal/mole). These results can be correlated with the structure of tRNA^{Phe} and more specially with the special concentration of the phosphates around a given base.

The residues which occur below the horizontal line at the bottom of table

1 are the modified guanines : 10 (N2-methylguanine), 26 (N2-dimethylguanine) and 46 (7-methylguanine). The average potentials of these bases and the site ordering are similar to the normal guanines. It is interesting to note that the positive charge carried by 7-methylguanine 46 does not distinguish this residue from its close neighbour, guanine 45, in terms of average potential. The dominating influence of the summed phosphate potentials is thus exemplified.

TABLE 2. Guanine atom accessibilities (\AA^2)

Residue	N1	N2	N3	N7	C8	O6	Order
1	0.0	0.09	1.21	5.33	2.11	2.34	N7 > O6 > C8 > N3
3	0.0	0.14	0.09	4.36	0.79	2.02	N7 > O6 > C8 > N2 > N3
4	0.0	0.09	0.60	2.60	0.53	0.97	N7 > O6 > N3 > C8 > N2
15	0.0	0.0	0.60	3.38	1.0	3.31	N7 > O6 > C8 > N3
18	0.0	0.0	0.37	4.13	0.53	3.55	N7 > O6 > C8 > N3
19	0.93	1.76	1.95	5.47	1.0	7.35	O6 > N7 > N3 > N2 > C8 > N1
20	1.02	1.81	1.39	5.75	2.27	5.17	N7 > O6 > C8 > N2 > N3 > N1
22	0.0	0.0	0.0	-	0.0	0.0	-
24	0.0	0.0	0.97	1.71	0.0	0.28	N7 > N3 > O6
30	0.0	0.32	0.0	3.11	0.11	1.90	N7 > O6 > N2 > C8
34	0.88	1.90	1.21	5.05	2.06	7.31	O6 > N7 > C8 > N2 > N3 > N1
42	0.0	0.0	0.56	4.03	0.74	2.71	N7 > O6 > C8 > N3
43	0.0	0.05	0.19	4.22	0.74	2.42	N7 > O6 > C8 > N3 > N2
45	0.0	0.0	0.0	3.75	1.05	6.30	O6 > N7 > C8
51	0.0	0.09	0.42	3.29	0.58	1.74	N7 > O6 > C8 > N3 > N2
53	0.0	0.0	0.0	3.57	0.79	2.46	N7 > O6 > C8
57	0.0	0.09	0.42	0.0	0.0	3.55	O6 > N3 > N2
65	0.0	0.05	0.19	4.17	0.69	2.26	N7 > O6 > C8 > N3 > N2
71	0.0	0.0	0.28	2.04	0.21	1.09	N7 > O6 > N3 > C8
10	0.0	0.0	0.74	0.0	0.79	0.0	C8 > N3
26	0.0	0.09	0.37	0.51	0.0	2.91	O6 > N7 > N3 > N2
46	0.0	0.0	0.0	-	0.95	0.28	C8 > O6

We now turn to the calculated accessibilities of the guanine atoms corresponding to sites for which the potentials have been studied. The results are shown in table 2, the presentation being similar to that for the potential sites. Note that zero accessibilities have not been included in the ordering in the last column of this table. It may be seen that the accessible areas of the guanines vary considerably passing from one residue to the next, both in terms of their average values and of their ordering for each base. There are, however, certain similarities between the bases, notably the inaccessibility of N1 in all but three cases (19, 20 and 34) and the high accessibilities of N7 and O6. C8 is also relatively accessible for most guanines, while N3 is somewhat less accessible. N2 has low or zero accessibility in all but three residues : 19, 20 and 34. It might be expected that the unpaired bases (ringed in table 2) would show particularly high accessibilities and this is so for residues 20 and 34, but not for 57 which has rather low accessibilities, apparently because of its internal position in the T Ψ C loop and its stacking interactions with bases 56 and 58. In contrast, guanine 19, which is paired with cytosine 56, has high accessibilities being exposed on the outer surface of the macromolecule between the D and T Ψ C loops, positioned similarly to the unpaired base 20. An exceptional case of steric hindrance is seen for guanine 22 for which no site was found to be accessible. This base is situated in the core of the tRNA^{Phe} molecule, is involved in the base triplet 22-13-46 and stacked between the pair 15-48 and the triplet 23-12-9.

Finally, comparing for each guanine the orderings of the potentials in table 1 and of the accessibilities in table 2, we remark a certain correlation between these two properties. In most residues N7 and O6 are sites associated with the deepest potentials and also with large accessibilities while N2 generally have weak potentials and small accessibilities. The remaining sites are intermediately placed in both tables with the exception of the particularly low accessibility of N1.

2) Adenine

The potentials are presented in table 3. The final entry at the bottom of this table is the modified base 1-methyladenine 58.

As for the guanines, the average potentials and the site orderings differ considerably for the various adenines. Thus, adenines 35, 36 and 76 have particularly weak average potentials, of the order of -600 to -800 kcal/mole, while those for residues 9, 64, 66 and 67 are strong, -1100 to -1250 kcal/mole. For the individual bases N7, N6 and, somewhat less commonly N3, have the most

negative potentials, while the potentials for N1 (this site is often occluded by hydrogen bonding) and C8 are generally weaker. For adenine 38, however, C8 becomes the most negative site. Contrasting these results with those for guanine one may note the relatively more negative potentials associated with the adenine amino groups in comparison with those of the guanine amino groups.

The accessibilities for adenines are given in table 4. Here we may note the generally large accessibilities at N7 and N3, and the small values at N6. The accessibilities at C8 are intermediate in magnitude with the exception of

TABLE 3. Adenine site potentials (kcal/mole)

Residue	N1	N3	N6	N7	C8	Order
5	-	-1145	-1176	-1223	-1165	N7 > N6 > C8 > N3
9	-1241	-1217	-1193	-	-1156	N1 > N3 > N6 > C8
14	-1138	-1106	-1147	-	-1170	N1 > C8 > N6 > N3
(21)	-1184	-1161	-1195	-1139	-1150	N6 > N1 > N3 > C8 > N7
23	-	-1132	-1167	-	-1154	N6 > C8 > N3
29	-	-1003	-1060	-1090	-987	N7 > N6 > N3 > C8
31	-921	-928	-973	-992	-922	N7 > N6 > N3 > N1 > C8
(35)	-742	-735	-744	-811	-783	N7 > C8 > N6 > N1 > N3
(36)	-757	-776	-736	-805	-725	N7 > N3 > N1 > N6 > C8
(38)	-836	-842	-831	-838	-843	C8 > N3 > N7 > N1 > N6
44	-	-1081	-1128	-1207	-1118	N7 > N6 > C8 > N3
62	-	-1064	-1169	-1205	-1119	N7 > N6 > C8 > N3
64	-	-1105	-1203	-1215	-1130	N7 > N6 > C8 > N3
66	-	-1173	-1220	-1246	-1165	N7 > N6 > N3 > C8
67	-	-1183	-1209	-1240	-1145	N7 > N6 > N3 > C8
(73)	-995	-948	-962	-994	-889	N1 > N7 > N6 > N3 > C8
(76)	-671	-679	-580	-614	-582	N3 > N1 > N7 > C8 > N6
58	-	-1118	-978	-	-1058	N3 > C8 > N6

the large value of 3.01 \AA^2 for the terminal acceptor end base, adenine 76. This exposed base is in fact highly accessible at all sites. The modified base 1-methyladenine 58 shows only a small accessibility at N3. Five adenines, residues 35, 36, 38, 73 and 76, show quite high accessible areas at N1. All these bases are unpaired. In contrast, adenine 21 which is also unpaired is inaccessible at N1, N3 and N6, apparently because of its orientation towards the core of the molecule and the consequent hindrances caused by uracil 47 and the stacking with guanine 22 and 7-methylguanine 46.

TABLE 4. Adenine atom accessibilities (\AA^2)

Residue	N1	N3	N6	N7	C8	Order
5	-	1.07	0.05	1.90	0.0	N7 > N3 > N6
9	1.11	0.70	0.70	-	0.0	N1 > N3 > N6
14	0.83	2.09	0.0	-	0.0	N3 > N1
(21)	0.0	0.0	0.0	2.22	0.42	N7 > C8
23	-	1.90	0.0	-	0.26	N3 > C8
29	-	0.88	0.37	3.11	0.74	N7 > N3 > C8 > N6
31	0.0	0.93	0.56	1.02	0.0	N7 > N3 > N6
(35)	2.09	1.11	0.0	0.0	0.0	N1 > N3
(36)	3.62	1.25	0.05	0.46	0.32	N1 > N3 > N7 > C8 > N6
(38)	2.50	0.88	0.23	0.46	0.0	N1 > N3 > N7 > N6
44	-	1.99	0.0	1.48	0.0	N3 > N7
62	-	1.44	0.09	1.99	0.05	N7 > N3 > N6 > C8
64	-	0.97	0.0	2.32	0.84	N7 > N3 > C8
66	-	0.74	0.05	2.69	0.69	N7 > N3 > C8 > N6
67	-	1.21	0.05	2.09	0.26	N7 > N3 > C8 > N6
(73)	2.46	0.93	0.14	1.44	0.21	N1 > N7 > N3 > C8 > N6
(76)	5.89	1.16	3.43	6.40	3.01	N7 > N1 > N6 > C8 > N3
58	-	0.14	0.0	-	0.0	N3

Comparing potentials and accessibilities for the adenines, N7 and N3 are seen to be well placed for most residues while the relatively deep potentials of N6 are contrasted by relatively low accessibilities.

3) Cytosine

The potentials at cytosine sites are contained in table 5. At the bottom of this table are included the modified bases 5-methylcytosines 40 and 49.

TABLE 5. Cytosine site potentials (kcal/mole)

Residue	N3	N4	O2	C5	Order
2	-	-1091	-985	-1074	N4 > C5 > O2
11	-	-1133	-1075	-1114	N4 > C5 > O2
13	-	-1205	-1131	-1188	N4 > C5 > O2
25	-	-1144	-1074	-1143	N4 > C5 > O2
27	-	-1111	-1053	-1069	N4 > C5 > O2
28	-	-1091	-994	-1049	N4 > C5 > O2
③②	-874	-921	-834	-865	N4 > N3 > C5 > O2
48	-	-1116	-1289	-1141	O2 > N4 > C5
56	-	-880	-839	-890	C5 > N4 > O2
⑥①	-1099	-1065	-1193	-1112	O2 > C5 > N3 > N4
61	-	-1179	-1049	-1139	N4 > C5 > O2
63	-	-1164	-1052	-1151	N4 > C5 > O2
70	-	-1104	-1012	-1042	N4 > C5 > O2
72	-	1017	-933	-970	N4 > C5 > O2
⑦④	-926	-891	-911	-885	N3 > O2 > N4 > C5
⑦⑤	-857	-842	-830	-837	N3 > N4 > C5 > O2
40	-	-1028	-978	-1058	C5 > N4 > O2
49	-	-1235	-1187	-1254	C5 > N4 > O2

For cytosine the site associated with N3 is occluded by hydrogen bonding for all but the four unpaired residues : 32, 60, 74 and 75. Of the three remaining sites N4 is most commonly associated with the deepest potential and O2 with the weakest. Once more the bases from differing regions of the macromolecule are associated with different average potentials, the weakest being those for cytosines 32, 56, 74 and 75 and the strongest those for cytosines 13, 48 and 49. It is interesting to note that in both the 5-methylcytosines

TABLE 6. Cytosine atom accessibilities (\AA^2)

Residue	N3	N4	O2	C5	Order
2	-	0.19	1.49	0.74	O2 > C5 > N4
11	-	0.0	1.13	0.42	O2 > C5
13	-	0.0	0.24	0.21	O2 > C5
25	-	0.0	0.52	0.0	O2
27	-	0.09	0.32	0.47	C5 > O2 > N4
28	-	0.0	0.73	0.74	C5 > O2
32	1.62	0.19	2.18	0.26	O2 > N3 > C5 > N4
48	-	0.0	0.32	0.11	O2 > C5
56	-	1.48	3.23	1.69	O2 > C5 > N4
60	1.07	0.14	4.48	0.42	O2 > N3 > C5 > N4
61	-	0.05	0.97	0.21	O2 > C5 > N4
63	-	0.05	1.05	1.11	C5 > O2 > N4
70	-	0.14	0.89	1.21	C5 > O2 > N4
72	-	0.0	0.85	0.16	O2 > C5
74	2.69	0.42	6.18	0.0	O2 > N3 > N4
75	3.62	1.71	7.95	1.69	O2 > N3 > N4 > C5
40	0.0	0.19	2.22	0.16	O2 > N4 > C5
49	0.0	0.56	1.01	0.32	O2 > N4 > C5

(40 and 49), the C5 site has the most negative potential, which for the unmodified bases is only the case for cytosine 56.

Passing to the cytosine accessibilities (table 6), the dominance of the values for O2 and C5 may be seen. N4 is only slightly accessible with the exception of cytosines 56 and 75 which are both exposed ; 56 between the D and T Ψ C loops and 75 in the acceptor end.

Sites D2 and C5 of cytosines are thus calculated to have both relatively strong negative potentials and large accessibilities for the majority of these residues.

TABLE 7. Uracil site potentials (kcal/mole)

Residue	N3	C5	O2	O4	Order
6	-1187	-1212	-1152	-1237	O4 > C5 > N3 > O2
7	-1201	-1226	-1174	-1241	O4 > C5 > N3 > O2
8	-1244	-1290	-1123	-1285	C5 > O4 > N3 > O2
12	-1109	-1113	-1067	-1183	O4 > C5 > N3 > O2
33	-853	-863	-737	-954	O4 > C5 > N3 > O2
41	-1032	-1063	-947	-1071	O4 > C5 > N3 > O2
47	-1086	-1092	-1098	-1037	O2 > C5 > N3 > O4
50	-1169	-1219	-1081	-1245	O4 > C5 > N3 > O2
52	-1086	-1112	-1015	-1206	O4 > C5 > N3 > O2
59	-1122	-1099	-1180	-1050	O2 > N3 > C5 > O4
68	-1145	-1142	-1101	-1205	O4 > N3 > C5 > O2
69	-1103	-1093	-	-1191	O4 > N3 > C5
16	-978	-	-1055	-943	O2 > N3 > O4
17	-800	-	-971	-792	O2 > N3 > O4
38	-928	^{N1} -958	-939	-918	N1 > O2 > N3 > O4
55	-1050	^{N1} -1028	-1101	-1004	O2 > N3 > N1 > O4
54	-1034	-1034	-1141	-1167	O4 > O2 > N3 = C5

4) Uracil

The potentials for uracil are contained in table 7. Included at the foot of this table are the modified bases dihydrouracil (16 and 17), pseudouracil (39 and 55) and thymine (54), which may be considered as a modified uracil in the present circumstances. We note that in the pseudouracils 39 and 55, which are linked to the sugar via C5, the potential associated with N1 replaces that for C5, which occurs in the equivalent geometrical position in the unmodified uracils.

The most negative potentials for the different uracils are found to be

TABLE 8. Uracil atom accessibilities (\AA^2)

Residue	N3	C5	O2	O4	Order
6	0.0	0.69	1.13	3.15	O4 > O2 > C5
7	0.0	0.84	2.30	1.45	O2 > O4 > C5
8	0.0	1.21	0.0	0.69	C5 > O4
12	0.0	0.16	2.18	0.44	O2 > O4 > C5
(33)	0.19	0.32	0.65	3.47	O4 > O2 > C5 > N3
41	0.0	1.32	1.41	2.06	O4 > O2 > C5
(47)	1.11	3.69	4.85	11.39	O4 > O2 > C5 > N3
50	0.0	1.27	1.41	2.42	O4 > O2 > C5
52	0.0	0.90	1.29	3.19	O4 > O2 > C5
(59)	0.0	0.0	5.13	3.96	O2 > O4
68	0.0	0.05	1.49	1.45	O2 > O4 > C5
69	0.0	1.32	-	4.85	O4 > C5
(16)	1.62	-	8.84	11.10	O4 > O2 > N3
(17)	1.67	-	3.39	11.14	O4 > O2 > N3
39	0.0	$\frac{N1}{0.14}$	3.03	2.38	O2 > O4 > N1
55	0.0	$\frac{N1}{0.56}$	7.19	0.0	O2 > N1
54	0.0	0.0	0.0	2.18	O4

those associated with O4. O2 is generally the weakest of the sites studied but in the case of five residues, 16, 17, 47, 55 and 59, this situation is reversed and it becomes the most negative site. For all these residues however, O2 is not involved in hydrogen bonding. N3 and C5 are generally associated with intermediate potentials with the exception of C5 for uracil 8, in which it is the most negative site.

The weakest average potentials for the uracil residues are those of the dihydrouracils 16 and 17 in the O loop and also of uracil 33, while the deepest average potentials are calculated for uracils 6, 7, 8, 50 and 68.

Table 8 shows the accessibilities at the uracil atoms. One may note the largest values for O2 and O4 for all residues with the exception of residues 8 and 55 where O2 and O4 are, respectively, inaccessible. For uracil 8, C5 has a large accessibility and it will be recalled that this atom is also associated with the most negative potential for this base. These phenomena are most probably explained by the unusual, reverse Hoogsteen, base pairing of uracil 8 with adenine 14 via its N3-H and O2 centers. The N3 atom is generally inaccessible, the exceptions being for the unpaired bases 16, 17, 33 and 47.

Comparing tables 7 and 8, O4 of uracil is seen to be associated most commonly with the strongest potentials and largest accessibilities, but this is also true for O2 when this atom is not involved in hydrogen bonding.

B) The effects of countercation screening

At this point, it may seem relevant to ask whether the potentials for tRNA^{Phe} in the presence of the four crystallographically located Mg^{2+} cations as presented in the preceding section, are relevant in describing tRNA^{Phe} in the presence of other cation arrangements, as could be the case for the molecule in solution. To tentatively answer this question we studied two new situations, namely, tRNA^{Phe} in the absence of any cation and tRNA^{Phe} in the presence of both the four Mg^{2+} and twenty Na^+ cations. Both these studies were carried for the test case of the guanine residues. The latter ions were bound to the bridge sites between the phosphate anionic oxygens, a position identical to that employed in studying the Na^+ cation screening of B-ONA (19). Their binding to given phosphates and their number was arrived at by an optimisation procedure which searched to locate, for each new ion, the bridge site with the most negative potential, avoiding at the same time too close approaches between the various countercations (limiting distances of $\text{Na}^+ - \text{Na}^+$ 7Å and $\text{Na}^+ - \text{Mg}^{2+}$, 14Å were imposed). The phosphates found in that way to bind Na^+ ions are residues 1, 3, 6, 15, 24, 25, 31, 44, 45, 51, 53, 57, 61, 63, 66, 68, 70, 72, 74

and 76. This screening does not pretend to represent the true countercation distribution in solution, but is rather intended to illustrate one of many possible models of a more complete screening of tRNA^{Phe} than that which exists in its orthorhombic crystal form in association with the four localized Mg^{2+} ions.

For reasons of economy of space imposed by the editorial rules we shall not reproduce the numerical results here. They may be obtained upon request. Here we simply wish to indicate that the order of potential sites for each base is not greatly changed between the three approximations investigated and that the same is true the ordering of the guanines in terms of their average potentials.

C) Correlations with experimentally observed chemical reactivities of tRNA^{Phe} .

It has been demonstrated, notably by our earlier studies of B-DNA (5, 12), that both electrostatic potentials and steric accessibility can, and often do, play a significant role in determining the chemical reactivity of sites within a macromolecule. We similarly investigated the possible role of these factors for several experimentally studied reactions of tRNA^{Phe} with electrophiles. The results of the previous sections indicate, however, by themselves the limitations of such an approach inasmuch as they show that strong negative potentials, which would imply greater affinity for a reaction with electrophiles, often occur for bases with low accessibilities, which disfavour reactions and that converse situation is also common, weak potentials being associated with bases having high accessibilities. It is also obvious a priori that these are not the only factors involved in the reactions and that as illustrated by ourselves on numerous occasions (13-15), polarisation and charge transfer are important components, besides the electrostatic one in the global interaction energy between two species. Nevertheless the previous success of this approach encourages its futur exploration.

We have studied four reactions of tRNA^{Phe} : with (i) kethoxal (16-17) (ii) I_2/HCl_3 (18), (iii) carbodiimide (19) and (iv) N-acetoxy-2-acetylaminofluorene (20-23). Each of these reactants interacts with specific sites on specific types of base within tRNA^{Phe} . We were concerned only with trying to understand to what an extent may the computed potentials and accessibilities account for the reactivity or non reactivity of certain bases in given structural regions of the macromolecule toward a given reactant. A full explanation of the site specificity of these reagents implies more complete quantum mecha-

nical studies of their respective reaction mechanisms, which is beyond the scope of this publication (see e.g. 13-15).

Because of space limitation we can present here the results for only one of the above quoted reactions. A more complete study will be presented separately.

The example selected concerns kethoxal. This compound reacts specifically with the guanine bases of tRNA^{Phe} and in the studies of Litt (16-17) only two guanines, numbers 20 and 34, are found to bind this reagent. The reaction product is a bridged adduct resulting from substitutions at guanines N1 and N2.

In figure 1 we have combined the calculated data for the accessibilities (horizontal axis) and potentials (vertical axis). Each point on this graphic therefore represents, simultaneously, the potential and the atom accessibility for a given base site. On this graphic we have, in this manner, placed dots representing the N1 and N2 sites of all the guanines of tRNA^{Phe}. Sites with zero accessibility, consequently unlikely to be reactive, are not included.

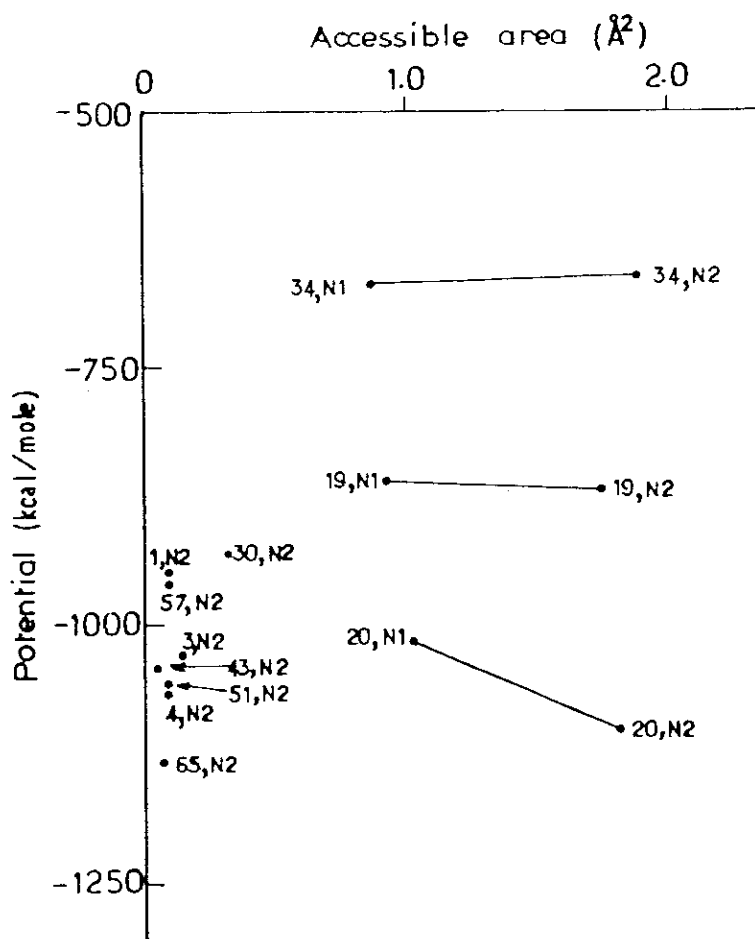


Fig. 1 . Potential and accessibility for the guanine N1 and N2 sites in tRNA^{Phe}.

When both the N1 and N2 sites of a given base appear on this diagram they have been joined by a line. It will be noted that this is the case only for guanines 19, 20 and 34. For the other bases the accessibilities of N1 or N2 or of both are zero (see table 2). The three remaining bases are seen to have high accessibilities for both sites, although N2, in each case, is somewhat more accessible than N1. The N1 and N2 sites of guanine 20 have high negative potentials, while those of guanine 34 are considerably lower, guanine 19 being intermediately placed.

The observed reactivity of guanines 20 and 34 is consequently understandable. The inactivity of guanine 19 may be due to the fact that while guanines 20 and 34 are unpaired, guanine 19 is bound to cytosine 56 with hydrogen bonds that involve its N1 and N2 atoms and these bonds would have to be disrupted before a reaction with kethoxal could take place.

CONCLUSIONS

In the present publication the electrostatic potentials and the steric accessibilities of all the sites on the nucleic acid bases of yeast tRNA^{Phe}, which are susceptible to electrophilic attack, have been calculated.

It is shown that the potentials and accessibilities of specific sites on each type of base are highly variable, depending strongly on the positioning of the base inside the macromolecule. This variability applies both to the average potential or accessibility associated with the sites of the bases and to the ordering of the sites, in terms of these values, for each individual base.

The effect of different counteraction screenings of the backbone phosphates has also been investigated. It is found that increased screening diminishes the absolute values of the base site potentials and also their spacing, but does not cause much reordering.

A clear correlation between structure and potential may be noted, bases in the acceptor end or the loops generally having much weaker potentials than those situated in the helical segments, closer to the core of the macromolecule. The accessibility of certain bases in the acceptor end or the loops is also particularly high compared to other regions, but the structural correlations with this property are less clear due to its sensitivity to the local structure surrounding each individual base atom.

Finally partial correlations between the combined potentials-accessibilities for the base sites and experimentally observed chemical reactivities of

tRNA^{Phe} have been noted. More detailed reactivity studies are required before definite conclusions on the relative roles of the potential and accessibility and their relation to other factors involved (e.g. polarization, charge transfer, etc.) can be made. Such studies are under way.

Acknowledgments - This work was supported by the contract n°14 of the Institut National de la Recherche Medicale of France with its ATP 77 79-109 in Chemical Carcinogenesis. One of us (M.O.N.) wishes to acknowledge the support of a doctoral fellowship from Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, Brazil.

References

- 1 Lavery, R., Pullman, A. and Pullman, B., (1980) Nucl. Acids Res. 8, 1061-1078.
- 2 Lavery, R., de Oliviera, M. and Pullman, B., (1980) J. Computational Chem., 1, 301-306.
- 3 Lavery, R., Pullman, A. and Pullman, B., Theoret. Chim. Acta, in press.
- 4 Sussman, J.L., Holbrook, J.B., Warrant, R.W., Church, G.M. and Kim, S-H., (1978) J. Mol. Biol. 123, 607-630. We have utilized in this work a further refined coordinate system obtained as a personal communication from Dr. S-H. Kim.
- 5 Pullman, B., Perahia, D. and Cauchy, D. (1979) Nucl. Acids. Res. 6, 3821-3830.
- 6 Pullman, A. in Mechanismes d'Alteration et de Reparation du DNA, Relations avec la Mutagenese et la Cancérogénese Chimique. Colloque du C.N.R.S. 256 Paris 1976, p. 103-113.
- 7 Thiyagarajan, P. and Ponnuswamy, P.K. (1979) Biopolymers, 18, 2233-2247.
- 8 C.J., Alden and S-H. Kim, (1979) J. Mol. Biol. 132, 411-434
- 9 Pullman, A. and Berthod, H., (1978) Theoret. Chim. Acta, 48, 269-277.
- 10 Kallenbach, N.R., Mandal, C. and Englander, S.W. in Stereodynamic of Molecular Systems (R.H. Sarma Ed.) Pergamon Press, New York, (1979) p. 271-282.
- 11 Lavery, R., Cauchy, D., Rojas, J. and Pullman, B. (1980) Int. J. Quant. Chem. Biol. Symp. 7, in press.
- 12 Lavery, R., Pullman, A. and Pullman, B., Bioch. Biophys. Acta, in press.
- 13 Pullman, A. and Armbruster, A.M., (1977) Theoret. Chim. Acta. 249-255.
- 14 Pullman, A. and Armbruster, A.M., (1979) Theoret. Chim. Acta. 5°, 359-361.
- 15 Perahia, D., Pullman, A. and Pullman, B., (1977) Theoret Chim. Acta. 43, 207-212.
- 16 Litt, M. (1969) Biochemistry, 8, 3249-3253.
- 17 Litt, M. and Greenspan, C.M. (1972) Biochemistry, 11, 1437-1442.
- 18 Batey, I.L. and Brown, D.M. (1977) Biochim. Biophys. Acta, 474
- 19 Rhodes, D. (1975) J. Mol. Biol. 94, 449-460.
- 20 Fujimura, S., Grunberger, D., Carvajal, G. and Weinstein, I.B. (1972) Biochemistry, 11, 3629-3635.
- 21 Pulkrabek, P., Grunberger, D. and Weinstein, I.B. (1974) Biochemistry, 13, 2414-2419.
- 22 Sprinzl, M., Grueter, F., Spelzhaus, A. and Gauss, D.H. (1980) Nucl. Acid. Res. B, r1-r22.
- 23 Massouh-Rizk, L., Thèse d'Etat, (1975), Université Louis Pasteur de Strasbourg, France.

A Preliminary Theoretical Study of the Acid Catalyzed Hydration of Glyoxal, Methylglyoxal, and Other Simple Aldehydes

RICHARD LAVERY, MARÇAL DE OLIVEIRA, AND BERNARD PULLMAN

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique associé au C.N.R.S., 75005 Paris, France

Abstract

On the basis of several parameters calculated quantum chemically an explanation is proposed for the relative reactivity of a range of aldehydes toward acid catalyzed hydration. In particular the biochemically interesting species of glyoxal and methylglyoxal are studied and compared to other aldehydes.

Introduction

In recent years a renewed interest has centered on the cellular role of glyoxal and methylglyoxal as a result of the proposals of Szent-Györgyi that the latter of these molecules may be involved in the control of cellular division and his interpretation of the observed cancerostatic ability of both molecules [1-5]. In this laboratory we have made several theoretical investigations of the reactions of these two species [6-8], including a recent model study of the transformation of methylglyoxal to lactic acid, mediated by the glyoxylase enzyme system [9].

As a complement to these latter studies we now present a preliminary theoretical inquiry into the hydration of glyoxal and methylglyoxal to form gem-diols, which is an important reaction to consider if the cellular concentrations of these molecules are to be understood. A particularly interesting aspect of this problem, which we attempt to clarify, is the observation that both glyoxal and methylglyoxal are readily hydrated, like formaldehyde and, in fact, to a greater extent than this latter molecule, whereas acetaldehyde and higher aldehydes form no stable hydrates [10, 11].

The hydration of an aldehyde is illustrated in Figure 1. In order to understand the comparative reactivity of the molecules mentioned above we have considered

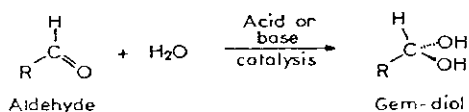


Figure 1. Aldehyde hydration reaction.

LAVERY, DE OLIVEIRA, AND PULLMAN

as -R groups -H, -CH₃, -CHO, and -COCH₃, that is, the molecules formaldehyde, acetaldehyde, glyoxal, and methylglyoxal. As indicated in the figure, the reaction may be acid or base catalyzed. In this preliminary study we limit ourselves to a consideration of the acid catalyzed mechanism.

Method

Standard bond lengths and bond angles have been used for all the molecular species studied, with the exception of certain bond length optimizations which are described shortly. When these optimizations were performed, a minimal STO-3G basis [12] was employed, as this has been shown to lead to satisfactory agreement with experimental bond lengths. For the remaining calculations we used the double zeta contracted to minimal basis of Ref. [13], which yields better total energies. All molecular conformations were optimized in this basis to reduce steric hindrance and improve intra- or intermolecular hydrogen bonding.

Calculations were carried out using the Gaussian 70 program [14].

Results and Discussion

A recent extensive study of the kinetics of formaldehyde hydration by Funderburk et al. [15] has concluded that earlier suggestions of a cyclic transition state involving several water molecules [16-18] are not supported by the kinetic evidence. The observation of an apparent third-order dependence on water concentration for hydrations carried out in aqueous dioxan or other nonpolar solvents is, it was concluded, likely to be due to the intimate involvement of two solvating water molecules stabilizing a proton transfer to the aldehyde carbonyl group, while only a single water molecule actually takes part in the transition state complex.

More particularly, in the case of the acid catalyzed hydration of formaldehyde it was concluded that the mechanism involves a concurrent protonation of the carbonyl group and complexation with a water molecule (a class *e* mechanism) [15]. This conclusion is based on the apparent instability of the protonated aldehyde and the evidence that water addition to this protonated species would exhibit little or no energy barrier [19]. Consequently no buffer catalysis by water (or, more generally, in the case of acetal formation, by an alcohol) would be observed, an inconsistency with the kinetic evidence.

To study such a three-species interaction quantum chemically is difficult, even when limited areas of the reaction energy hypersurface are of interest. Consequently, we have separated the protonation and water complexation phases of the transition state formation for computational convenience. Such a division need not influence our final interpretation of aldehyde reactivity, as this will largely be based on the structure and stability of the transition state complex rather than the kinetics of its formation.

We began our studies by attempting to calculate a transition state complex structure between a molecule of formaldehyde and a molecule of water. In order

ACID CATALYZED REACTION OF ALDEHYDES

to lead to the desired reaction the oxygen atom of water in this complex should be bound directly to the carbon atom of formaldehyde which would, as suggested by the kinetic evidence [20], be distorted to a tetrahedral form, the two hydrogens and the carbonyl oxygen atom being bent away from the approaching water molecule. Such a structure has been studied by semiempirical methods, but no details of geometry or binding energies are included in the publication [21].

In order to simplify the study of the distortion of formaldehyde which, in practice, is a continuous progression of structures as the transition state is formed, we considered only four conformations of the molecule, termed I, II, III, and IV. Conformation I was the normal planar molecule, while conformation IV has a tetrahedral carbon atom with the three remaining atoms bent back symmetrically with respect to an axis, henceforth termed the z axis, passing through the carbon atom and perpendicular to the original plane of the molecule (Fig. 2). Conformations II and III were intermediate distorted geometries equispaced between conformations I and IV. Consequently the angles formed between the formaldehyde ligands and the z axis for conformations I, II, III, and IV are, respectively, 90° , 96.5° , 103° , and 109.5° .

A water molecule was subsequently allowed to approach the formaldehyde molecule along the z axis (Fig. 2), and for each of the formaldehyde conformations described a value of r_{CO} , the distance between the formaldehyde carbon atom and the water oxygen atom, for which the interaction energy between the two molecules would be optimal, was searched for. The results of these calculations showed, however, that no attraction between the two molecules was observed for any of the formaldehyde conformations, and the complex became less stable with decreasing r_{CO} . The only exception to this result was a small attractive minimum of 0.7 kcal/mole between the water molecule and conformation I at $r_{CO} = 3 \text{ \AA}$. This is, however, possibly an artifact, a basis extension effect due to the use of a minimal atomic basis.

Faced with this lack of complexation we turned to an investigation of an initial protonation of the formaldehyde oxygen atom and, subsequently, to the investigation of complexes between this protonated species and a water molecule.

Protonation of formaldehyde in this way was favorable. An optimized geometry yielded the proton bound to oxygen with a bond length of 1.01 \AA , having an angle of 117° with the carbonyl bond axis and lying in the plane of the mol-

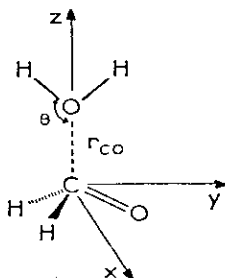


Figure 2. Geometry of the formaldehyde-water complex.

LAVERY, DE OLIVEIRA, AND PULLMAN

ecule. The reoptimized carbonyl bond length increased by only 0.05 to 1.27 Å, implying that this bond still had considerable double bond character. This deduction is supported by the observation that rotating the protonated bond to be perpendicular to the molecular plane results in a loss of 36 kcal/mole in total energy. At the optimized geometry the protonation energy was -143.0 kcal/mole, virtually identical to the protonation energy of water in this atomic basis (-142.3 kcal/mole).

We proceeded to investigate complexes between a water molecule and the protonated formaldehyde, using the conformations I-IV previously described for formaldehyde distortion, and approaching the water molecule as in Figure 2. The proton linked to the carbonyl was maintained perpendicular to the *z* axis as, even for the distorted conformations II-IV, a large energy loss resulted from its rotation.

The optimum energy for the complex was found with conformation III when $r_{CO} = 1.60$ Å, at which geometry there was a binding interaction between the two molecules of 33.4 kcal/mole. If one considers that this energy conceals the unfavorable distortion energy of protonated formaldehyde in order to pass from conformation II to conformation III (14.3 kcal/mole), it is seen that a rather strong complex has been formed. Further optimization of the complex geometry was attempted by rotating the water molecule about the *z* axis and changing its orientation by the angle θ (Fig. 2), but no significant improvement in complexation energy was found.

Preliminary investigations now indicated that deprotonation of the water molecule to yield the final acetal (after necessary geometrical changes) was a strongly endothermic step and would clearly involve the assistance of another molecule. Because such assistance is difficult to model quantum chemically, especially if several solvent molecules are intimately involved, we did not attempt further studies on this step but passed immediately to the final acetal.

When the conformation of this species was optimized (both hydroxyl protons were found to lie in the O-C-O plane, one of them having a weak hydrogen bond interaction with the neighboring hydroxyl oxygen and C-OH bond lengths equal to 1.43 Å), the overall hydration reaction was observed to be exothermic by 33.5 kcal/mole. This, however, implies that to deprotonate the water complex to arrive at the final reaction product requires 142.9 kcal/mole. This is a large expense of energy but is feasible with the assistance of one or more water molecules (protonation energy = -142.3 kcal/mole) or other more strongly proton-accepting anions which may be present in the reaction medium.

As a result of these studies it would seem that the start of protonation of the aldehyde carbonyl group is more likely to trigger water complexation than the reverse process, but because of the model division of these steps which we adopted we cannot give any further precision as to the kinetics of the transition state formation.

Figure 3 represents the steps considered in the computations. The associated energetic parameters are illustrated schematically in Figure 4 and listed in detail in Table I.

ACID CATALYZED REACTION OF ALDEHYDES

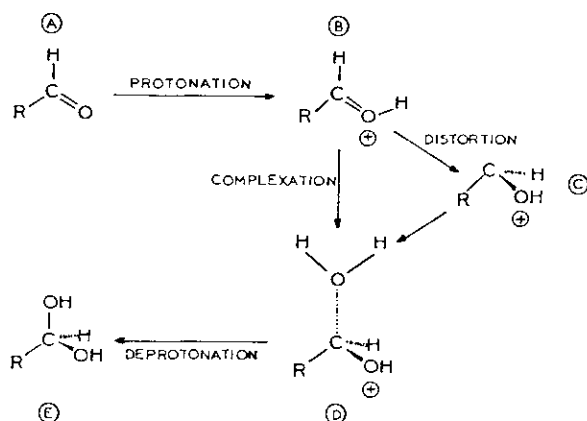


Figure 3. Computed steps in the acid catalyzed aldehyde hydration.

Based on these results for formaldehyde we now proceed, using an exactly similar technique, to a comparative study of the other aldehydes.

We have considered a first and second hydration of glyoxal, the hydration of acetaldehyde, the two possible first hydrations of methylglyoxal (at the formyl group and at the acetyl group, these sites of attack being indicated by arrows for the results contained in Table I), and the second hydration following the more favorable of the two initial reactions. The results of these calculations are shown, in the order described, in Table I.

Interestingly enough it was found that for all these species the most favorable water complex is formed with aldehyde conformation III and that the optimum r_{CO} varies little from the value found for formaldehyde. (The maximum change is to 1.68 Å for the second hydration of methylglyoxal.)

In comparing the likelihood of a given hydration we may make reference to

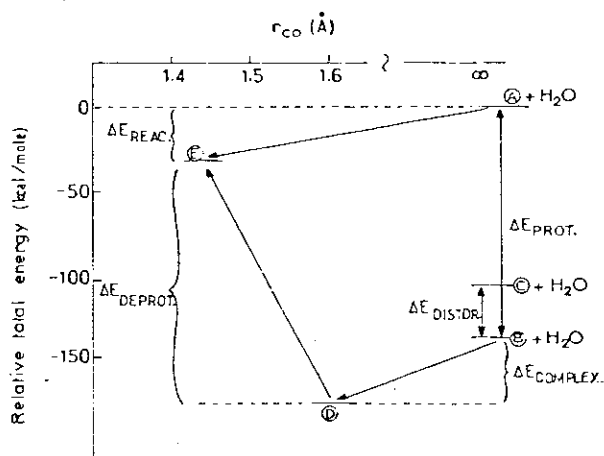









Figure 4. Energetics of the computed steps in acid catalyzed aldehyde hydration.

LAVERY, DE OLIVEIRA, AND PULLMAN

TABLE I. Calculated thermodynamic data for acid catalyzed aldehyde hydration (kcal/mole).

REACTANT	$\Delta H_{\text{PROTONATION}}$	$\Delta H_{\text{DISTORTION}}$	$\Delta H_{\text{COMPLEXATION}}$	$\Delta H_{\text{DEPROTONATION}}$	$\Delta H_{\text{REACTION}}$
	-143.0	+14.3	-33.4	+142.9	-33.5
	-141.8	+16.6	-29.0	+136.3	-34.5
	-146.6	+13.9	-26.7	+133.8	-39.5
	-156.8	+16.6	-24.4	+149.7	-31.5
	-143.8	+11.8	-32.7	+141.3	-35.2
	-151.8	+12.5	-26.6	+150.3	-28.1
	-164.9	+16.1	-18.4	+145.7	-37.6

three of our calculated quantities, namely, the overall exothermicity of the reaction, the energy of complexation, and the energy of deprotonation of the complex. Table I also lists the aldehyde protonation energy and its distortion energy (from conformation I to conformation III) for completeness, although we shall not refer directly to these quantities.

Considering then the first hydration of glyoxal we see that although its complexation energy is somewhat less than that of formaldehyde, its complex deprotonation energy is almost 7 kcal/mole less, and the overall reaction is more favorable by 1 kcal/mole. From Figure 4 it seems clear that the deprotonation of the complex is likely to be the rate-limiting step of the reaction, and consequently as glyoxal yields a deprotonation value considerably smaller than that of formaldehyde and well within the possibility of even a water molecule to achieve (protonation energy = -142.3 kcal/mole), we can expect a ready hydrate formation, as is indeed observed.

Passing to the second hydration of glyoxal the complexation energy has again fallen slightly, but the deprotonation is still easier than for the first glyoxal hydration product, and the overall exothermicity of the reaction has increased by 5 kcal/mole. Thus, again in line with experiment, we would expect this hydration to occur readily.

The picture with acetaldehyde is different. Compared to the hydration of formaldehyde, the overall exothermicity has decreased (by 2 kcal/mole), the complexation energy has decreased (by almost 7 kcal/mole), and the deprotonation energy has increased (by almost 7 kcal/mole) and is now considerably greater than the protonation energy of a water molecule, although it could be achieved by a more strongly proton-accepting anion in solution. Nevertheless, on the basis of the relative unfavorable condition of the three parameters it seems

ACID CATALYZED REACTION OF ALDEHYDES

justified to expect little stable hydrate to be formed for acetaldehyde, which is the experimental observation.

Considering, finally, methylglyoxal and comparing the two possible sites of first hydration, it is clearly seen that hydration at the formyl group is favored. For this site all parameters are slightly more in favor of the reaction than for formaldehyde, with the exception of a complexation energy which is 0.7 kcal/mole less. On the other hand, for hydration at the acetyl group the complexation energy falls by 6 kcal/mole with respect to reaction at the formyl group, the overall exothermicity falls by 7 kcal/mole, and the deprotonation energy is considerably greater at 150.3 kcal/mole. Consequently the first hydration of methylglyoxal would clearly seem to be favored at the formyl group.

We then considered a second hydration of methylglyoxal at the acetyl group, following the hydration at the formyl group just described. This reaction, although it is slightly more favorable than the former on the grounds of exothermicity, shows an exceptionally low complexation energy (-18.4 kcal/mole) and an increased deprotonation energy. Its occurrence therefore seems doubtful. We have not been able to locate any experimental evidence to compare with this estimation.

Conclusion

By our preliminary quantum chemical study of acid catalyzed hydration of aldehydes we have been able to calculate several energy parameters which seem successful in explaining the relative reactivity of a range of aldehydes, including the biochemically interesting species glyoxal and methylglyoxal. However, more remains to be done at a more sophisticated level, to elucidate the details of the reaction mechanism and also to treat the problem of base catalyzed hydration.

Acknowledgment

The authors wish to express their thanks to Professor Albert Szent-Györgyi for his stimulative activity on the biological role of glyoxal and its derivatives and for many exciting discussions on this subject. This work was supported by the National Foundation for Cancer Research.

Bibliography

- [1] A. Szent-Györgyi, L. G. Együd, and J. A. McLaughlin, *Science* **155**, 539 (1967).
- [2] L. G. Együd, J. A. McLaughlin, and A. Szent-Györgyi, *Proc. Nat. Acad. Sci. US* **57**, 1422 (1967).
- [3] A. Szent-Györgyi, *Science* **161**, 988 (1968).
- [4] A. Szent-Györgyi, *Proc. Nat. Acad. Sci. US* **74**, 2844 (1977).
- [5] S. Bonc, J. J. Lewis, R. Pethig, and A. Szent-Györgyi, *Proc. Nat. Acad. Sci. US* **75**, 315 (1978).
- [6] D. Demoulin, A.-M. Armbruster, and B. Pullman, *Theor. Chim. Acta* **48**, 143 (1978).
- [7] D. Demoulin, A.-M. Armbruster, and B. Pullman, *Theor. Chim. Acta* **50**, 75 (1978).
- [8] D. Demoulin, A.-M. Armbruster, and B. Pullman, *Int. J. Quant. Chem.* (in press).

LAVERY, DE OLIVEIRA, AND PULLMAN

- [9] R. Lavery and B. Pullman, *Int. J. Quant. Chem.* (in press).
- [10] L. F. Fieser and M. Fieser, *Organic Chemistry*, 2nd ed. (Heath, Boston, 1950).
- [11] J. D. Roberts and M. C. Caserio, *Basis Principles of Organic Chemistry*, 2nd ed. (Benjamin, New York, 1977).
- [12] L. Radom, P. C. Hariharan, J. A. Pople, and P.v.R. Schleyer, *J. Am. Chem. Soc.* **95**, 6531 (1973).
- [13] A. Pullman, H. Berthod, and N. Gresh, *Int. J. Quant. Chem.* No. 10, 59 (1976).
- [14] M. D. Newton and J. A. Pople, Gaussian 70, Quantum Chemistry Program Exchange, Indiana University, 1973.
- [15] L. H. Funderburk, L. Aldwin, and W. P. Jencks, *J. Am. Chem. Soc.* **100**, 5444 (1978).
- [16] M. Eigen, *Disc. Faraday Soc.* **39**, 7 (1965).
- [17] Y. Pocker and J. E. Meany, *J. Phys. Chem.* **71**, 3113 (1967).
- [18] R. P. Bell and P. E. Sørensen, *J. Chem. Soc. (Perkin II)*, 1740 (1972).
- [19] P. R. Young and W. P. Jencks, *J. Am. Chem. Soc.* **99**, 8238 (1977).
- [20] E. A. Hill and S. A. Milosevich, *Tetrahed. Lett.* **50**, 4553 (1976).
- [21] N. D. Thang, P. Hobza, J. Paneř, and R. Zahradnik, *Coll. Czech. Chem. Comm.* **43**, 1372 (1978).

Received March 16, 1979

APÊNDICE D

CBPF-NF-040/85

RAPID LOCATION OF THE PREFERRED INTERACTION SITES BETWEEN
SMALL POLAR MOLECULES AND MACROMOLECULES. I. BINDING OF
WATER TO THE COMPONENT UNITS OF NUCLEI ACIDS*

by

Marçal de Oliveira Neto

Centro Brasileiro de Pesquisas Físicas - CNPq/CBPF
Rua Dr. Xavier Sigaud, 150
22290 - Rio de Janeiro, RJ - Brasil

*In part this work was carried out at the Institut de Biologie Physico-Chimique,
Laboratoire de Biochimie Théorique, associé au Centre National de la Recherche
Scientifique, 13, rue Pierre et Marie Curie, 75005, Paris, France.

ABSTRACT

An approximate procedure for the rapid detection of favoured sites for the location of bound water on macromolecules has been developed with the aid of accurate electrostatic energy calculations. The method also enables to picture the lability of the bound water molecules around the substrate. As an application the method has been used to study the interaction surface between nucleic acid components and a water molecule.

Key-words: Nucleic acid constituents; Quantum mechanical computations; Electrostatic interaction; Hydration.

I INTRODUCTION

In the last few years, a large number of theoretical studies have been concerned with solute-solvent interactions. Two theoretical approaches are widely employed in such studies: the continuum¹⁻⁴ and the discrete models⁵⁻⁷. Computer simulations based on Monte Carlo calculations using analytical potential functions for the potential energy have been also used⁸⁻¹⁰.

A thorough investigation of the hydration scheme of numerous fundamental biomolecules using the *ab initio* SCF method (within the "supermolecule" procedure) has been carried out applying the discrete approach¹¹⁻¹⁶. The majority of such works has been directed to the main component units of the essential biopolymers, i.e., nucleic acids (pyrimidine bases¹¹⁻¹³; phosphate group¹⁴; sugar ring¹⁵) and proteins (amino acid side chains¹⁶).

However, the extension of these studies to larger molecular systems within the same framework is prohibitively expensive. Further investigations have shown, nevertheless, that a correct picture of the hydration scheme of a substrate mimicking *ab initio* SCF results, may be obtained with the help of an electrostatic approximation alone, with the right cautions about the closest distance between water and substrate¹¹.

In this work we extend the applicability of the electrostatic approach to comprise more complex systems. Our purpose is to develop a procedure by which the general features of the interaction between small dipolar molecules (in particular the very important water molecule) and macromolecules are promptly depicted. Of course, such a technique may not provide the same

precision as an *ab initio* SCF calculation. Nevertheless, one may expect to obtain significant informations on the affinity of biological macromolecules towards water molecule, for example the detection of their specific binding sites. During the development of this procedure it was proposed not only to establish the optimum position of the water molecule and its respective energy, but also to characterize its lability, by investigating throughly all possible orientations of water molecule at each binding site.

In part I of this work we illustrate the procedure to describe the surface interaction between the nucleic acid constituents (guanine, adenine, cytosine, thymine, uracil, phosphate group and sugar ring) and a water molecule. These results may then be used to deduce the effect of inserting these constituents into macromolecular biopolymers. Such effects are described in part II of this work. As a model compound, model segments of poly(dA).poly(dT) and poly(dG).poly(dC) were considered assuming them on the double helical structure of B-DNA.

II METHOD

In order to bring a water molecule into contact with the substrate, we use a technique similar to that previously developed for the calculation of steric accessibility¹⁷. This first consists of surrounding both species with envelopes formed by the juxtaposition of van der Waals spheres centered on all the constituent atoms. Two atoms in mutual contact are then considered: an atom of the molecule to be hydrated (receptor

sphere) and either the oxygen atom or one of hydrogen atoms of the water molecule (attacking sphere). We keep the interaction distance at $2A^{\circ}$ (see ref. 11) for this contact (the radius of receptor and attacking spheres in contact taken as $1 A^{\circ}$). We impose the same approach limit to the neighbouring atoms of the receptor susceptible of forming hydrogen bonds with the water molecule, but between other atom pairs the approach limit is set as the sum of their van der Waals radii. In this way the binding of water to any atom belonging to the receptor molecule can be studied. The orientation of the water is varied by rolling the surface envelope of its attacking atom over the sphere of the receptor atom. Any position where the envelopes of the two species intersect is taken as inaccessible, while for every other position the electrostatic energy of interaction is calculated. The exact positions to be studied for a given atom pair contact are determined by placing a uniformly spaced grid of points defined by a Korobov distribution (see, e.g., ref. 18) on the surface of the receptor and attacking spheres, and restricting contacts to pairs of these points. For any such contact it is further necessary to consider a rotation of the water molecule around an axis joining the centers of the two atoms in contact. Tests have shown that 144 points on the receptor sphere combined with 89 points on the attacking sphere of water and 24 rotation steps lead to convergence of results both regarding the interaction energy and the most favourable positions for the bound water molecules.

The evaluation of the electrostatic interaction between the receptor and water molecules is made by using electron density distributions for both species. For water and for receptors of

normal molecular size this is obtained directly from *ab initio* wave functions. When the receptor is a macromolecule, a single wavefunction cannot be obtained directly and we employ an approximation developed previously for the study of intrinsic electrostatic properties of the nucleic acids¹⁹⁻²². This consists of subdividing the macromolecule into a number of subunits small enough to be studied quantum mechanically and chosen so as to minimize as much as possible the resulting electronic perturbations. Thus for the nucleic acids these subunits are the bases, sugar and phosphate, the points of subdivision being the backbone linkages C_3-O_3 , C_5-O_5 , and the glycosidic bonds. The electrostatic properties of the macromolecules are obtained by a superposition of properties of the subunits appropriately positioned in space.

The wavefunctions are obtained by *ab initio* SCF calculations using a 7s3p/3s Gaussian basis set (10s6p1d for phosphorus) contracted to minimal basis²³. The continuous electron density distributions of each subunit and of the water molecule are replaced by a discrete multicenter overlap multipole expansion, up to quadrupoles, termed OMT^{11,24-25} which has been shown to reproduce accurately the properties of the continuous distribution. The interaction energy between the receptor and water molecule is then calculated as the total interaction between the two sets of multipoles, using the classical electrostatic expressions of the charge-charge, charge-dipole, charge-quadrupole, dipole-dipole, dipole-quadrupole and quadrupole-quadrupole interactions (see, e.g., 26, 27).

There remains a technical problem to be solved before these calculations can be made. The choice described above of

144 receptor points, 89 attacking points and 24 rotation steps implies that roughly 310.000 configurations should be tested for a water molecule in contact with only a single atom of the receptor molecule. Although this number is considerably reduced by forbidding the intersection of the two envelopes, there remain nevertheless between 50000 and 100000 configurations to be studied for a relatively accessible atom. This number is much too high to envisage calculating the full electrostatic interaction energy in all cases, even with our OMTP expansions. The problem is overcome by studying the accessible configurations in several steps with increasing levels of accuracy. The first step consists of representing the water molecule simply by a single dipole and a quadrupole, both calculated from the water SCF wavefunction and placed at the barycenter of the molecule. The energy of a given configuration is then approximated by the product of the dipole and the local electrostatic field vector of the receptor added to the product of the quadrupole and the local field gradient. In this step, moreover, the field and the field gradient are not recalculated for each water position, but rather calculated once for a grid of points surrounding the receptor atom, the point closest to the water molecule barycenter being used in the energy calculation. By using points along radial vectors originating from the center of the receptor atom and passing through accessible points in its Korobov grids, it is possible to ensure that there is always a point within a few tenths of an Angström of the water molecule barycenter. At this level, all the accessible configurations are analyzed, ordered and the best 500 are selected. The second level of approximation reconsiders these selected configurations, keeping the single

center multipole model of water but employing the full OMTP representation of the receptor molecule to calculate the interaction energies. From the resulting energies the best 200 configurations are selected. Finally, for these latter configurations the interaction energy are recomputed, using now the full OMTP representations both for the water molecule and for the receptor. A final reordering yields the best possible water position at this site. Tests showed that this three-step procedure can obtain at least the first 150 water positions correctly, while saving a great deal of computer time. These calculations give not only the optimal energy configuration of the bound water at a given substrate site, but also a whole range of configurations with closely related energies. In this way a view of the lability of the bound water may be obtained: we employ a simple graphic representation of the water positions around the receptor molecule in which we include all the configurations at a given site having energies within 1 Kcal/mole of the optimum energy. With this energy limit between roughly 10 and 150 water configuration were found at each site.

III RESULTS AND DISCUSSIONS

The results obtained by the present technique for each of the subunits studied are displayed in figures 1,3-8. We give also as an example the electrostatic field on the surface envelope of guanine (figure 2). The remaining results and the detailed technique of calculating the field on surface envelopes are described in earlier works (see, e.g., ref. 28).

For the nucleic acid bases the hydration schemes have a central diagram showing the interaction zones in the plane of the base, each of the distinct spatial zones being numbered in order of decreasing binding energy for the optimum configuration belonging to the zone. In order to simplify the hydration schemes we have drawn the water molecule only at its optimal energy position for each zone and for the other water configurations we only indicate with a point the oxygen atom position. Note that we do not show the results of water binding in all weak sites where the optimum interaction energy is less than -3 Kcal/mole. These diagrams were drawn using a graphic plotter connected to a microcomputer system.

The results are reported in table I where we give the interaction energies, the receptor atoms of the subunit involved with each zone and the number NC of water configurations having energies within 1 Kcal/mole of the optimum energy.

Out-of-plane views of the water interaction zones with the bases (figures 1,3-6) are given in two diagrams: at the top the upper half of the central in-plane diagram and at the bottom the lower half of the central diagram. The corresponding zones in the different diagrams are easily associated with one another by their numbering, which remains unchanged. The separation of the out-of-plane views into two diagrams avoids confusion due to the superposition of the interaction zones on opposite sides of the bases. In order to simplify the comparison, the surface field envelopes for guanine (figure 2) are similarly disposed: at the center, a view on the base plane, at the top a view of the upper edge of the base and, at the bottom, a view of its lower edge.

On these envelopes we show both the intensities of the fields, indicated by various degrees of shading (Table II) and their vectorial directions indicated by one of three symbols. These symbols are: a) an arrow \nearrow for vectors which point within 30° of the local envelope tangent; b) a triangle Δ for vectors pointing outwards from the envelope by more than 30° ; c) a distorted cross Δ for vectors pointing inwards by more than 30° . In addition, each of these symbols indicate the direction of the component of the field tangential to the surface envelope. (In the case of the triangle or the distorted cross this direction is from the broader end of the symbol towards its narrower end)²⁸.

The phosphate subunit hydration is treated in a similar manner to that of the nucleic bases; but in this case the "plane" of this unit is taken to be the one containing the phosphorus atom together with two anionic oxygens. For the sugar subunit the plane is taken to be that containing the atoms $C_{1'}$, $C_{2'}$, and $C_{3'}$. In this case only one other view of its hydration perpendicular to the plane was necessary, as only a single water binding site was involved. It should be noted that because of the subdivision scheme we adopt for DNA, the sugar subunit lacks its hydroxyl groups at $C_{3'}$ and $C_{5'}$. It is thus not a complete deoxyribose and it will be used for comparison with the polymer results.

The hydration zones around guanine are shown in figure 1. Five binding zones are seen, of which the strongest (zone 1) is situated between the N_1 -hydrogen and O_6 . The configuration distribution of the water molecule in this region seems to be concentrated in the plane of the base; actually, as shown in the

upper diagram of figure 1, it covers a wide zone on both sides of the base plane. Thus, this strongly bound water molecule (-11.9 Kcal/mole) can be displaced out of the base plane very easily. The same remarks apply to the second most favourable site of guanine which falls between N_7 and O_6 (-11.3 Kcal/mole). The third site is on the opposite side of the base, associated with N_3 , the N_9 -hydrogen and one of the N_2 amino group hydrogen. This interaction zone is rather interesting because it is split into two distinct bands, one on each side of N_3 . The fourth binding site falls between the N_1 -hydrogen and the remaining N_2 -hydrogen, the water configurations having a radial distribution around the former atom. As Table I shows, this zone includes 94 configurations. These configurations, as seen in this region, are divided in four distinct groups: for each group the oxygen atoms are in the same position and the hydrogen atoms in different ones. This division is a consequence of the restriction of the method, for we make use of a discrete number of points on the spheres of the attacking and receptor atoms. Then, it may be expected that these configurations will be displayed in a continuous way in space when the number of points tends to infinite. We verify the same radial distribution for other examples in the present study. In a general way, they do not correspond to water molecules strongly bound to the bases. The remaining hydration site (zone 5) which has a rather weaker binding energy (-7.0 Kcal/mole) is associated with the C_8 -hydrogen.

In the right hand diagrams of figure 2 we show the electrostatic field on guanine surface. With the aid of these diagrams we can see that the four strongest hydration zones can be well correlated with the regions of strongest field. In figure 2a

we can see that three particularly strong field regions are manifest and correspond nicely to the hydration zones 1, 2 and 4. The third group corresponds to the strong field seen in figure 2c and it should be noted that the strong field is divided into two vertical zones in exactly the same way as the hydration zone. The binding site number 5 corresponds to a zone of intermediate field strength. The directions of the field vectors on the envelope of guanine are shown in the left hand diagram of figure 2. The water molecule may be expected to align its dipoles with these field vectors. This helps to explain quite subtle effects, such as the orientation of the water in the two regions forming zone 3 at receptor atom N_3 . Figure 2f shows a rapid change in field direction on each side of N_3 due to the opposing influences of this high electron density atom and of the hydrogens at both sides of it, one on N_9 and the second attached to N_2 . These oppositions, which lead to rapid potential changes, also explain why the field (and thus the hydration at this receptor) is separated in two strong interaction zones, instead of, for example, in a single zone along the external bisector of the angle $C_2-N_3-C_4$. Note also that the radial distribution found for zone 4 correlates very well with the field directions in this region of guanine (figure 2d).

The strongest binding site on adenine is close to N_3 (figure 3). Note that there is no division into two regions at this site as is the case for the corresponding position on guanine. Another binding site can be seen between N_3 and the C_2 -hydrogen (zone 5), but this has a much lower optimum energy (only -5.7 Kcal/mole compared to -12.3 Kcal/mole for zone 1). The second strongest binding is between N_7 and one of the N_6 amino

hydrogens. The resulting configuration presents an important change with respect to the situation in guanine. The water molecule in this region of guanine tends to remain close to the base plane; in adenine, the attraction of N_7 results in the rotation of the water plane so as to turn its hydrogens towards this atom. In the third interaction zone (-10.8 Kcal/mole) the association involves N_1 and the remaining hydrogen of N_6 . The zone 4 of adenine (-6.0 Kcal/mole) involves the C_8 -hydrogen and presents a similar situation to zone 5 in guanine.

The overall comparison of the two purine bases should make clear that the nature of hydration around any given functional group is by no means transferable. The neighbouring atoms always play an important role and often entirely modify the nature of the water binding configurations.

The hydration zones of cytosine are shown in figure 4. Three of these are associated with strong binding energies. The first is shared between N_3 and one of the N_4 amino group hydrogens (-12.7 Kcal/mole); the second lies between O_2 and the N_1 -hydrogen (-11.8 Kcal/mole) and the third between O_2 and N_3 (-11.2 Kcal/mole). The binding site associated to zone 4 (-6.4 Kcal/mole) is located between the remaining amino group hydrogen and the C_5 -hydrogen, this distribution sweeping a great deal of space around the base.

The hydration scheme of thymine (figure 3) shows two relatively strong interactions zones. For the strongest binding, the water molecule is shared between O_2 and the N_1 -hydrogen (-10.8 Kcal/mole). In the case of the second zone (-9.4 Kcal/mole) it is interesting to remark that the binding site is located on each side of N_3 and implies that the water molecule in this region can move very easily around the N_3 -hydrogen so as to interact with

O_2 or O_4 . The third zone, as shown in figure 5, is situated between the hydrogens of C_6 and N_1 . The methyl group interaction with water involves two distinct regions. The first one is associated with the interaction with the methyl hydrogen closest to O_4 and gains a supplementary interaction with this atom (zone 4). The second region is associated with the remaining two hydrogens (zone 5) and has a very weak interaction energy (-3.0 Kcal/mole). However, it is interesting to note that water molecule configurations are radially distributed around the atoms and show one of the largest zones investigated in these studies. It should be noted that this hydration zone results show a very good correlation with those found by Langlet *et al.*²⁹ using simplified formulas to study the hydration of the dimethylphosphate anion (DMP^-); a spherical distribution of weakly bound water molecules was observed around the methyl group of this molecular system.

The hydration of uracil (fig. 6) is very similar to that of thymine. The water molecule configurations around the carbonyl oxygens O_2 and O_4 are distributed in the same way, as well as those for the water molecule between the N_1 - and C_6 hydrogens. The only significant change occurs at C_5 where the absence of the methyl group allows the bound water to approach the base more closely and leads to a slight gain in interaction energy with the adjacent center O_4 (-6.5 Kcal/mole for uracil zone 4 against -6.2 Kcal/mole for thymine zone 4).

The results of our study of the phosphate group $H_2PO_4^-$ in its geometry from B-DNA show high lability associated with strongly bound water molecules. In fact, the phosphate water binding energies are roughly 50% stronger than those of the best among the base sites. The hydration scheme for the phosphate

(figure 7) shows three distinct zones, in which the most strongly associated binding, zone number 1, the water lies between the O_1 and O_2 atoms (-19.0 Kcal/mole). As indicated in Table I zones 2 and 3 are associated with water binding between one anionic and one esteric oxygen (O_1-O_3 , and O_2-O_5 , respectively), the binding energy of the former group being somewhat favoured by the position of the hydrogen bound to O_3 .

Finally, concerning the sugar subunit, we recall that the studied molecule is not a complete deoxyribose and thus its interest is principally limited to comparisons with our next polymer studies. In consequence there is only one interesting hydration site to be studied, namely, around the ring oxygen O_1 . As figure 8 shows the water bound at this site has a high lability, but the interaction energy is rather low (-5.7 Kcal/mole). The sugar out-of-plane view shows that the water positions are distributed more or less uniformly around the external bisector of the $C_1'-O_1'-C_4'$ angle.

IV CONCLUSIONS

The main purpose of the present work is to develop a method which enables, with relatively moderate effort and cost, the description of the interaction surface between macromolecules and small, neutral, dipolar molecules, specifically water molecules. This approach has been applied to describe the interaction surface between the nucleic acid components and a water molecule. The results satisfactorily reproduce the salient features of hydration schemes of these systems brought to light by *ab initio* SCF

computations: pyrimidine bases¹¹⁻¹³, phosphate group¹⁴ and sugar rings¹⁵.

Despite of the great number of theoretical works concerning hydration of nucleic acid components the present study reveals interesting aspects of this process such as the visualization of the water molecule lability; this is shown by employing simple graphical representations of the water positions around the receptor molecule. It has been found that many bound water molecules can be displaced over relatively large regions around their optimum binding position with only very small weakening of their interactions energies. Important spatial zones associated with weakly bound water molecules were observed, for instance, that situated between one of the N_4 amino group hydrogen and the C_5 -hydrogen of cytosine, as well as that associated with the methyl group of thymine. On the other hand, relatively large spatial zones associated with strongly bound water molecules were also observed; for instance, the interaction zone determined around the phosphate group oxygens.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. A. Pullman and Professor B. Pullman for fruitful discussions during the course of this work. He also thanks M.S. de Giambiagi, M. Giambiagi and H.V. Ferreira for several helpful discussions.

This work was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil).

FIGURE CAPTIONS

- FIG. 1 - Interaction zones of a water molecule with guanine (the out-of-plane distributions for sites on the upper edge of the central base are show above and those on the lower edge are show below. For the corresponding interaction energies see Table I. These remarks also apply to figures 3, 4, 5 and 6).
- FIG. 2 - Surface field intensities of guanine (the views of the base are disposed as in the preceeding figure. For the definition of the shading se Table II).
- FIG. 3 - Interaction zones of a water molecule with adenine.
- FIG. 4 - Interaction zones of a water molecule with cytosine.
- FIG. 5 - Interaction zones of a water molecule with thymine.
- FIG. 6 - Interaction zones of a water molecule with uracil.
- FIG. 7 - Interaction zones of a water molecule with phosphate.
- FIG. 8 - Interaction zones of a water molecule with subunit.

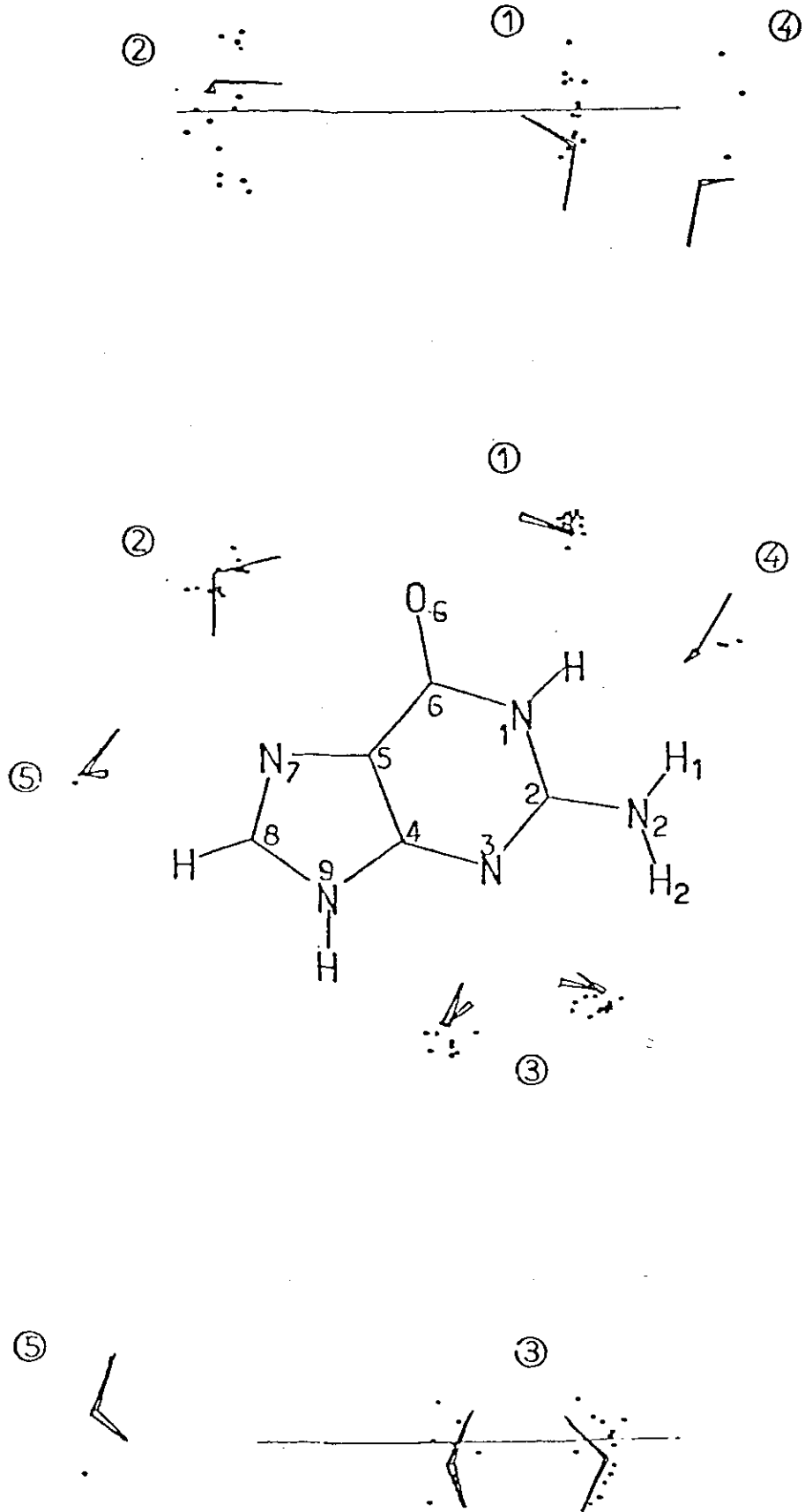


Fig. 1

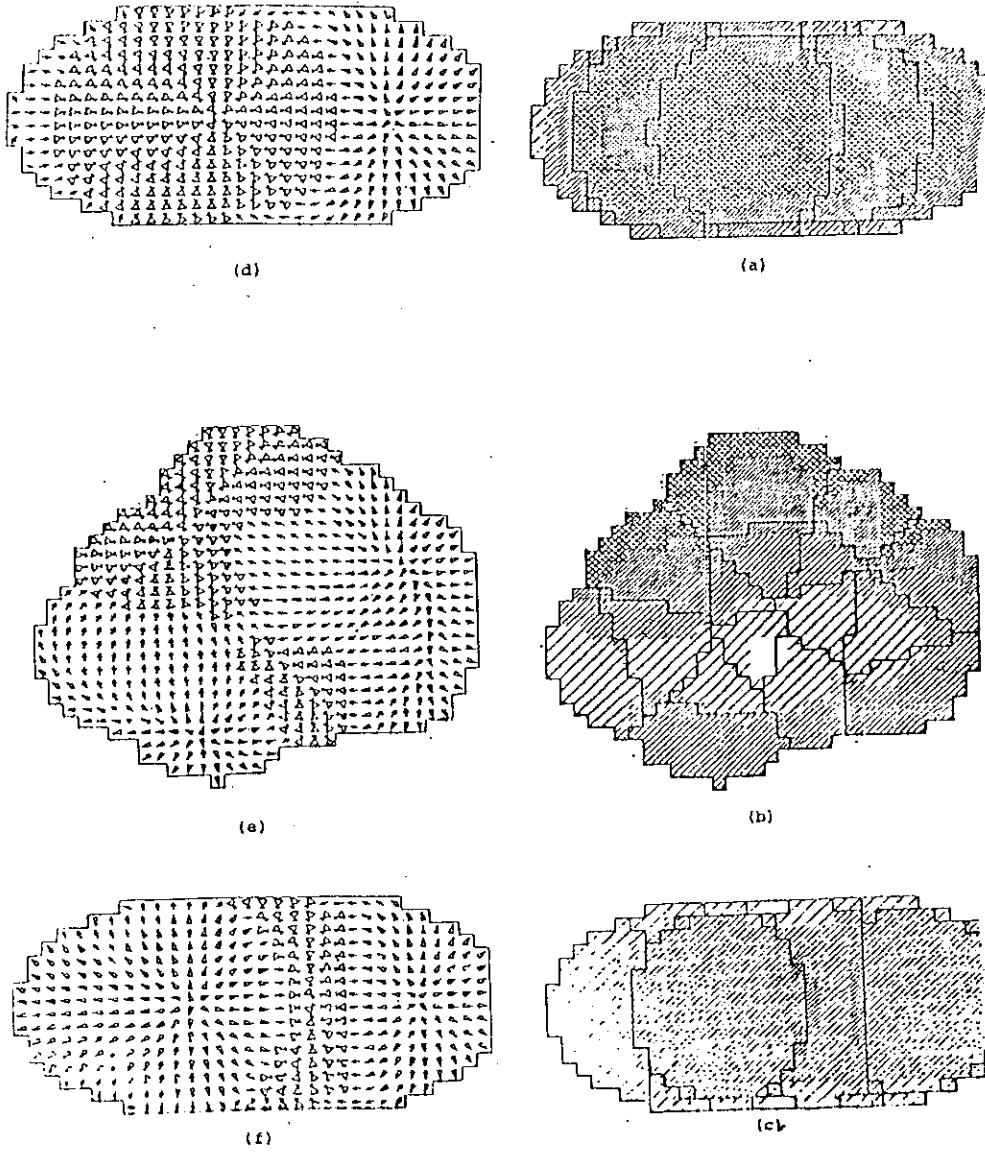


Fig. 2

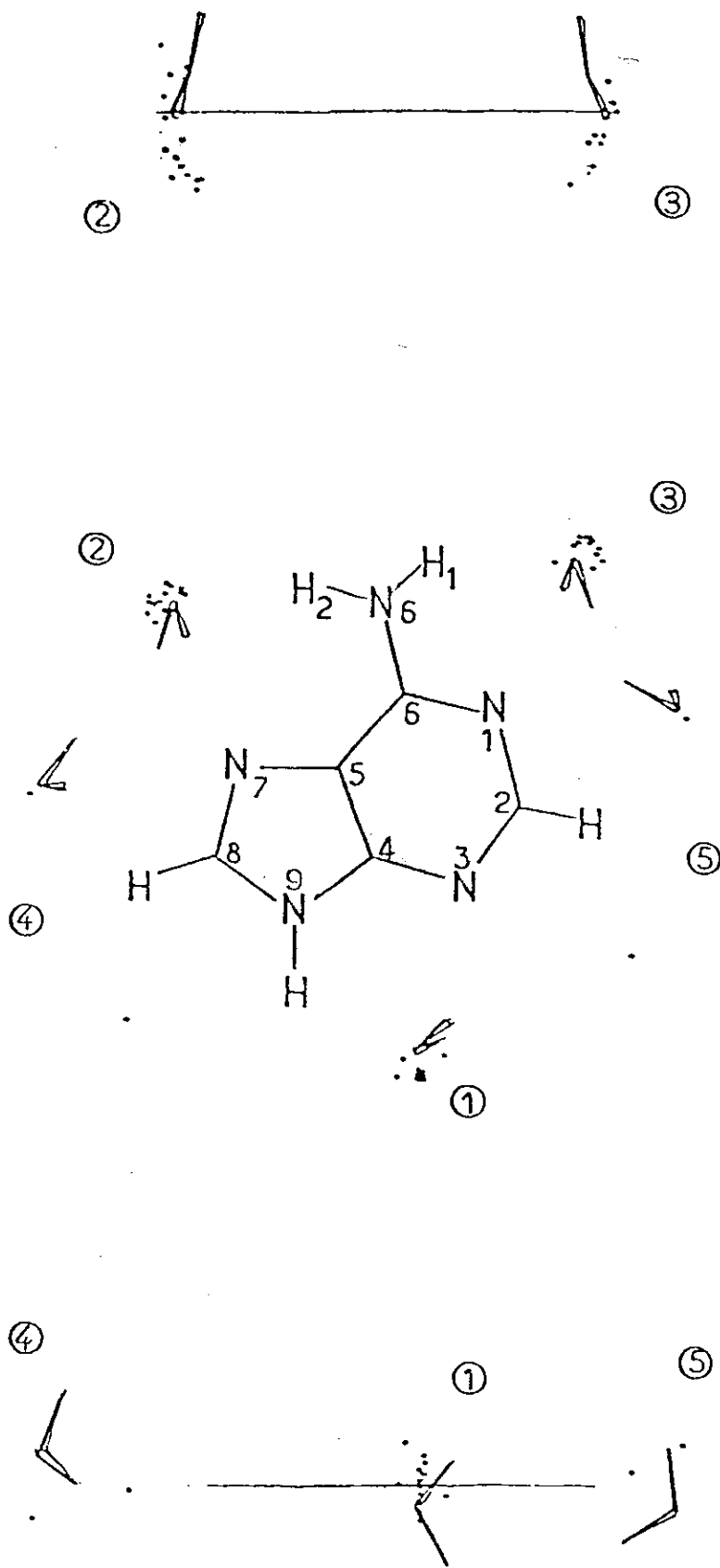


Fig. 3

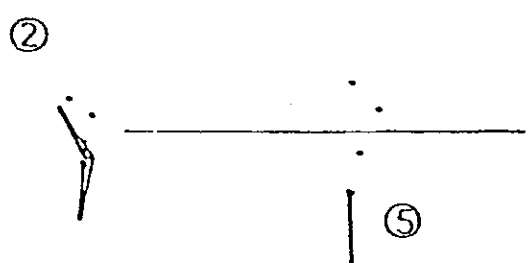
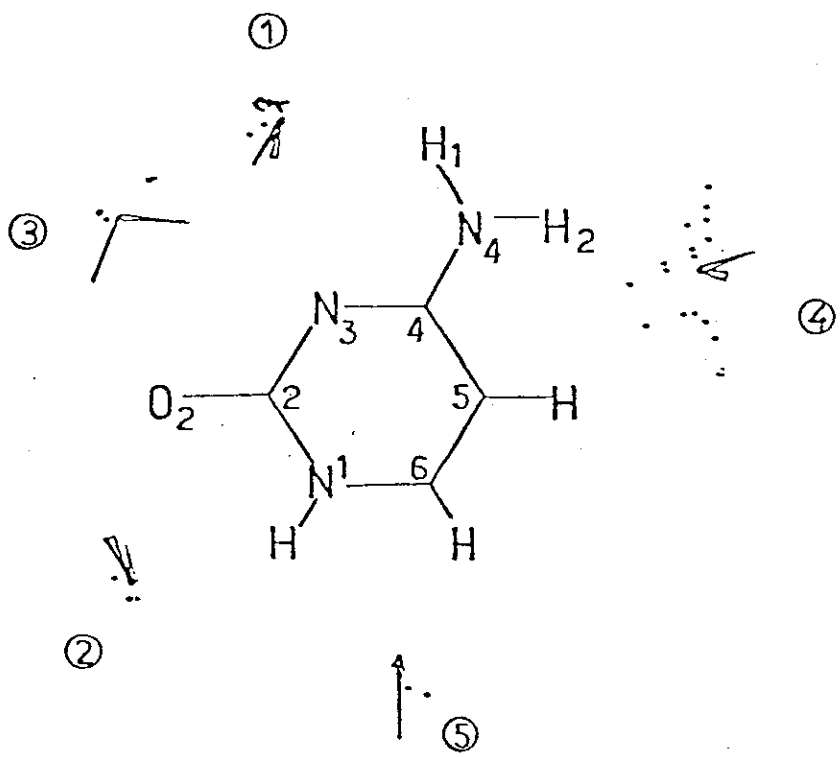
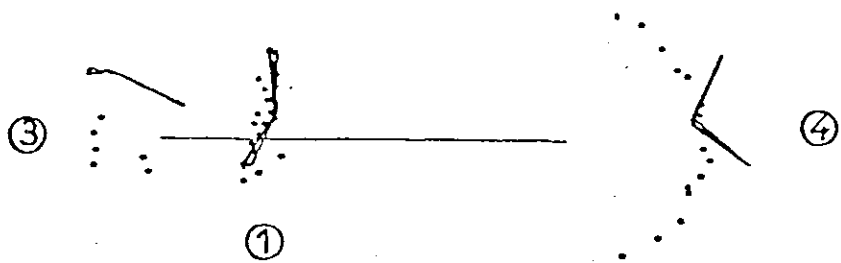


Fig. 4

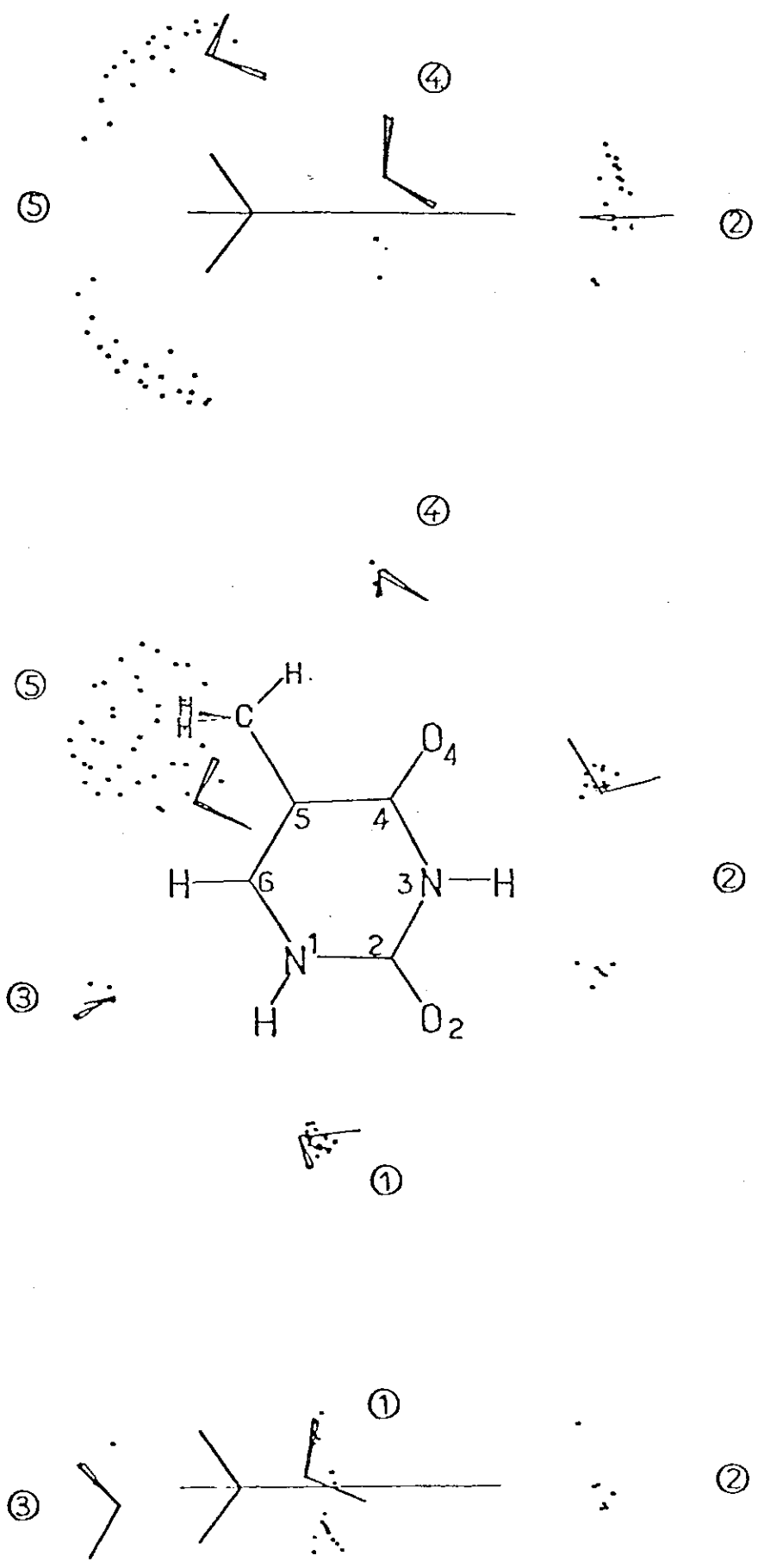


Fig. 5

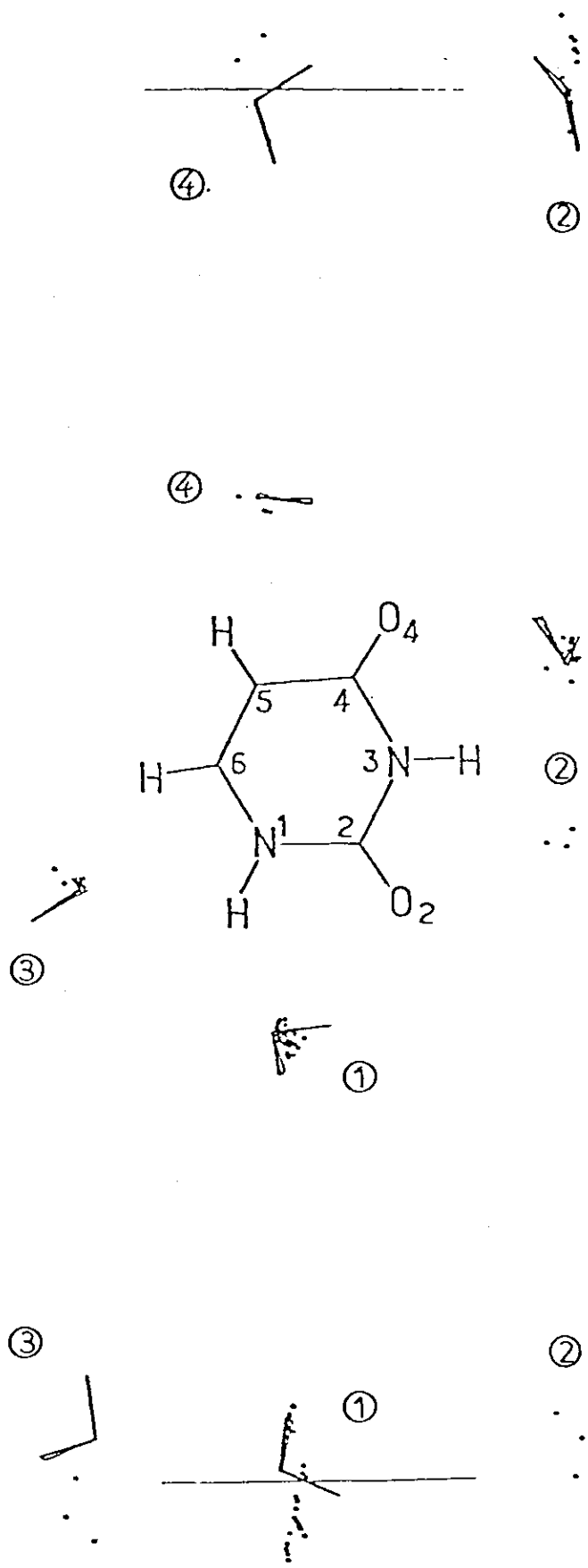


Fig. 6

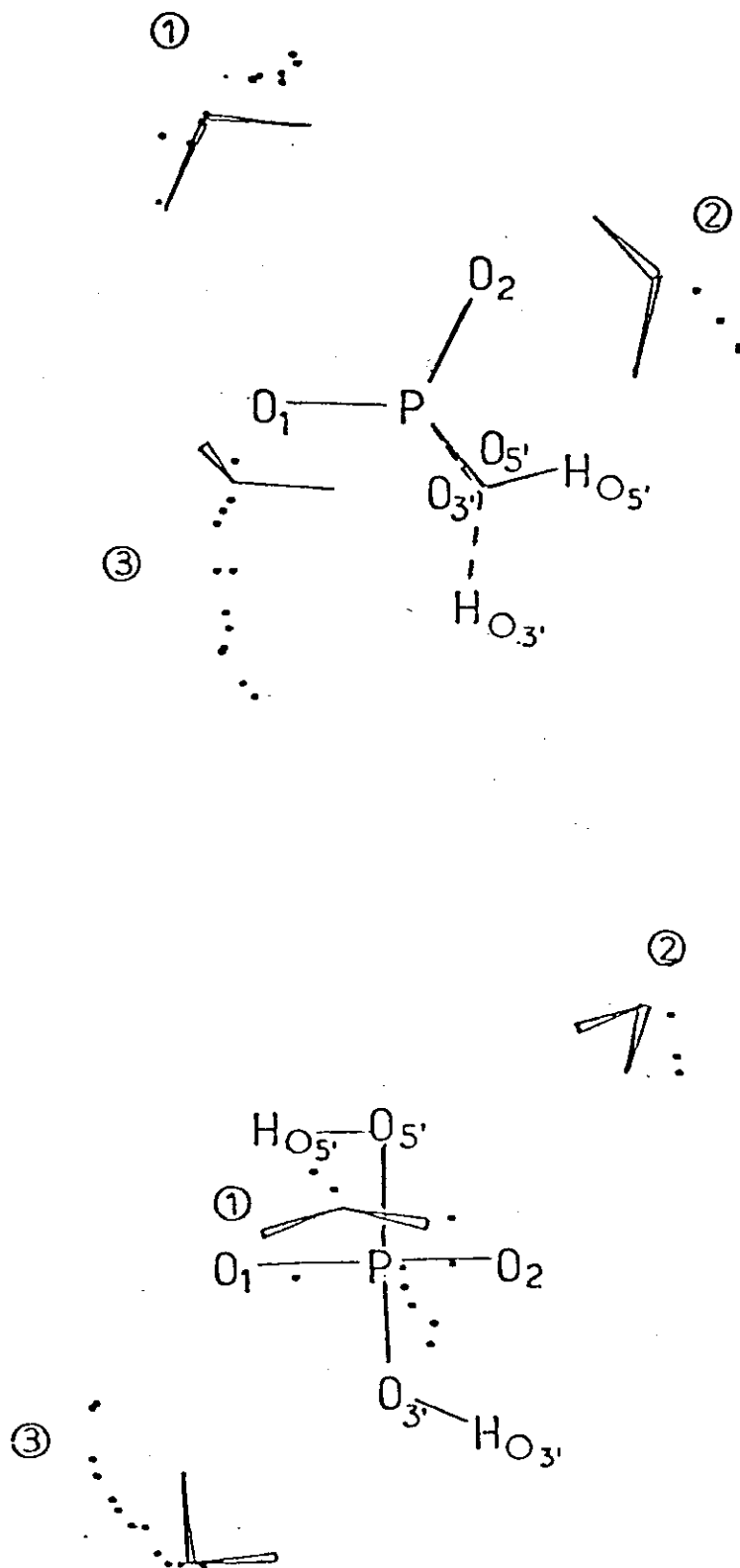


Fig. 7

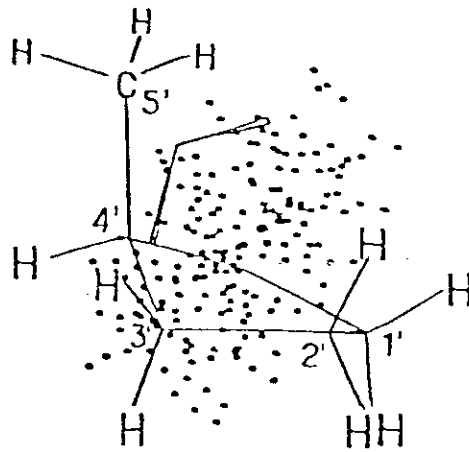
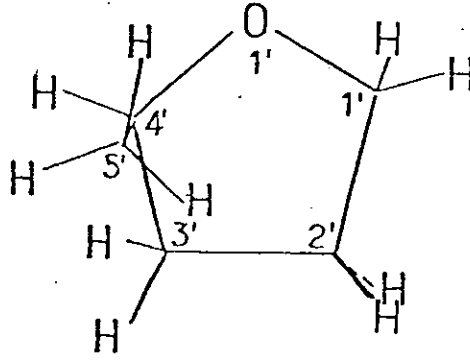


Fig. 8

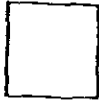






TABLE I

Energies (Kcal/mole) and interaction zones of a water molecule with the nucleic acid components. N_c indicates the number of configurations having energies within 1 Kcal/mole of the optimum energy.

Subunit	Zone	Optimum energy	Receptor atoms involved	N_c
Guanine	1	-11.9	$O_6, H(N_1)$	24
	2	-11.3	N_7, O_6	13
	3	-11.3	$H(N_9), N_3, H_2(N_2)$	22
	4	-10.1	$H(N_1), H_1(N_2)$	94
	5	- 7.0	$H(C_8)$	28
Adenine	1	-12.3	$H(N_9), N_3$	25
	2	-11.4	$N_7, H_2(N_6)$	21
	3	-10.8	$N_1, H_1(N_6)$	22
	4	- 6.0	$H(C_8)$	54
	5	- 5.7	$H(C_2)$	16
Cytosine	1	-12.7	$N_3, H_1(N_4)$	19
	2	-11.8	$H(N_1), O_2$	17
	3	-11.2	N_3, O_2	6
	4	- 6.4	$H_2(N_4), H(C_5)$	349
	5	- 5.9	$H(C_6)$	183
Thymine	1	-10.8	$H(N_1), O_2$	27
	2	- 9.4	$H(N_3), O_4, O_2$	90
	3	- 6.5	$H(C_6)$	90
	4	- 6.2	$H_1(CH_3)$	13
	5	- 3.0	$H_2(CH_3), H_3(CH_3)$	400
Uracil	1	-11.1	$H(N_1), O_2$	48
	2	- 9.1	$H(N_3), O_4, O_2$	54
	3	- 6.7	$H(C_6)$	155
	4	- 6.5	$H(C_5)$	18
Phosphate	1	-19.0	O_1, O_2	12
	2	-17.0	$O_2, O_5,$	4
	3	-16.0	$O_1, O_3,$	13
Sugar	1	- 5.7	$O_1,$	194

TABLE II

Shading used for the surface field intensities (volt/Å⁰) of guanine.

Shading	Field
	0.0
	0.11
	0.22
	0.32
	0.43
	0.54
	0.65

REFERENCES

1. M.J. Huron and P. Claverie, *J. Phys. Chem.*, 76, 2123 (1972).
2. O. Sinanoglu, in "The World of Quantum Chemistry", (R. Daudel and B. Pullman, Eds.), p. 265, Reidel publishing, Dordrecht, Netherlands (1974).
3. D.L. Beveridge, in "Molecular and Quantum Pharmacology", (E. D. Bergman and B. Pullman, Eds.), p. 153, Reidel Publishing, Dordrecht, Netherlands (1974).
4. J. Hylton McCreery, R.E. Christoffersen and G.G. Hall, *J. Am. Chem. Soc.* 98, 7191 (1976).
5. A. Pullman and B. Pullman, *Q. Rev. Biophys.* 7, 505 (1975).
6. A. Pullman, in "The New World of Quantum Chemistry", (B. Pullman and R.G. Parr, Eds.), p. 149, Reidel Publishing, Dordrecht, Netherlands (1976).
7. P. Claverie, J.P. Daudey, J. Langlet, B. Pullman, D. Piazzola and M.J. Huron, *J. Phys. Chem.* 82, 405 (1978).
8. E. Clementi and G. Corongiu, *J. Chem. Phys.* 72, 3979 (1980).
9. E. Clementi and G. Corongiu, in "Biomolecular Stereodynamics" (R. Sarma, Ed.) Adenine Press, New York, p. 209 (1981).
10. E. Clementi, G. Corongiu, M. Gratarola, P. Habitz, C. Lupo, P. Otto and D. Vercauteren, *I.J.Q.C.: Quantum Chemistry Symposium*, 16, 409 (1982).
11. A. Pullman and D. Perahia, *Theoret. Chim. Acta*, 48, 29 (1978).
12. J.E. Del Bene, *J. Comp. Chem.* 2, 188 (1981).
13. J.E. Del Bene, *J. Chem. Phys.* 76, 1058 (1982).
14. A. Pullman, B. Pullman and H. Berthod, *Theoret. Chim. Acta*, 47, 175 (1978).
15. H. Berthod and A. Pullman, *Theoret. Chim. Acta*, 47, 59 (1978).
16. G.N.J. Port and A. Pullman, *Int. J. Quantum Chemistry, Quantum Biol. Symp.* 1, 21 (1974).

17. R. Lavery, A. Pullman and B. Pullman, *Int. J. Quantum Chem.*, 20, 49 (1981).
18. A.H. Stroud, in "Digital Computers User's Handbook", (M. Klerer and G.A. Korn, Eds.), Mc Graw-Hill, N.Y. (1967).
19. A. Pullman, K. Zakrzewska and D. Perahia, *Int. J. Quantum Chem.* 16, 295 (1979).
20. R. Lavery, M. de Oliveira and B. Pullman, *J. Comp. Chem.*, 1, 301 (1980).
21. R. Lavery, A. Pullman, B. Pullman and M. de Oliveira, *Nucleic Acid Res.*, 8, 5095 (1980).
22. A. Pullman and B. Pullman, *Quart. Rev. Biophysics*, 14, 289 (1981).
23. B. Pullman, N. Gresh, H. Berthod and A. Pullman, *Theoret. Chim. Acta* 44, 151 (1977).
24. M. Dreyfus, These de 3^{ème} Cycle, Paris (1970).
25. G.N.J. Port and A. Pullman, *FEBS Letters* 31, 70 (1973).
26. R. Rein, *Adv. Quantum Chem.* 7, 335 (1973).
27. A.D. Buckingham, *Adv. Chem. Phys.* 12, 107 (1967).
28. R. Lavery, A. Pullman and B. Pullman, *Biophys. Chem.*, 17, 75 (1983).
29. J. Langlet, P. Claverie, B. Pullman and D. Piazzola, *Int. J. Quantum Chemistry, Quantum Biol. Symp.*, 6, 409 (1979).

APÊNDICE E

CBPF-NF-041/85

RAPID LOCATION OF THE PREFERRED INTERACTION SITES
BETWEEN SMALL POLAR MOLECULES AND MACROMOLECULES.
II. BINDING OF WATER TO A MODEL SEGMENT OF B-DNA*

by

Marçal de Oliveira Neto

Centro Brasileiro de Pesquisas Físicas - CNPq/CBPF
Rua Dr. Xavier Sigaud, 150
22290 - Rio de Janeiro, RJ - Brasil

* In part this work was carried out at the Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique, associé au Centre National de la Recherche Scientifique, 13, Rue Pierre et Marie Curie, 75005, Paris, France.

ABSTRACT

The procedure developed in Part I of this series is applied to the homopolymeric sequences poly(dA).poly(dT) and poly(dG).poly(dC) on the double helical structure of B-DNA. Some aspects of the base sequence influence on the polymer's attraction for water molecule are described. The results are used to discuss the general hydration features of those systems in relation to recent experimental studies of DNA single crystals.

Key-words: B-DNA; Major groove; Minor groove; Interaction energy; Hydration.

I INTRODUCTION

Chemical or biological properties of macromolecules are often strongly influenced by their surroundings. Among the various environmental factors, the detailed nature of hydration certainly plays an important role. However, due to the size and complexity of macromolecules, a detailed picture is difficult to obtain either from the experimental or the theoretical viewpoint.

Nowadays, since the advent of fast digital computers, it has been possible to study theoretically the hydration of macromolecules such as DNA molecules or proteins¹⁻⁵. These investigations have been carried out using the Monte Carlo method with different atom-atom pair potentials obtained from quantum mechanical calculations.

In part I of this series⁶ a procedure was developed to take into account both the sterical and energetic aspects of the interaction between small polar molecules and macromolecules. The intermolecular interaction energies were computed through the overlap between the multipole expansions of the charge distributions of the interacting entities⁷⁻⁹. The macromolecule is built from its constituent fragments and the multipolar expansions (up to quadrupoles) of the charge distributions of each constituent are derived from *ab initio* SCF wave functions. This procedure is the same as that adopted in a large number of theoretical studies devoted to the reactivity of nucleic acids based on their molecular electrostatic potentials or fields¹⁰⁻¹³.

As an example, the above procedure was applied, in part I, to the interaction surface between the component units of nucleic acids and water molecules. By investigating thoroughly

all orientations of a water molecule at each binding site, it was possible to characterize its lability.

Following the line of the previous work⁶, an extension of this study to a model fragment of B-DNA is presented now. We show primarily the results regarding model segments of the homopolymers poly(dA).poly(dT) and poly(dG).poly(dC), for which it is assumed a double helical structure of B-DNA. These two homopolymeric sequences allows a description of some aspects of the base sequence influence on the polymer's affinity towards a water molecule.

Subsequently, a model is proposed for a relatively complete first hydration layer in the minor groove of the homopolymeric sequences. This attempt to go a step further is motivated by recent X-ray diffraction studies of single crystals of oligomeric DNAs; fragments of this molecule in different conformations, B¹⁴⁻¹⁸, A¹⁹⁻²¹ and Z^{22,23} have been studied and in several cases the positions of cations, polycations and water molecules appearing in the crystals have been also located with precision. We further discuss the present theoretical results and those concerning hydration of single crystals of oligomeric B-DNA to establish the significant correlations between them.

The present study is restricted to the standard B-DNA conformation²⁴ of the oligomers.

II RESULTS

The results regarding water binding to the major groove, minor groove and phosphates groups of the B-DNA model are presented in Sections II.1, II.2 and II.3, respectively.

The model polymer consists of one turn of the double helix with 11 phosphates in each strand and homopolymeric base sequences $(GC)_n$ and $(AT)_n$. In section II.4 we describe the model adopted in order to obtain a profile of the second hydration layer in the minor groove of the oligomeric sequences.

Unless otherwise stated water molecules were bound to the central nucleotides of the model double helices to minimize end effects. The binding of water molecule to bases electro-negative heteroatoms, to amino group hydrogens (other than those involved in the hydrogen bonds of the nucleic acid base pair), to furanose and ester oxygens of the phosphodiester linkage and to phosphate group anionic oxygens are investigated. Standard notations are employed for the atoms of the various components of DNA, for example, $N_7(G)$ for the atom N_7 of guanine, $O_{1'(S)}$ for the sugar oxygen, $H_6(A)$ for the N_6 -adenine free amino hydrogen and so on.

Table I shows the studied oligomer regions, DNA component atoms involved in the interaction site, the optimum energies and the number NC (see ref. 6) of water configurations having energies within 1 Kcal/mole of the optimum energy.

Figures 1 to 6 depict the first layer water molecule interaction sites in poly(dA).poly(dT) and poly(dG).poly(dC) grooves for which we have employed only three nucleotides of each double helix strand. In order to simplify the visualization of these diagrams we indicate the water molecule at its optimum energy position; the oxygen atoms of the other water molecule configurations are indicated by a shaded zone. All figures in the present work were drawn with the help of a graphic plotter connected to a Apple II microcomputer.

11.1 MAJOR GROOVE

As indicated in table I, the strongest binding site in poly(dA).poly(dT) major groove lies between the $N_7(A)$ and $H_6(A)$ atoms. This site is virtually the same as that located for isolated adenine⁶ but in the polymer its binding energy becomes more favourable by 3.2 Kcal/mole. However, for the second site located in the major groove, this extra stabilization is not found when a water molecule interacts with $O_4(T)$; in this case the interaction energy falls by 1 Kcal/mole in comparison to that of the isolated base. In figure 1 the bound water habits (that in a bridge position between $N_7(A)$ and $H_6(A)$ and the one bound to $O_4(T)$ are denoted by W_1 and W_2 , respectively) can be visualized: W_1 is not very labile whereas W_2 can be displaced over a wide region in the major groove of the polymer.

In poly(dG).poly(dC) major groove the strongest binding site giving rise to an interaction energy of -15.3 Kcal/mole is associated with a bound water molecule lying between the $N_7(G)$ and $O_6(G)$ atoms (denoted by W_3 in figure 2). This configuration is more stable by 4 Kcal/mole than the corresponding one in isolated guanine⁶ (although in equivalent binding site). The second located site involves the N_4 free amino hydrogen of cytosine (W_4 in figure 3). We recall that for isolated cytosine the optimum interaction energy of a water molecule with this atom is -6.4 Kcal/mole; in the macromolecule this value increases to -11 Kcal/mole despite a relatively high water lability.

11.2 MINOR GROOVE

If we first consider binding to the poly(dA) strand of

poly(dA).poly(dT), energy optimization indicates that in preferred location the water molecule is shared by $N_3(A)$ and O_1' atom of the sugar bound to the consecutive adenine in the sequence $5'A_pA_3'$. The optimum energy in this case is -12.3 Kcal/mole which coincides with that found for water binding between $N_3(A)$ and the N_9 -hydrogen in isolated adenine.

As to bindings to the poly(dT) strand our calculations show that the resulting configuration bridges the $O_2(T)$ and $O_1'(S)$ atoms, with an interaction energy of -11.9 Kcal/mole. The resulting bond is oriented similarly to that obtained for the poly(dA) strand.

It may be remarked that in figure 3 the water molecules bound to poly(dA) and to poly(dT) strands (denoted by W_5 and W_6 respectively) presents labilities significantly smaller than the corresponding water molecule lability found for the association with $O_1'(S)$ in the isolated sugar subunit⁶. For this latter case the interaction energy is -5.7 Kcal/mole; however, this water molecule gains an extra stabilization when embedded in the macromolecular system.

In the case of poly(dG).poly(dC) minor groove the stronger binding (-11.8 Kcal/mole) is associated with a water molecule (W_7 in figure 4) interacting simultaneously with three centers: one of these centers is the $N_2(G)$ free amino hydrogen of the central guanine and the other two are the $O_2(C)$ and $O_1'(S)$ atoms both belonging to the nucleotide on the 3' side of the central cytosine. As reported in table I the resulting interaction energy involving a water molecule and the $N_3(G)$ atom in poly(dG).poly(dC) minor groove is less favourable than that mentioned above and the binding site (W_8 in figure 4) is located in the neighbourhood of the guanine amino group. Analyse of the

distance between the water molecules W_7 and W_8 shows that both sites cannot coexist due to steric hindrance. Hydration of $O_{1,(S)}$ atom on poly(dG) strand yields a water configuration very similar to that found for $N_3(G)$, the binding energy being, however, 2 Kcal/mole less favourable. According to these results it would seem that only one binding configuration can be found in this region of poly(dG).poly(dC) minor groove. We shall return to this point later.

Figure 5 and 6 depict a view of the bound water configurations in the major and minor groove of poly(dA).poly(dT), and poly(dG).poly(dC), respectively, on a plane perpendicular to their helical axis.

II.3 PHOSPHATE GROUPS

The strongest binding energies within the two studied sequence models are associated with phosphate groups (see table I). The binding of water to anionic oxygens is favoured both from sterical and energetic considerations. For instance, O_3 atom is sterically highly hindered preventing it from directly attacking a water molecule. The resulting configurations show high water labilities associated with the anionic oxygens. A water molecule may bridge the O_1 and O_2 atoms and also the O_2 and O_5 atoms through a hydrogen bond, although the latter is less favoured. Figure 6 shows, as an example, the poly(dG).poly(dC) sequence in which the water molecule W_9 is situated between the two anionic oxygens of the phosphate group. As indicated in table I there are about a hundred possible configurations for W_9 ; such configurations spread over a wide region (also depicted in figure 6) around the anionic oxygens.

II.4 A MODEL TO INVESTIGATE A SECOND HYDRATION LAYER IN MINOR GROOVE

Such a model consists in the preliminary construction of a relatively complete first hydration layer. In the case of poly(dA).poly(dT) minor groove we have located the two bound water configurations W_5 and W_6 . These two molecules are distant enough from each other, as the minimal distance between two of their atoms is approximately 5.5 \AA ; consequently it may be assumed these binding associations do not interfere with subsequent hydration. In this way our model for a first hydration layer in minor groove has been built with the following approximation: in view of the homopolymeric poly(dA).poly(dT) sequence we inserted into the polymer at each base pair water molecules identical to W_5 and W_6 , i.e., the configurations at the minimal energy position found for the central base pair. Figure 7 shows the resulting scheme: W_5' , W_5 and W_5'' represent the binding configurations on poly(dA) strand; W_6' , W_6 and W_6'' the corresponding ones on poly(dT) strand. Now, the water molecules are integrated to the macromolecular system and their multipoles are also taken into account for the interaction energies computations. In order to obtain a profile of the second hydration layer we then studied a subsequent binding with the two central W_5 and W_6 water molecules. The resulting hydration scheme indicates that it is constituted by bridging the binding configurations on poly(dA) strand and those on poly(dT) strand. However, this bridge does not link water molecules associated with one base pair, but instead a water molecule bound to adenine and the following one in the 5'-3' direction bound to thymine. The interaction energy for this binding (-17.8 Kcal/mole) is markedly larger than those obtained

for water molecules of first hydration layer. In figure 7 it is shown the resulting configurations W_{10} (bridging W_5 and W_6'') and W_{11} (bridging W_5' and W_6).

As to poly(dG).poly(dC), we recall that apparently only the water molecule W_7 may associate with the central nucleotides in minor groove. In order to investigate the structure of water molecules in this region we have adopted a similar procedure to that employed for the poly(dA).poly(dT) case. Nevertheless, it is necessary to study the possibility of binding at the hydrophilic centers $N_3(G)$ and $O_1(S)$ of the poly(dG) strand, taking into account the presence of other water molecules identical to W_7 in the minor groove of poly(dG).poly(dC). The results indicate that binding to O_1 , sugar oxygen (-9 Kcal/mole) is preferred to binding to $N_3(G)$ atom by an amount of interaction energy about 3 Kcal/mole. In figure 8 W_7' , W_7 and W_7'' denote the resulting configurations on poly(dC) strand; W_8'' , W_8' and W_8''' those on poly(dG) strand.

Subsequent bindings to W_7 and W_8' molecules shows that in contrast to the poly(dA).poly(dT) minor groove, no cross strand bridging occurs and that, in fact, the binding configurations on poly(dC) strand are inaccessible to direct attacking of subsequent water molecules. Yet, the corresponding ones on poly(dG) strand can accept another strong hydrogen bonded water molecule (-20.5 Kcal/mole) locating in the neighbourhood of the backbone phosphate groups (W_{12} in figure 8).

A view of both poly(dA).poly(dT) and poly(dG).poly(dC) hydration scheme on a plane perpendicular to the helical axis is presented in figures 9 and 10, respectively.

III DISCUSSIONS AND CONCLUDING REMARKS

The above results indicate that when comparing hydration features in the minor and major grooves of the B-DNA model, as functions of its base sequence, relevant differences can be established. We remark that the major groove of both poly(dA).poly(dT) and poly(dG).poly(dC) have nearly the same general hydration features, the binding being associated with a single base. Moreover, the water molecule can bridge between two base atoms, which is the case for the site involving $N_7(A)$ and $H_6(A)$ or that involving $N_7(G)$ and $O_6(G)$. The highest labilities in the major groove are associated with the water molecule bound to $O_4(T)$ and that bound to $H_4(C)$ for which there appears to be no preferential neighbouring atom available to render these associations bidentates.

However, bidentate bound water configurations can be formed between atoms of different component units of the double helix and such states have been found in the two oligomer's minor grooves. This is the case for the bridge formed between base edge atoms $O_2(T)$ or $N_3(A)$ and $O_1(S)$ in poly(dA).poly(dT) and that involving $O_2(C)$, $H_2(G)$ and $O_1(S)$ in poly(dG).poly(dC) minor grooves.

A tentative investigation of a subsequent layer hydration in the minor groove of the B-DNA model reveals clear contrast in their hydration features and only for poly(dA).poly(dT) a cross-linking of the first hydration layer is observed. The presence of the guanine amino group in the poly(dG).poly(dC) minor groove appears to play an instrumental role in inhibiting such cross-linking.

Through the comparison of the two studied sequences we may deduce that not only the interaction energy of bindings in the grooves are strongly correlated with the base sequence involved, but also the geometrical arrangement of the water molecules; this effect is emphasized in the minor groove case.

As we have mentioned Dickerson and collaborators X-ray studies of a single crystal oligomeric DNA allowed to locate water molecules in different regions of the B-DNA self-complementary dodecamer CGCGAATTCGCG¹⁶⁻¹⁸. Here, it is important to remark that we have carried out the present investigations regarding oligomers with homopolymeric base sequence having the classical B-DNA conformation. In the case of Dickerson's dodecamer geometry the base sequence has a more complicated arrangement and presents noticeable local structural heterogeneities. Nevertheless a theoretical interpretation of its NMR spectrum suggests that, in solution, its conformation may be considerably closer to that of a regular B-DNA²⁶.

The results of the dodecamer crystal, indicate that one of the most striking hydration feature observed is a regular spine of ordered water molecules binding between successive adenine N₃ and thymine O₂ atoms in the minor groove in the AT-rich center of the polymer. Moreover, the second hydration layer is found to bridge this first layer in an approximate tetrahedral bound configuration. This regularity is disrupted on passing from the AT center of the self-complementary dodecamer towards its GC ends. In contrast, hydration of major groove showed that most water molecules in the first layer interact with nitrogen or oxygen atoms on the edges of the base pair, and the majority of the binding observed are associated

with one single base. The hydration geometry in the major groove did not present a regularity comparable to the hydration in the minor groove.

Since the classical work of Falk *et al.*²⁶⁻²⁸ the preferential hydration of the anionic oxygen atoms in the phosphate groups of DNA is well established. This feature was nevertheless invisible in the primitive Dickerson's X-ray results at room temperature and insufficiently resolved spectrum of the dodecamer crystal CGCGAATTCGCG, indicating thermal or static disorder of the backbone atoms¹⁶. A further investigation of the same dodecamer at 16°K¹⁷ and of a bromo derivative d(CGCGAATT^{Br}CGCG) in a 60% MPD (2-methyl 2,4-pentanediol) solution¹⁸ enabled a refinement of the initial conclusions, indicating the phosphates as the sites of strongest hydration. This correlates with our findings since strong binding may be observed favouring the anionic oxygens of the phosphate groups. Moreover the largest water labilities associated to these binding sites seems to be one of the most important factors rendering difficult the observation of water molecules in this regions of the dodecamer crystal.

Although the present theoretical investigation refers to classical B form of DNA it seems to account satisfactorily for experimental observations such as those relevant to the structure of hydration in major and minor groove as well as those relevant to the affinity of this biomolecule for water molecules.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. A. Pullman and Professor B. Pullman for fruitful discussions during the course of this work. He also thanks M.S. de Giambiagi, M. Giambiagi and H.V. Ferreira for several helpful discussions.

This work was partially supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil).

FIGURE CAPTIONS

- FIG. 1 - Preferred interaction sites of a water molecule in poly(dA).poly(dT) major groove.
- FIG. 2 - Preferred interaction sites of a water molecule in poly(dG).poly(dC) major groove.
- FIG. 3 - Preferred interaction sites of a water molecule in poly(dA).poly(dT) minor groove.
- FIG. 4 - Preferred interaction sites of a water molecule in poly(dG).poly(dC) minor groove.
- FIG. 5 - A view on a plane perpendicular to the helical axis of poly(dA).poly(dT). The three base pair planes are placed at the distances:
 Z (———), Z-h (-----), Z+h (———), o oxygen atoms,
 • nitrogen atoms, hydrogen bonds involved.
- FIG. 6 - A view on a plane perpendicular to the helical axis of poly(dG).poly(dC).
- FIG. 7 - A schematic representation of the partial first and second hydration layer in poly(dA).poly(dT) minor groove.
- FIG. 8 - A schematic representation of the partial first and second hydration layer in poly(dG).poly(dC) minor groove.
- FIG. 9 - A view on a plane perpendicular to the helical axis of the scheme obtained in poly(dA).poly(dT) minor groove.
- FIG. 10 - A view on a plane perpendicular to the helical axis of the scheme obtained in poly(dG).poly(dC).

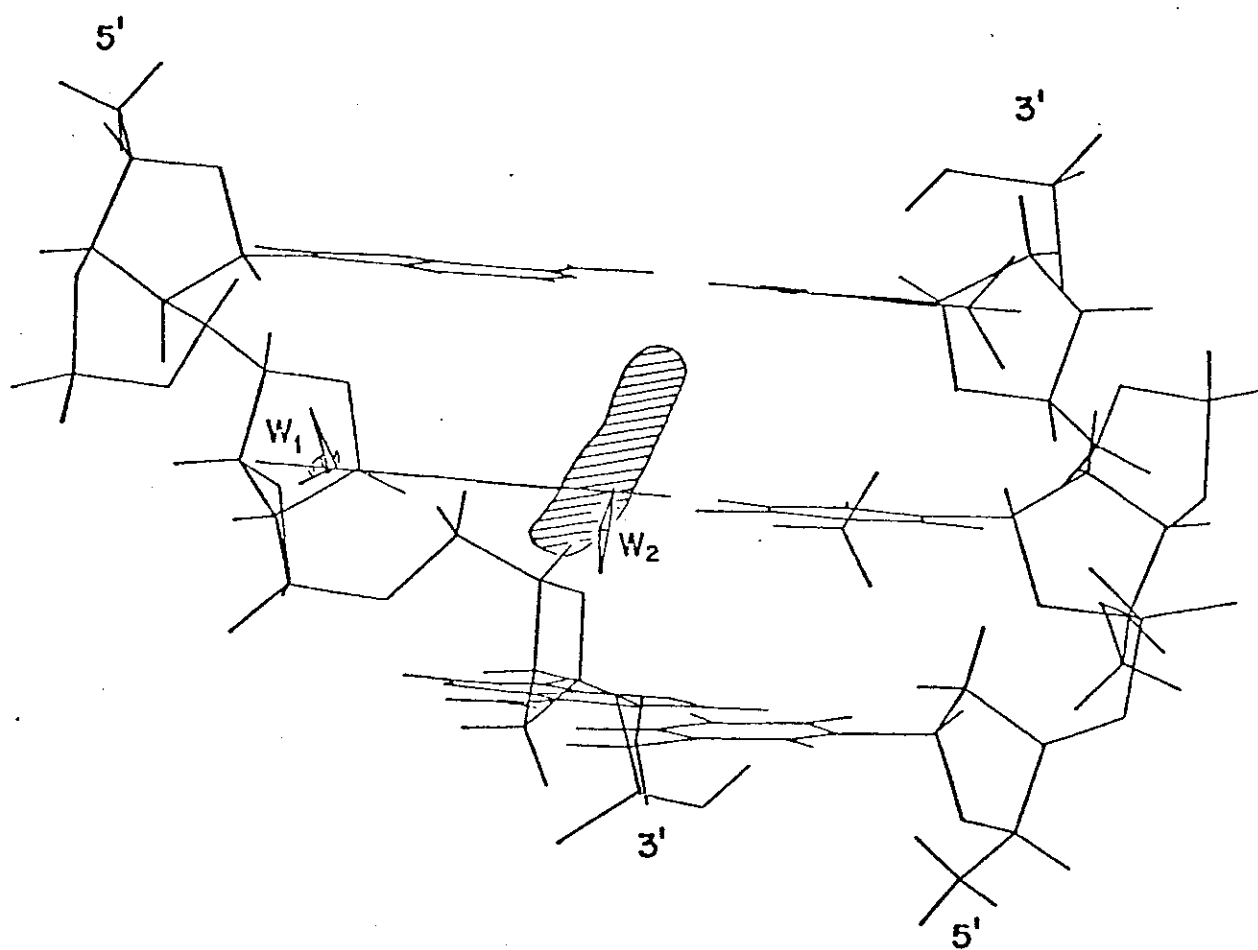


FIG. 1

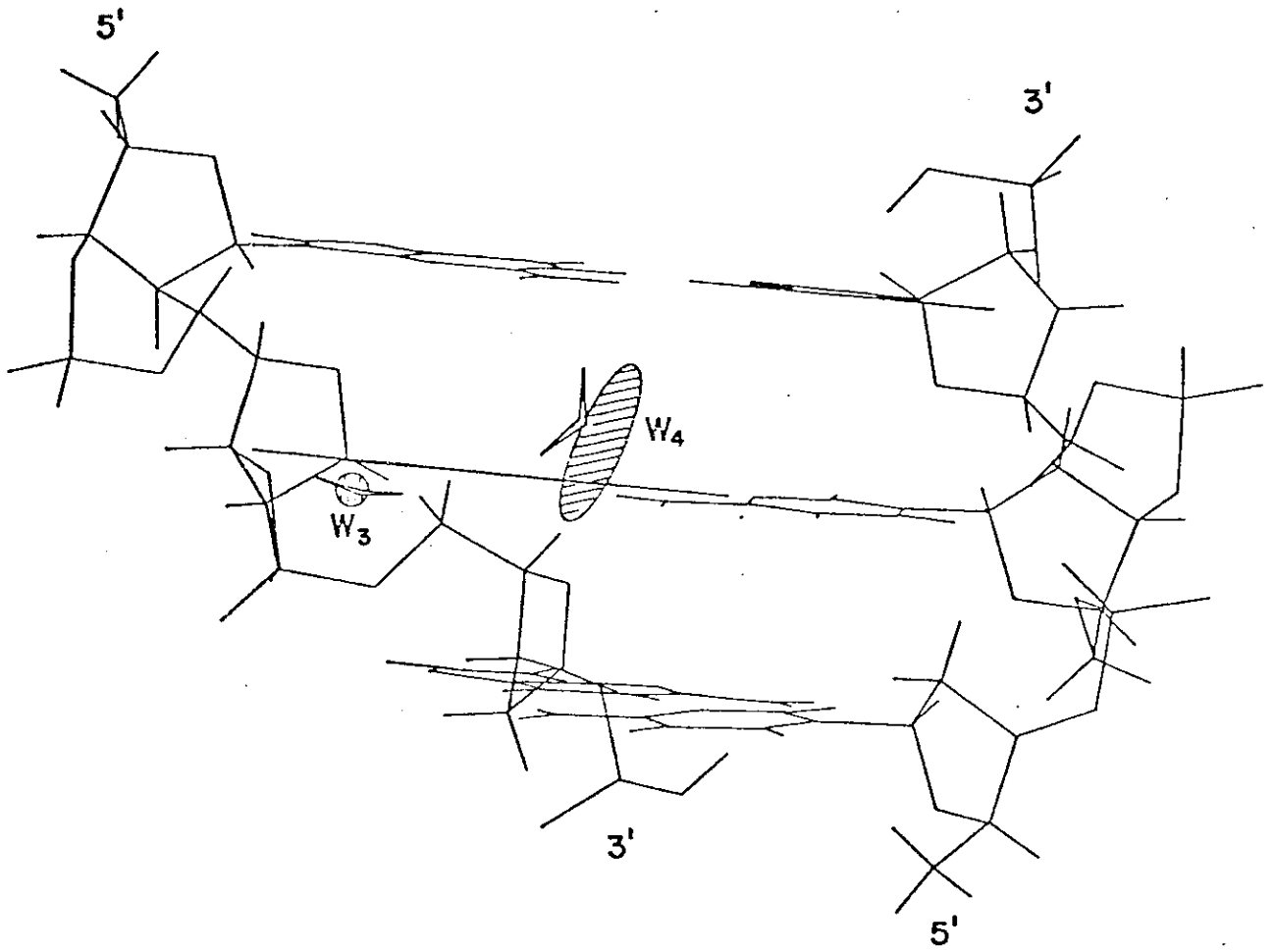


FIG. 2

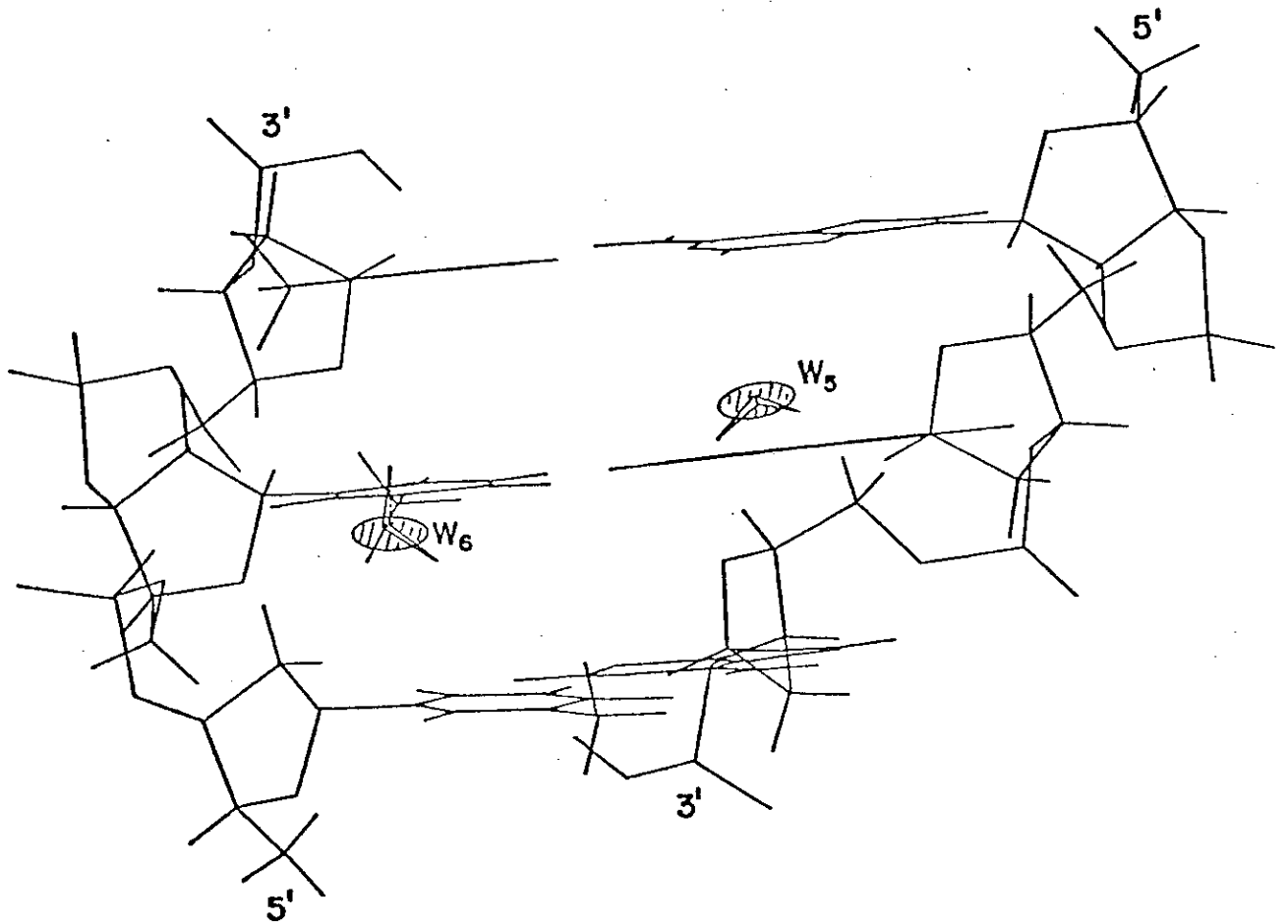


FIG. 3

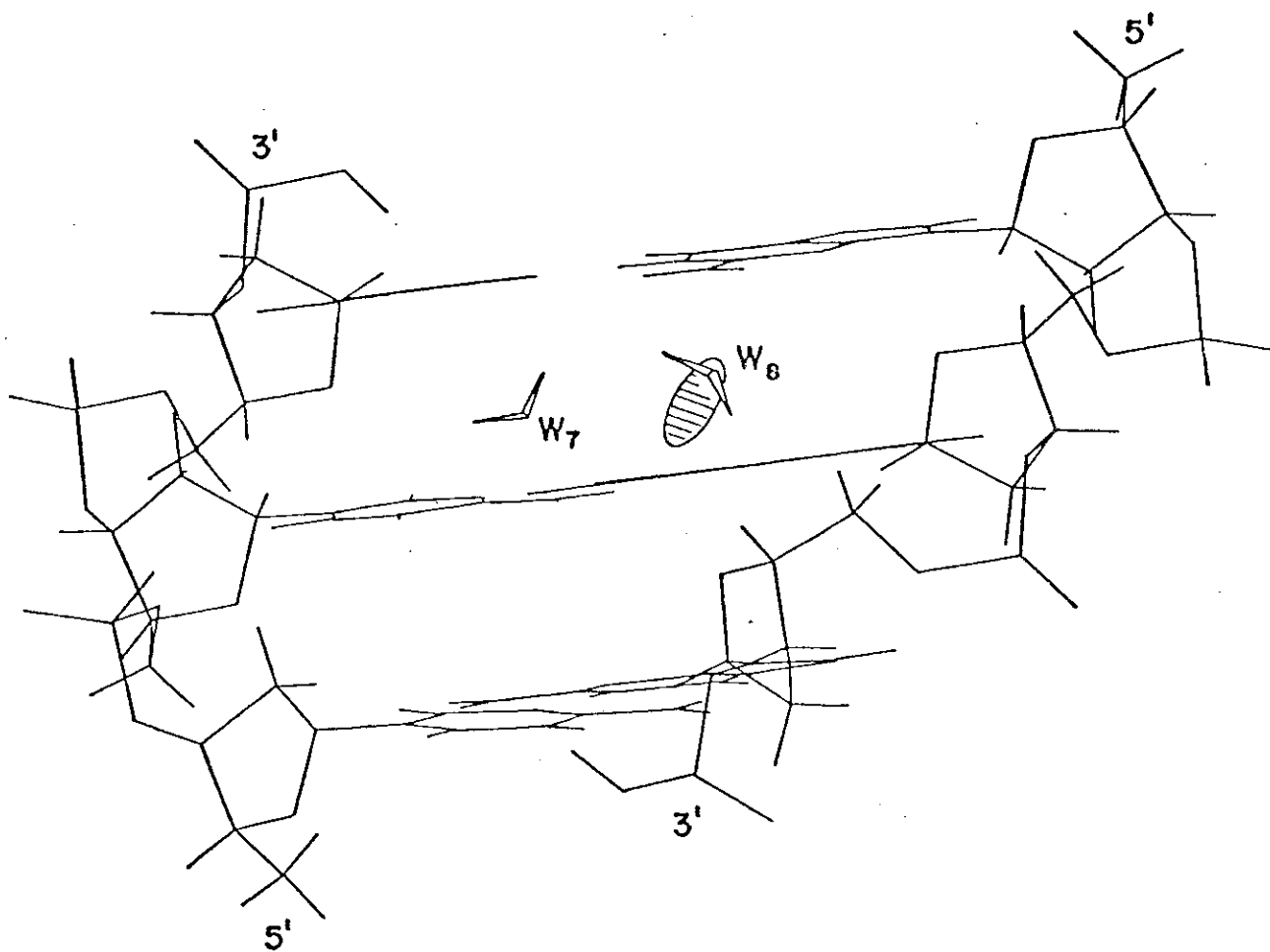


FIG. 4

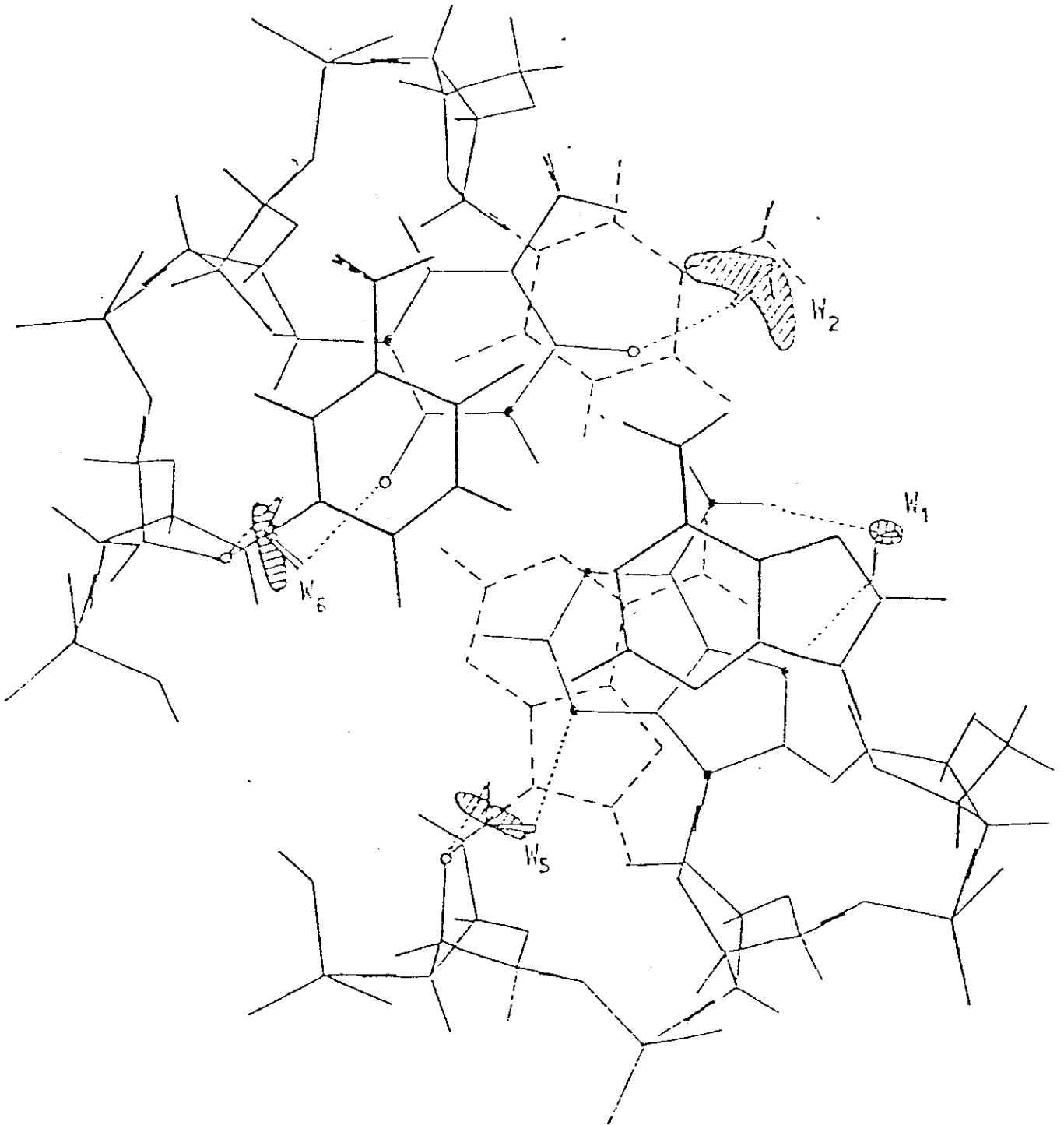


FIG. 5

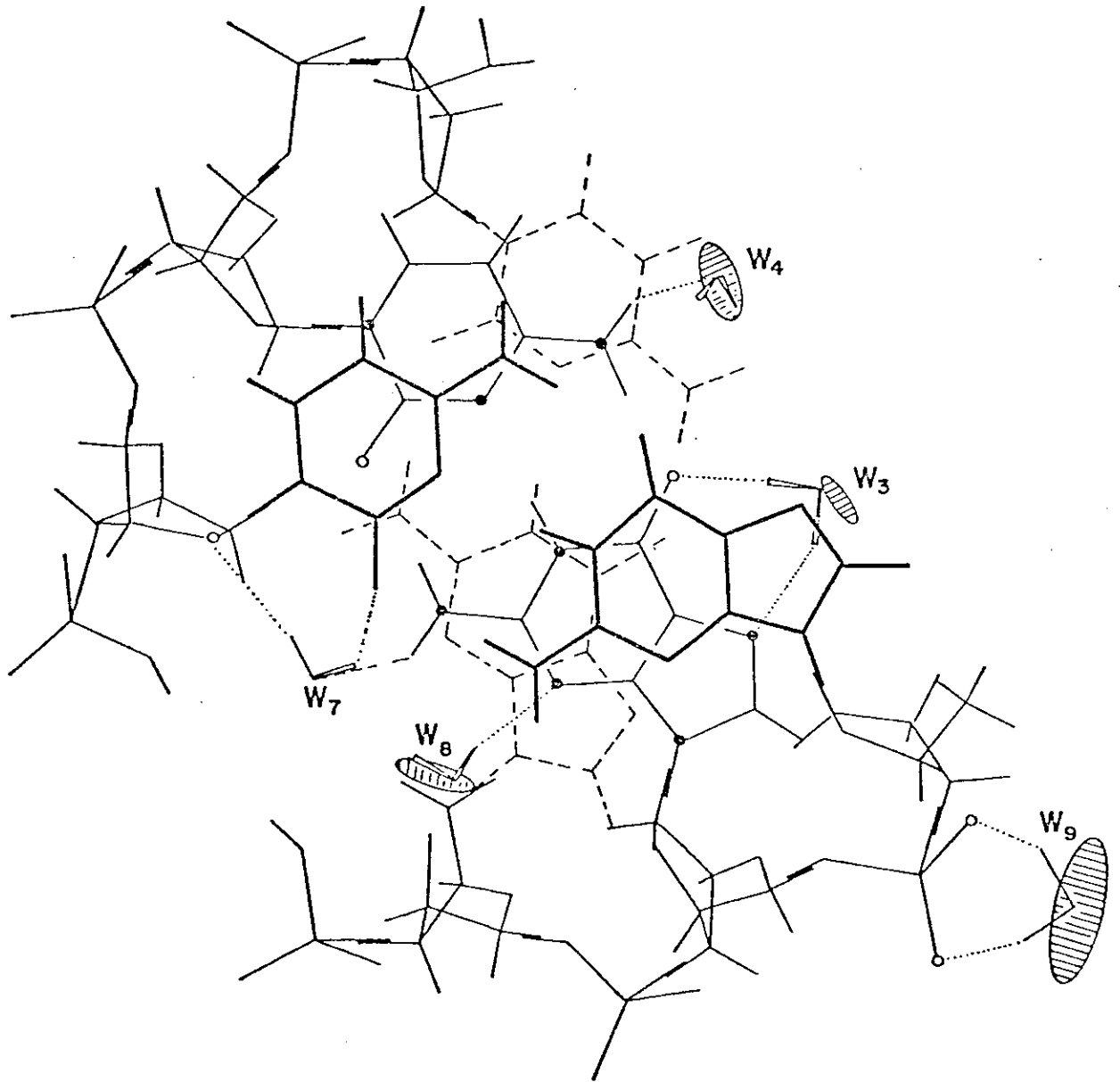


FIG. 6

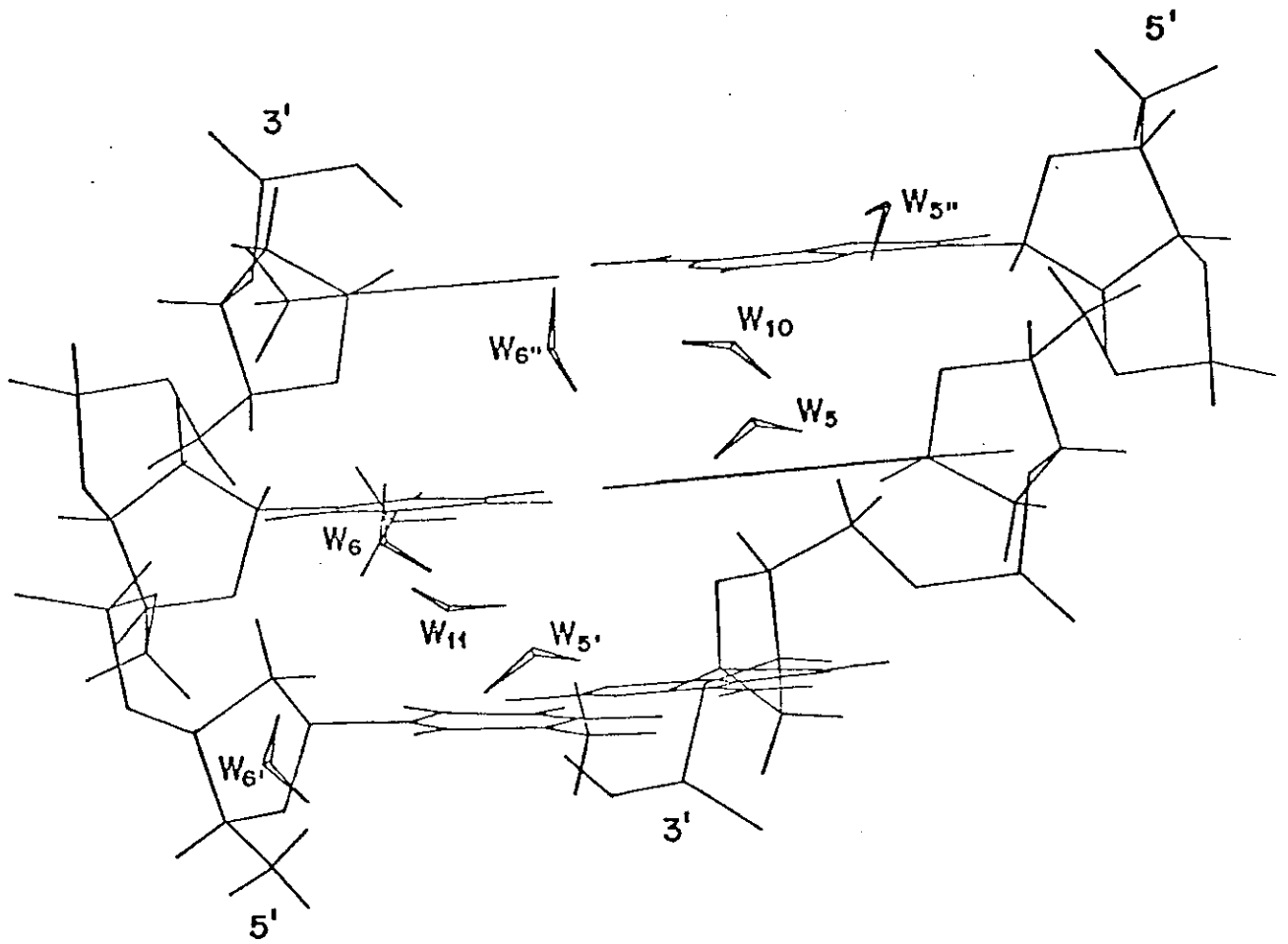


FIG. 7

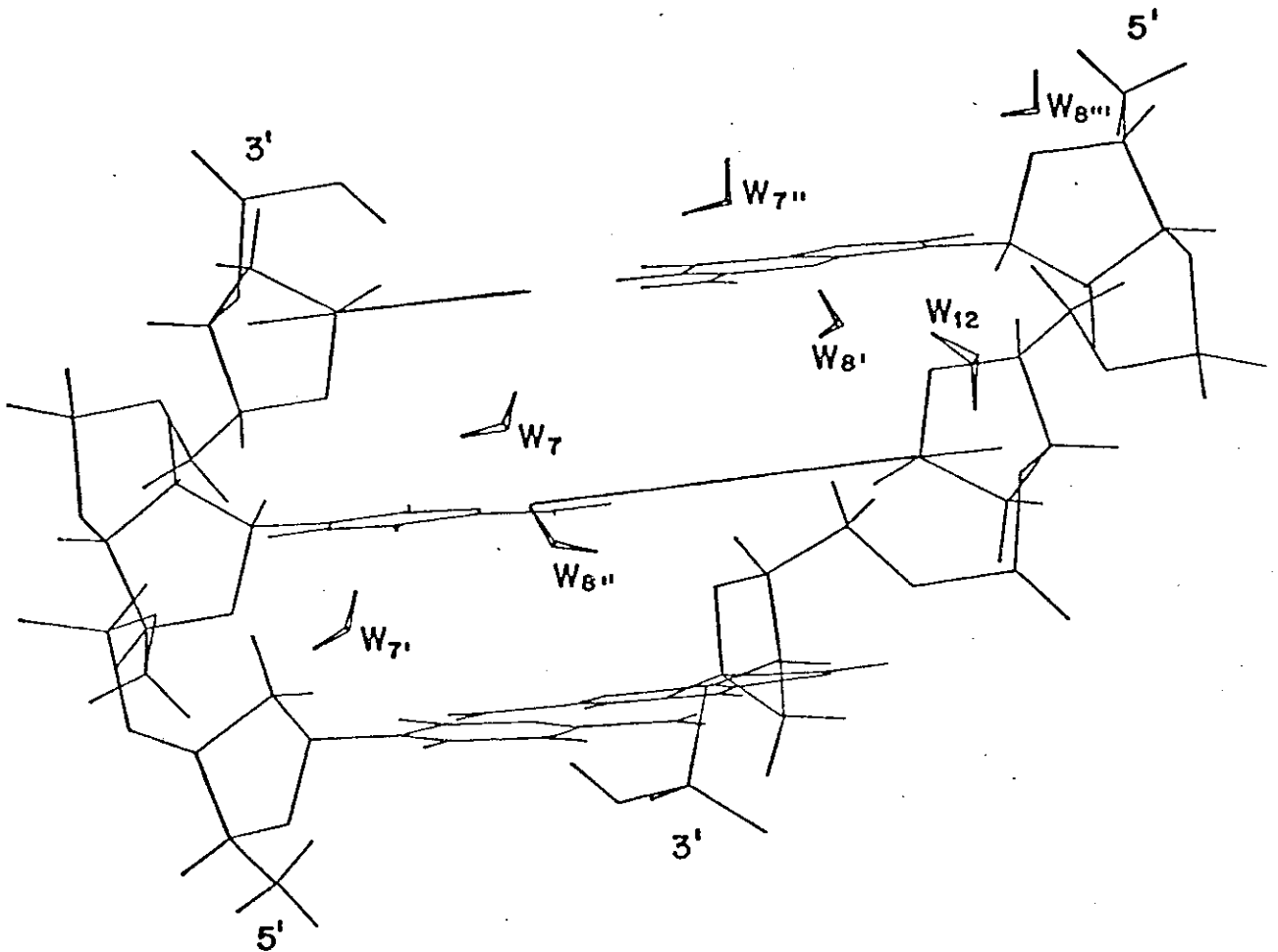
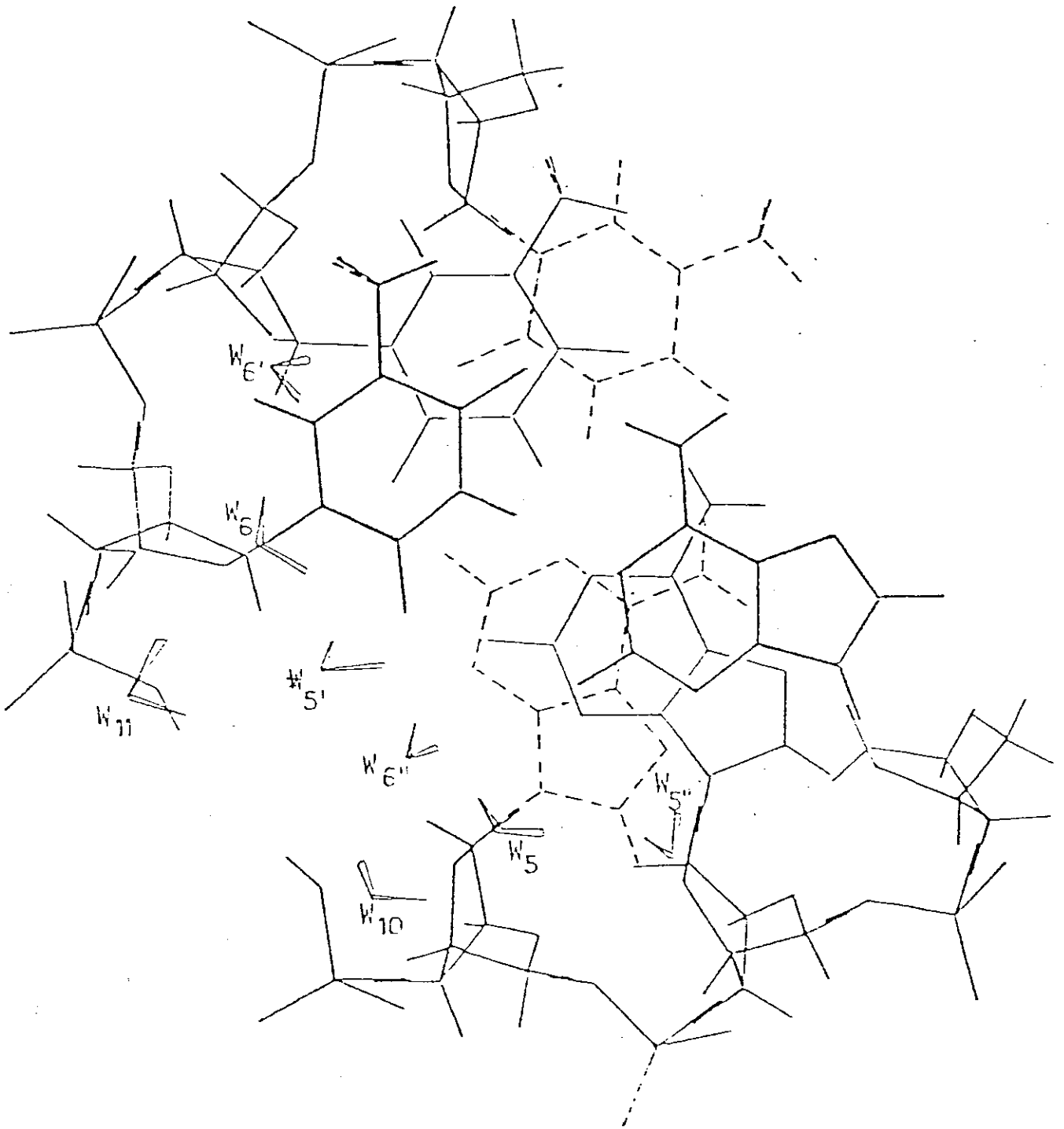


FIG. 8

FIG. 9

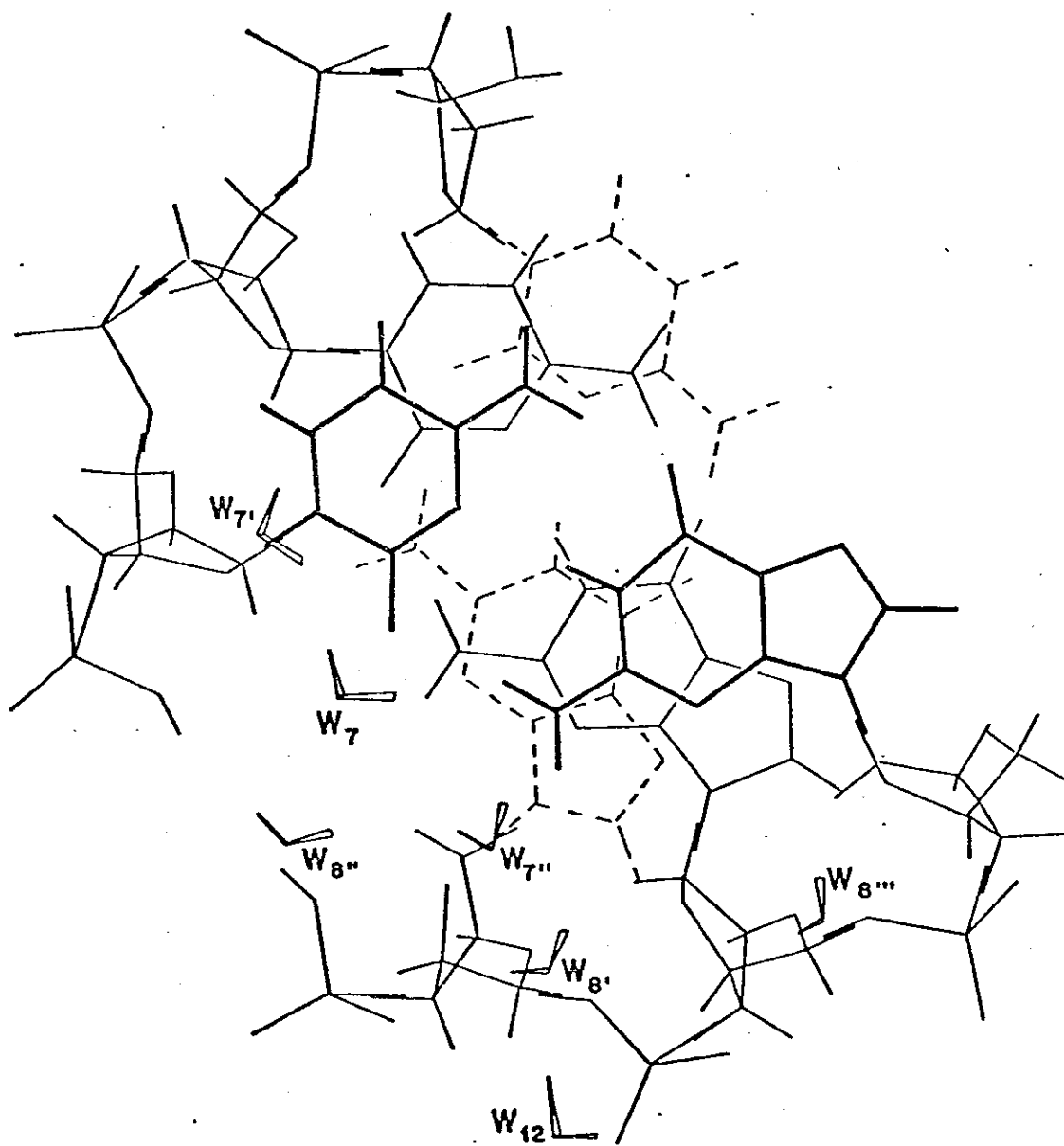


FIG. 10

Table 1

Preferred interaction sites between a water molecule and a model fragment of B-DNA. NC denote the number of a water molecule configurations at each binding site having energies within 1 Kcal/mole of the optimum energy.

Region	Base pair sequence	Receptor atoms involved	ΔE (Kcal/mole)	NC
Major Groove	AT	N_7 (A), H_6 (A)	- 14.6	9
		O_4 (T)	- 8.4	24
	GC	N_7 (G), O_6 (G)	- 15.3	3
		H_4 (C)	- 11.0	96
Minor Groove	AT	N_3 (A), $O_{1'}$ (S)	- 12.3	15
		O_2 (T), $O_{1'}$ (S)	- 11.9	20
	GC	O_2 (C), H_2 (G), $O_{1'}$ (S)	- 11.8	5
		N_3 (G)	- 11.0	4
Phosphate Group		$O_{1'}$, O_2	- 25.5	103
		O_5'	- 20.8	2
		O_3'	—	—

REFERENCES

1. E. Clementi and G. Corongiu, *Biopolymers*, 18, 2431 (1979).
2. E. Clementi and G. Corongiu, *Int. J. Quant. Chem.*, 16, 897 (1979).
3. E. Clementi and G. Corongiu, *Chem. Phys. Lett.*, 60, 175(1979).
4. G. Corongiu and E. Clementi, *Biopolymers*, 20, 551(1981).
5. E. Clementi, G. Corongiu, M. Gratarola, P. Habitz, C. Lupo, P. Otto and D. Vercauteren, *Int. J. Quant. Chem. Symp.*, 16, 409(1982).
6. Submitted to publication in this Journal.
7. G.N.J. Port and A. Pullman, *FEBS Lett.*, 31, 70(1973).
8. A. Pullman and D. Perahia, *Theoret. Chim. Acta*, 48, 29(1978).
9. A. Goldblum, D. Perahia and A. Pullman, *FEBS Lett.*, 91, 213 (1978).
10. R. Lavery, M. de Oliveira and B. Pullman, *J. Comp. Chem.*, 1, 301(1980).
11. R. Lavery, A. Pullman, B. Pullman and M. de Oliveira, *Nucleic Acid Res.*, 8, 5095 (1980).
12. A. Pullman and B. Pullman, *Quart. Rev. Biophysics*, 14, 289 (1981).
13. A. Pullman, B. Pullman and R. Lavery, in "Nucleic Acids: The Vectors of Life", (B. Pullman and J. Jortner, Eds.), p. 75, Reidel Publishing, Dordrecht, Netherlands (1983).
14. R. Wing, H.R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, *Nature* 287, 755 (1980).
15. H.R. Drew, R. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, *Proc. Natl. Acad. Sci. USA*, 78, 2179 (1981).
16. H.R. Drew and R.E. Dickerson, *J. Mol. Biol.*, 151, 535(1981).
17. H.R. Drew, S. Samson and R.E. Dickerson, *Proc. Natl. Acad. Sci. USA*, 79, 4040 (1982).

18. M.L. Kopka, A.V. Fratini, H.R. Drew and R.E. Dickerson, *J. Mol. Biol.*, 163, 129 (1983).
19. B.N. Connor, T. Takano, S. Tanaka, K. Itakura and R.E. Dickerson, *Nature*, 295, 294 (1972).
20. A.H.J. Wang, S. Fukii, J.H. van Boon and A. Rich, *Proc. Natl. Acad. Sci. USA* 79, 3968 (1982).
21. B.N. Conner, C. Yoon, J.L. Dickerson and R.E. Dickerson, *J. Mol. Biol.*, 174, 663 (1984).
22. H.R. Drew, T. Takano, S. Tanaka, K. Itakura and R.E. Dickerson, *Nature*, 286, 567 (1980).
23. S. Arnott, R. Chandrasekaran, R. Birdsall, A.G.W. Leslie, and R.L. Ratliff, *Nature*, 283, 743 (1980).
24. S. Arnott and D.W.L. Hukins, *Biochem. Biophys. Res. Comm.*, 47, 1504 (1972).
25. C. Giessner - Prettre and B. Pullman, *Biochem. Biophys. Res. Comm.*, 107, 1539 (1982).
26. M. Falk, K.A. Hartman and R.C. Lord, *J. Am. Chem. Soc.*, 84, 3843 (1962).
27. M. Falk, K.A. Hartman and R.C. Lord, *J. Am. Chem. Soc.*, 85, 387 (1963).
28. M. Falk; K.A. Hartman and R.C. Lord, *J. Am. Chem. Soc.*, 85, 391 (1963).

BIBLIOGRAFIA

- 1 - E.C. Miller, *Cancer Research*, 38, (1978) 1479
- 2 - J.A. Miller e E.C. Miller, em "Origins of Human Cancer";
Cold Spring Harbor Laboratory, (1977) 605
- 3 - S. Neidle, *Nature*, 283, (1980) 135
- 4 - G. de Murcia, M.C. Lang, A.M. Freund, R. Fuchs, M. Daune,
E. Sage e M. Leng, *Proc. Natl. Acad. Sci. USA*, 76, n^o
12 (1979) 6076
- 5 - E. Scrocco e J. Tomasi, *Topics in Current Chemistry* 42,
(1973) 95
- 6 - E. Scrocco e J. Tomasi, *Adv. Quant. Chemistry*, 11, (1978)
115
- 7 - A. Pullman e B. Pullman, em "Chemical Applications of Atomic
and Molecular Electrostatic Potentials", Eds. P. Politzer
e D.G. Truhlar, Plenum Press, N.Y., (1981) 381
- 8 - A. Pullman e B. Pullman, *Quart. R v. Biophysics* 14, (1981)
289
- 9 - B. Pullman e A. Pullman, em "Carcinogenesis: Fundamental
Aspects and Environmental Effects". Proceedings of the
13th Jerusalem Symposium in Quantum Chemistry and
Biochemistry. Eds. B. Pullman, P.O.P.Ts'o and H. Gelboin,
Reidel Publishing Co. Dordrecht, Holanda, (1981) 55.
- 10 - J. D. Watson e F.H.C. Crick, *Nature*, 171, (1953) 737
- 11 - R.E. Franklin e R.G. Gosling, *Nature*, 171, (1953) 740
- 12 - R.E. Franklin e R.G. Gosling, *Acta Cryst.*, 6 (1953) 673
- 13 - J. Texter, *Prog. Biophys. Molec. Biol.*, 33, (1978) 83

- 14 - R. Langridge, H.R. Wilson, C.W. Hooper, M.H. Wilkins e L. D. Hamilton, *J. Mol. Biol.*, 2, (1960) 19
- 15 - S. Arnott, R. Chandrasekaran, D.W.L. Hukins, R.J.C. Smith e L. Watts, *J. Mol. Biol.*, 88, (1974) 523
- 16 - A.H.J. Wang, G.J. Quigley, F.J. Kolpak, J.L. Crawford, J. H. Van Boom. G. Van der Marel e A. Rich, *Nature*, 282, (1979) 680
- 17 - M. Falk, K.A. Hartman e R.C. Lord, *J. Am. Chem. Soc.*, 84, (1962) 3843
- 18 - M. Falk, K.A. Hartman e R.C. Lord, *J. Am. Chem. Soc.* 85, (1963) 387
- 19 - M. Falk, K.A. Hartman e R.C. Lord, *J. Am. Chem. Soc.*, 85, (1963) 391
- 20 - L.D. Hamilton, *Nature* 218, (1968) 633
- 21 - B. Lee e F.M. Richards, *J. Mol. Biol.*, 55, (1971) 379
- 22 - C.J. Alden e S.H. Kim, em "Stereodynamics of Molecular Systems", Ed. R.H. Sarma, Pergamon Press, N.Y. (1979)
- 23 - E. Clementi e G. Corongiu, em "Biomolecular Stereodynamics, Ed. R.H. Sarma, Adenine Press, N.Y., (1981) 209
- 24 - E. Clementi, G. Corongiu, M. Gratarola, P. Habitz, C. Lupo, O. Otto e D. Vercauteren, *Int. J. Quant. Chem.: Quant. Chem. Symp.*, 16 (1982) 409
- 25 - J.A. Pople e D.L. Beveridge, "Approximate Molecular Orbital Theory", McGraw-Hill Book Company, (1970)
- 26 - V. Fock, *Z. Physik*, 61, (1930) 126
- 27 - C.C.J. Roothan, *Rev. Mod. Phys.*, 23, (1951) 69
- 28 - S.F. Boys, *Proc. Roy. Soc. (London)* A 258, (1960) 402

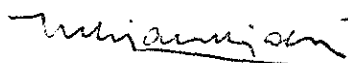
- 29 - E. Clementi, J. M. André, M.C. André, D. Klint e D.Hahn, Acta. Phys. Hung., 27, (1969) 493
- 30 - Ver por exemplo, H.C. Longuet Higgins, J. Chim. Phys., 61, (1964) 13
- 31 - J.N. Murrell, G. Shaw, J. Chem. Phys. 19 (1970) 521; A. Conway, J.N. Murrell, Mol. Phys., 23 (1972) 1143
- 32 - P. Claverie, em "Intermolecular Interactions; from Diatomics to Biopolymers", Ed. Bernard Pullman, Wiley, N.Y., (1978) 69
- 33 - A. Amos e R. Crispin, Mol. Phys. 34, (1976) 669
- 34 - F. Mulder e C. Huiszoon, Mol. Phys. 34, (1977) 1215
- 35 - J. Rabinowitz e R. Rein, Int. J. Quant. Chem., 6, (1972) 669
- 36 - M. Stamatiadou, T. Swisler, J. Rabinowitz e R. Rein, Biopolymers, 11, (1972) 1217
- 37 - R. Rein, Adv. Quant. Chem. 7, (1973) 735
- 38 - A. Pullman e D. Perahia, Theor. Chim. Acta, 48, (1978) 29
- 39 - C.F.J. Böttcher, "Theory of Electric Polarization", Elsevier, N.Y. (1973)
- 40 - Ver por exemplo; J. Langlet, P. Claverie, F. Caron e J.C. Boeue, Int. J. Quant. Chem., 19 (1981) 299
- 41 - A. Goldblum, D. Perahia e A. Pullman, Int. J. Quant. Chem., 15, (1979) 121
- 42 - F.L. Suddath, G.J. Quigley, A. McPherson, D. Sneden, J.J. Kim, S.H. Kim e A. Rich, Nature 248 (1974) 20
- 43 - J.D. Robertus, J.E. Ladner, J.T. Finch, D. Rhodes, R.S. Brown, B.F.C. Clark e A. Klug, Nature, 250 (1974) 546
- 44 - J.L. Sussman, S.R. Holbrook, R.W. Warrant, G.M. Church e S.H. Kim, J. Mol. Biol., 123 (1978) 607

- 45 - D. Perahia, B. Pullman, D. Vasilescu, R. Cornillon e H. Broch, *Biochim. Biophys. Acta*, 478, (1977) 244
- 46 - K. Kim e M. S. John, *Biochim. Biophys. Acta*, 565, (1979) 131
- 47 - B. Pullman, *Int.J. Quantum Chem., Quantum Biol. Symp.*, 4, (1977) 3
- 48 - M.S. de Giambiagi, M. Giambiagi e D.M.S. Esquivel, *Z. Naturforsch.* 38c, (1983) 621
- 49 - A. Pullman e H. Berthod, *Int. J. Quantum Chem., Quantum Biol. Symp.* 4, (1977) 327
- 50 - A. Pullman e H. Berthod, *Theoret. Chim. Acta* 48, (1978) 269
- 51 - A. Pullman, K. Zakrzewska e D. Perahia, *Int. J. Quantum Chem.*, 16 (1979) 295
- 52 - A. Pullman, B. Pullman e R. Lavery, em "Nucleic Acids: The Vectors of Life", Eds. B. Pullman e J. Jortner, Reidel Publishing Co. (1983) 75
- 53 - B.F.C. Clark, *Prog. in Nuc. Acids Res. and Mol. Biol.* 20, (1977) 1
- 54 - R. Lavery, A. Pullman e B. Pullman, *Nucl. Acids Research*, 8, (1980) 1061
- 55 - B. Roos e P. Siegbahn, *Theoret. Chim. Acta*, 17 (1970) 209
- 56 - B. Pullman, N. Gresh, H. Berthod e A. Pullman, *Theoret. Chim. Acta* 44 (1977) 151
- 57 - R. Lavery e A. Pullman, *Int. J. Quantum Chem.*, 20 (1981) 49
- 58 - L. Pauling, em "The Nature of the Chemical Bond", 3^a edição, Cornell press (1960) 260
- 59 - R. Lavery, B. Pullman e K. Zakrzewska, *Biophys. Chem.* 15 (1982) 343
- 60 - M. Litt, *Biochemistry*, 8 (1969) 3249

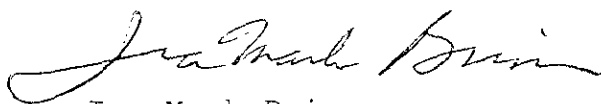
- 61 - M. Litt e C.M. Greenspan, *Biochemistry*, 11 (1972) 1437
- 62 - D. Demoulin, A.M. Armbruster e B. Pullman, *Theoret. Chim. Acta*, 48 (1978) 143
- 63 - D. Rhodes, *J. Mol. Biol.*, 94 (1975) 449
- 64 - L. Massouh-Rizh, These de doctorat d'Etat, Université Louis Pasteur de Strasbourg, France (1975)
- 65 - D.M. Brown, em "Basic Principals in Nucleic Acid Chemistry", vol. 2, Ed. P.O.P Ts'O, Academic Press, N.Y.
- 66 - S. Fujimura, D. Grunberger, G. Caravajal e I.B. Weinstein, *Biochemistry*, 11 (1972) 3629
- 67 - P. Pulkrabek, D. Grunberger e I.B. Weinstein, *Biochemistry* 13 (1974) 2414
- 68 - D. Grunberger e I.B. Weinstein, *Prog. Nucl. Acid. Res. and Mol. Biol.* 23 (1979) 105.
- 69 - M. Staehlin, *Biochim. Biophys. Acta*, 31, (1959) 448
- 70 - M. Litt, *Biochemistry*, 10 (1971) 2223
- 71 - N.E. Broude e E.I. Budowsky, *Biochim. Biophys. Acta*, 254, (1971) 380
- 72 - N.E. Broude e E.I. Budowsky, *Biochim. Biophys. Acta*, 294, (1973) 378
- 73 - A. Szent-Györgyi, *Proc. Acad. Sci. USA*, 74, (1977) 2844
- 74 - A. Szent-Györgyi, em "Search and discovery", Ed. B. Kaminer, New York: Academic Press (1977) 329
- 75 - D. Demoulin, A.M. Armbruster e B. Pullman, *Theoret. Chim. Acta*, 50, (1978) 75
- 76 - M.S. de Giambiagi, M. Giambiagi e C. Alarcon Chamarelli, *Il Nuovo Cimento*, 3D, N.3 (1984) 522

- 77 - L.F. Fieser e M. Fieser, "Organic Chemistry", 2^a edição, Heath. Boston (1950)
- 78 - J. D. Roberts e M.C. Caserio, "Basic Principles of Organic Chemistry", 2^a edição, Benjamin, New York (1977)
- 79 - L. H. Funderburk, L. Aldwin e W.P. Jencks, J. Am. Chem. Soc., 100, (1978) 5444
- 80 - L. Radom, P.C. Hariharan, J.A. Pople e P.v. R. Schleyer, J. Am. Chem. Soc., 95, (1973) 6531
- 81 - E.A. Hill e S.A. Milosevich, Tetrahed. Lett., 50 (1976) 4553
- 82 - A.H. Stroud, em "Digital Computer User's Handbook", Eds. M. Klerer e G.A. Korn, Mc Graw-Hill, N.Y. (1967)
- 83 - R. Wing, H.R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, Nature 287, (1980) 755
- 84 - H.R. Drew, R. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, Proc. Natl. Acad. Sci. USA, 78, (1981) 2179
- 85 - H.R. Drew and R.E. Dickerson, J. Mol. Biol., 151, (1981) 535
- 86 - H.R. Drew, S. Samson and R.E. Dickerson, Proc. Natl. Acad. Sci. USA, 79, (1982) 4040
- 87 - M.L. Kopka, A.V. Fratini, H.R. Drew and R.E. Dickerson, J. Mol. Biol. 163, (1983) 129

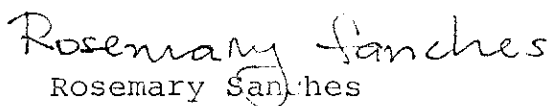
Tese apresentada ao Centro Brasileiro de Pesquisas Físicas do Conselho Nacional de Desenvolvimento Científico e Tecnológico, fazendo parte da Banca Examinadora os seguintes professores:



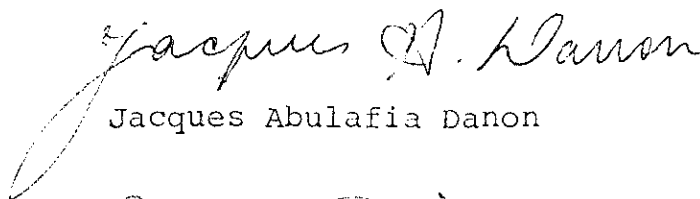
Mario Giambiagi - Presidente



Ira Mark Brinn



Rosemary Sanchez



Jacques Abulafia Danon



Paulo Mascarello Bisch

Rio de Janeiro, 14 de outubro de 1985