Combined Quantum and Molecular Mechanical Study of DNA Cross-Linking by Nitrous Acid

Adrian H. Elcock, Paul D. Lyne, Adrian J. Mulholland, Ankush Nandra, and W. Graham Richards*

Physical Chemistry Laboratory, University of Oxford South Parks Road, Oxford OX1 3QZ, U.K.

Received March 2, 1994
Revised Manuscript Received December 21, 1994

Interstrand cross-linking of DNA is believed to be responsible for the biological activity of a number of antitumor agents, many of which exhibit some degree of sequence selectivity in their interactions with DNA.[1] Despite the experimental interest in cross-linking reactions, their theoretical study has up to now received only limited attention. Molecular mechanics (MM) calculations, for example, have been used to provide some insight into the likely structural effects of cross-linking,[2a-d] but studies of the reaction processes themselves have been prevented by the need for a quantum mechanical (QM) treatment. Recently developed combined QM-MM potential functions[3] allow reactions in large systems to be studied by considering a small QM region influenced by a surrounding MM region. Here such an approach is applied to reactions in DNA and is used to investigate the sequence preferences of DNA cross-linking by nitrous acid.

Nitrous acid cross-links DNA through the exocyclic amino groups of guanine bases.[2b,4a,b] This reaction is believed to proceed via diazotization of guanine, which subsequently undergoes nucleophilic attack (accompanied by expulsion of the leaving group N_2) by a second guanine amino group (Figure 1). It has further been shown that this cross-linking occurs with a pronounced preference for 5′-CG sites over 5′-GC sites,[4a] an effect which, from inspection of standard B-DNA structures, has been ascribed to differences in the proximity of the reactive centers. This explanation essentially equates the distance between the reactive centers of the reactants with the activation energy for their reaction. While this is intuitively appealing, there has up to now been no quantitative evidence that such a relationship exists. Kirchner *et al.*[4a] have pointed out that the cross-linking preference may equally be explained by preferential diazotization of the 5′-CG site, but calculations of the electrostatic potential from solutions to the nonlinear Poisson-Boltzmann equation show little difference between the two sites, suggesting that diazotization should occur to equal extents.[5]

The cross-linking stage of the reaction, shown in Figure 1, was studied for each of the sequences investigated experimentally,[4a] all calculations being performed with the QM-MM potential[3a] implemented in the molecular simulation program CHARMM.[6] Standard B-DNA oligonucleotides of sequence

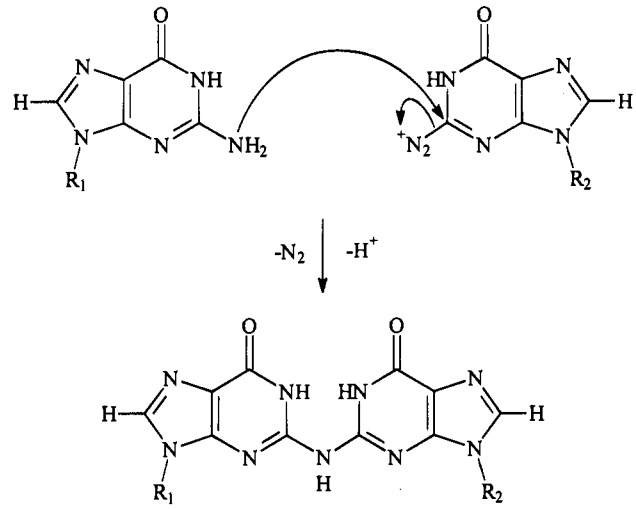

Figure 1. Cross-linking of guanine bases through nucleophilic attack at C2 of a diazotized guanine by the exocyclic amino group of a second guanine.

d(TAATN$_4$ATTA)$_2$ (where N_4 is ACGT, AGCT, CCGG, GGCC, TCGA, and TGCA) were constructed using the molecular modeling package QUANTA[7] with one of the two central guanines being altered to its diazonium form. For the MM region, the all-atom parameters developed for use in CHARMM version 23 were adopted,[6] with one important change being that the total charge on each nucleotide residue was reduced from −1.0 to −0.25 au by scaling the charges of the phosphate group and O3′ and O5′ oxygens. This modification, which approximates the screening of charges by solvent and dissolved ions, is in line with that used in other MM studies.[8] The QM region, which was described by the semiempirical molecular orbital AM1 method,[9] in all cases consisted only of the two guanine bases directly involved in the cross-linking reaction; "link" atoms used to define the boundary of the quantum and molecular mechanical regions[3a] were placed along the glycosidic bond. Two minor additional modifications to the force field were made. Firstly, the charges on the ribose ring attached to the QM bases were set to 0 to prevent unrealistic polarization effects.[3a] Secondly, an additional harmonic restraint (in the form of an improper torsion) was used to maintain a planar geometry at N9 in order to correct for the known weakness of semiempirical methods in the description of such systems.[10]

Prior to reaction path calculations being performed, energy minimization of the modeled structures was achieved by combining 500 steps of steepest descent and 1500 steps of Powell's conjugate gradient methods.[11] An adiabatic mapping procedure[12] was used to calculate approximate energy barriers to reaction. A harmonic potential (k = 10 000 kcal mol^{-1} Å^{-1}) was applied to restrain a suitable coordinate to a series of points along the pathway from reactants to products. At each successive point 200 steps of steepest descent minimization of the whole system were carried out. Initially, the distance between the amino nitrogen of the attacking guanine and the diazonium carbon was varied; in the region of the energy maximum, a better reaction coordinate was found to be given

(1) Hopkins, P. B.; Millard, J. T.; Woo, J.; Weidner, M. F.; Kirchner, J. J.; Sigurdsson, S. Th.; Raucher, S. *Tetrahedron* **1991**, *47*, 2475.
(2) (a) Haworth, I. S.; Lee, C.-S.; Yuki, M.; Gibson, N. W. *Biochemistry* **1993**, *32*, 12857. (b) Kirchner, J. J.; Solomon, M. S.; Hopkins, P. B. In *Structure and Function: Proceedings of the Seventh Conversation in Biomolecular Stereodynamics*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: Albany, NY, 1992. (c) Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Colvin, O. M. *Anti-Cancer Drug Des.* **1989**, *4*, 281. (d) Norman, D.; Live, D.; Sastry, M.; Lipman, R.; Hingerty, B. E.; Tomasz, M.; Broyde, S.; Patel, D. J. *Biochemistry* **1990**, *29*, 2861.
(3) (a) Field, M. J.; Bash, P. A.; Karplus, M. J. *J. Comput. Chem.* **1990**, *11*, 700. (b) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718. (c) Stanton, R. V.; Hartsough, D. S.; Merz, K. M. *J. Phys. Chem.* **1993**, *97*, 11868.
(4) (a) Kirchner, J. J.; Hopkins, P. B. *J. Am. Chem. Soc.* **1991**, *113*, 4681. (b) Kirchner, J. J.; Sigurdsson, S. Th.; Hopkins, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4012.
(5) Elcock, A. H. D. Phil. Thesis, University of Oxford, Oxford, U.K. 1994.
(6) CHARMM version 23 (1993), contact: Prof. M. Karplus, Department of Chemistry, Harvard University, Oxford Sreet, Cambridge, MA.
(7) QUANTA (1993), Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA.
(8) McConnell, K. J.; Nirmala, R.; Young, M. A.; Ravishanker, G.; Beveridge, D. L. *J. Am. Chem. Soc.* **1994**, *116*, 4461.
(9) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(10) Stewart, J. J. P. *MOPAC 93.00 Manual*; Fujitsu Limited: Tokyo, Japan, 1993.
(11) Powell, M. J. D. *Math. Programming* **1977**, *12*, 241.
(12) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, UK, 1987.

0002-7863/95/1517-4706$09.00/0

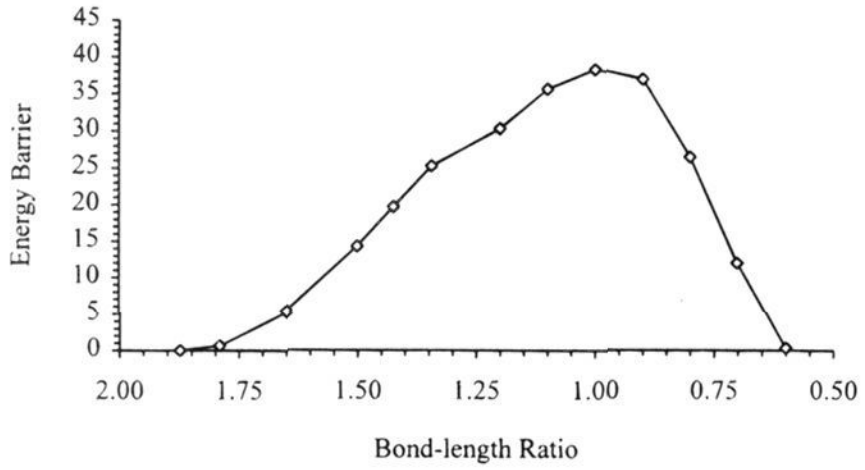

Figure 2. Reaction profile obtained for cross-linking at CCGG sequence as a function of the ratio of forming to breaking bond lengths. Points with bond-length ratio > 1.2 were obtained using a reaction coordinate defined as the forming C−N bond length (see text for details). The energy barrier is given in kcal mol^{-1}.

Table 1

sequence[a]	$E_{barrier}$[b]	C−N distance[c]	sequence	$E_{barrier}$	C−N distance
ACGT	33.6 (31.9)	2.68	AGCT	113.8	4.03
CCGG	45.0 (39.1)	2.72	GGCC	99.2	4.06
TCGA	46.0 (39.1)	2.72	TGCA	117.1	4.50

[a] Refers to the central four bases of the DNA 12-mer. [b] Energy barrier in kcal mol^{-1} relative to the minimized reactants, numbers in parentheses being those obtained from the use of a ratio reaction coordinate (see text for details). [c] Distance between reacting centers in the minimized structure prior to reaction (in Å).

by the ratio of the forming and breaking bondlengths for the particular case of 5′-CG sites. A typical reaction profile is shown in Figure 2; the computed energy barriers for all sequences are given in Table 1.

The calculated energy barriers for the six duplexes fall into two categories distinguished by the identity of the central site. In all cases, the energy barrier for cross-linking at 5′-GC sites is much higher than for the 5′-CG sites, in agreement with the experimental results. The barriers for all reactions are high, in accord with the low yields obtained experimentally, but are almost certainly overestimated because only partial relaxation of the DNA structure is allowed by the rather limited nature of the minimizations. Use of more extensive minimizations or simulated annealing procedures would almost certainly reduce the barriers, but is highly unlikely to alter qualitatively the relative energies associated with 5′-CG and 5′-GC sites. The

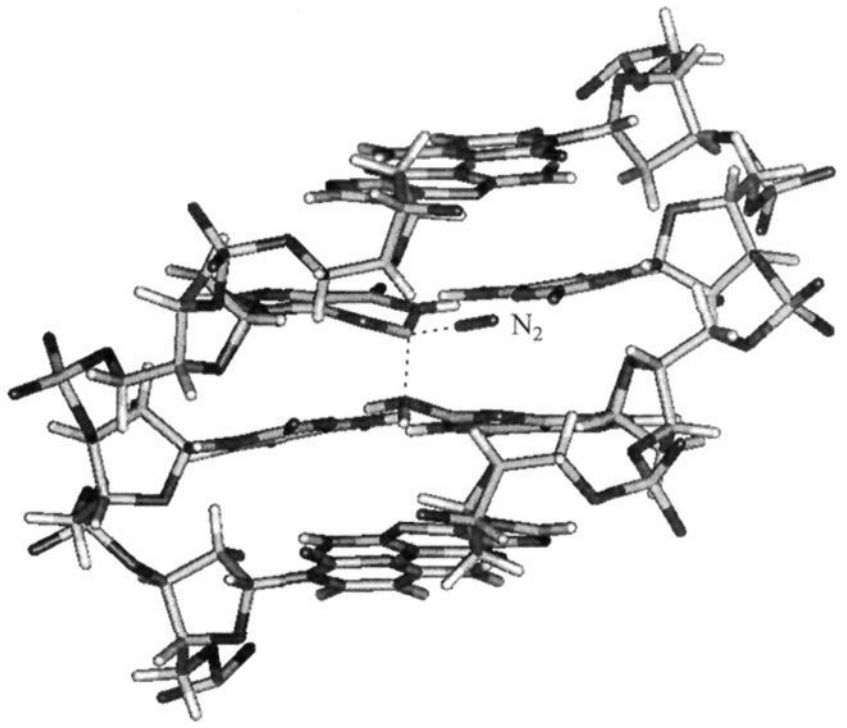

Figure 3. Structure of highest energy along the reaction coordinate for cross-linking of the CCGG sequence. Forming and breaking bonds are indicated by dotted lines.

"transition state", or more properly, the structure of highest energy along the reaction coordinate, for cross-linking with a central CCGG sequence is shown in Figure 3. The forming and breaking bond lengths are almost equal, at 1.70 Å, with the diazonium carbon adopting an approximately tetrahedral geometry, but displaced below the plane of the purine ring system. This latter effect is perhaps an artifact of the limited minimization procedure, but the structure is otherwise realistic. It is worth noting that the energy barriers obtained are, as suggested by Hopkins and co-workers,[4a,b] strongly dependent on the C−N distance in the minimized structures (Table 1).

A more complete description of the cross-linking process would require dynamic effects to be taken into account. Nonetheless, the results reported here do provide a clear explanation for the sequence preferences of DNA cross-linking induced by nitrous acid; in particular, they support the proposal of Kirchner *et al.*[4a,b] that a higher activation energy results from the greater structural change required in cross-linking at 5′-GC sites. As a more general point, the present results illustrate the potential utility of methods combining quantum and molecular mechanical treatments for the study of reactions involving DNA.

JA9406703