The origin of enantioselectivity for intramolecular Friedel-Crafts reaction catalyzed by supramolecular Cu / DNA catalyst complex

Galina P. Petrova,[a] Zhuofeng Ke,[a] Soyoung Park,[b] Hiroshi Sugiyama,[b] and Keiji Morokuma[a,*]

[a] Fukui Institute for Fundamental Chemistry, Kyoto University, Kyoto 606-8103, Japan

[b] Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

Abstract

The present theoretical investigation aims at understanding the origin of enantioselectivity of intramolecular Friedel-Crafts reaction, catalyzed by supramolecular Cu / DNA catalyst. 28 conformations of the supramolecular L-Cu(II)-R / $d(CAAAAATTTTTG)_2$ complex were thoroughly modeled to estimate their stability and structural features depending on the metal complex conformation and its intercalation position. The preferred formation of *S*-product can be rationalized by the higher binding energy of pro-*S* conformations to DNA. Pro-*S* conformations are structurally closer to the expected C3-C2' bond formation TS and usually not deeply buried into DNA, which would facilitate TS formation by decreasing the energy for conformational changes.

Keyword:

supramolecular Cu / DNA catalyst, asymmetric unimolecular Friedel-Crafts reaction, MD and QM/MM model, structure and stability of Cu / DNA catalyst

1. Introduction

The idea of harnessing natural chiral biomolecules, instead of synthetic chiral organic ligands, in asymmetric organic synthesis is intriguing and proved to be quite versatile. Recently such novel "metal-complex / DNA" based catalysts have been developed and their catalytic behaviors and properties have attracted wide interest due to their high enantio- or stereoselectivity in many chemical reactions [1-6]. These catalysts combine the catalytic power of transition metal complexes and the chiral architecture of biopolymers resulting in both efficiency and high enantioselectivity for the catalysis. Another important advantage is that these systems are water-soluble and do not require the usage of expensive and environment unfriendly organic solvents. The DNA component is cheap and readily available, and desired oligo- and polymers with designed nucleotide sequence and length could be prepared relatively easily. The potential of such hybrid supramolecular catalysts was first demonstrated by Roelfes & Feringa in the case of Diels-Alder reactions [7], resulting in extremely high enantioselectivity as the catalyst was simply formed *in situ* from a double stranded DNA oligomer and a Cu(II)-containing complex with achiral binding ligand. The DNA was found not only to provide a chiral micro-environment for the reaction but to also increase the reactivity. The range of chemical reactions successfully performed in the presence of "metal-complex / DNA" catalyst includes: Diels–Alder reactions [7-9], Michael addition [10], Friedel-Crafts alkylation [11,12], fluorination [13]. Although the performance of the Cu / DNA catalyst has been investigated in detail, the intimate structure of the catalyst and the exact reasons for the high enantioselectivity still remain unclear and, thus, a challenging subject for a theoretical investigation. In a very recent experimental study, Park et al. [12] reported the successful application of Cu / DNA based catalysts in a unimolecular Friedel-Crafts reaction (FC) by using a single reactant molecule with 2-acyl imidazole part, coordinating to Cu(II), and indole moiety (Scheme 1). This system provoked our interest due to the narrower range of possible initial configurations of the system, and the availability of some experimental information on the structure of similar metal complexes and their interaction with DNA [14,15]. Ternary Cu(II) amino-acid complexes were studied previously as DNA cleavage agents by combining experiment and docking simulation [14]. Crystallographic data [14,15] show that in L-tryptophan Cu(II) complexes the indole moiety is located over the base of the square pyramid defined by the coordination of the organic ligands to the metal cation and the coordinated water / counter anion resulting in pseudo octahedral coordination of the cupric ion. The formation of the structure is easily explained by intramolecular π-π stacking interactions and electrostatic attraction between the Cu(II) and the π-orbitals of the indole moiety. L-tryptophan Cu(II) complexes show good DNA binding propensity and intercalate in the DNA strand with the L-tryptophan ligand oriented on the side of

the minor groove. During the process the tryptophan ligand may loose the intramolecular stacking and unfold in order to form strong intermolecular H-bond between the indole-NH and the DNA-phosphate groups [14].

In this paper, we present a combined MD and QM/MM theoretical study on the binding of L-Cu(II)-R complex to d(CAAAAATTTTTG)$_2$ dodecamer. The L-Cu(II)-R complex consists of bipyridine (L) and the reactant (R) that undergoes the unimolecular FC reaction (Scheme 1, Fig. 1) [12]. The main goal of the study is to find the preferred conformation of the L-Cu(II)-R / DNA reactant complex and to explain the influence of DNA micro-environment on the enantioselectivity. For this purpose we modeled 28 structures of the Cu / DNA catalyst with different initial conformation of L-Cu(II)-R and different intercalation position considering all symmetrically unique base pairs layers (BP layers). In the comparison of different Cu / DNA supramolecular complexes, several important features were taken into account: (i) energetic characteristics of the system as binding energy and energy of deformation; (ii) deformation of the DNA strand; and (iii) deformation of the metal complex.

2. Theoretical approach

The model system consists of L-Cu(II)-R complex (Fig. 1a) and a double stranded DNA dodecamer. In the experimental studies of the modeled Cu / DNA system, it was proposed that the metal complex intercalates between the DNA BP layers via the binding ligand, 5,6-dimethyl-1,10-phenanthroline [12]. Considering this, we focused the theoretical modeling on intercalated species but using a simpler bipyridine ligand L. It is, however, possible that 5,6-dmp exhibits other type of interaction with the DNA strand mainly hydrophobic interactions (not considered here) as commented in some experimental studies [16]. Another reason for choosing bpy as ligand was that bpy-based ligands were found to give good results in the case of bimolecular reactions studied by Feringa and coauthors [1-3,7,10,11]. The reactant used in the experimental study was (2E)-6-(1H-indol-3-yl)-1-(1-methyl-1H-imidazol-2-yl)hex-2-en-1-one. In the theoretical investigation we simplified the reactant, R, by replacing the methyl substituent at the imidazole N1' atom with a hydrogen atom. The DNA was initially created as ideal B-type AT-rich dodecamer double strand, d(CAAAAATTTTTG)$_2$, with 6 unique positions for intercalation of L-Cu(II)-R and the gaps between the DNA BP layers for intercalation were manually created gaps. In the experiment this DNA sequence resulted in relatively high enantioselectivity (60 % *ee*).

Four possible conformations of bare L-Cu(II)-R complex were considered, two pro-*S* (denoted as *S*1 and *S*2) and two pro-*R* (denoted as *R*1 and *R*2), as shown on Fig. 1b. According to previous docking results [14], R was located in the minor groove of the DNA; such orientation

would provide for additional interaction between the indole and the phosphate groups and better micro-environment for enhanced enantioselectivity. There are six symmetrically unique gaps of DNA BP layer, denoted by *nm* later, i.e. 12 between C1 and A2, 23 between A2 and A3,.. , and 67 for A6 and T7, where L-Cu(II)-R conformation can intercalate. Thus, a total 24 initial L-Cu(II)-R / DNA supramolecular complexes, DNA_*nm*_*R*1, *R*2, *S*1 and *S*2, were manually constructed via intercalating each L-Cu(II)-R conformation into every gap of DNA BP layers. Additional 4 initial structures, DNA_67maj_*R*1, *R*2, *S*1 and *S*2, were constructed by intercalating L-Cu(II)-R between A6 and A7 but with R located on the side of the major groove. The open space on the side of the major groove of such a small DNA fragment is not likely to provide large enough asymmetric micro-environment. However, recent experimental studies have shown that copper complexes with itrastrand ligands may prefer major groove location and also exhibit high enantioselectivity [17].

Each of the 24 supramolecular complex structures was then neutralized by adding 20 Na^+ counter cations applying Åqvist model [18] (the total negative charge of the DNA dodecamer is 22 e; however, the presence of the positively charged Cu(II) complex decreases this charge to -20 e.). The systems were further solvated into 10 Å truncated octahedral shell of ~ 6000 TIP3P H_2O's. During the MD simulations with boundary conditions, the long-range electrostatic interactions were treated with the particle mesh Ewald (PME) [19] method using a real space cutoff distance 10 Å. SHAKE algorithm [20] was applied to constrain bonds involving hydrogen. Each initial structure was then relaxed via a trajectory of 1 ns MD simulation (detailed explanation is provided in Section SII, where S indicates Supporting Information). The MD simulation time of 1 ns is very short compared to recent MD standards of 10 ~ 1000 ns. However, in the current case the MD simulations were performed mainly for relaxing the generated conformations of the supramolecular complexes, assuming that the metal complex is intercalated between DNA BP layers at different positions on the minor groove side of the double strand.

All MD simulations were performed with the sander module of Amber 9.0 [21,22]. The DNA structure was described with ff99SB force field (FF) without modifications [23], the organic components L and R were described by a general AMBER FF (GAFF). The Cu(II)-related FF parameters were generated via three-point approach [24] or literature data [25,26] (for more information see Section SI). The binding energy, BE^{MD}, was estimated by applying MM-PBSA algorithm [27]. The DNA structure in Cu / DNA complexes was analyzed by using Curves+ web server [28,29].

Next, we optimized the local structure of L-Cu-R intercalated into the middle of four layers of DNA base pairs at the QM/MM level using ONIOM(M06-2x/6-31G*:AMBER) with mechanical embedding [30] in Gaussian 09 program [31] starting from *the most stable* L-Cu(II)-

R / DNA structure within each MD trajectory. The quantum-mechanical (QM) part includes only L-Cu(II)-R (59 atoms, q = +2e, doublet). The DNA (764 atoms, q = -22e), the counter cations (20×Na^+) and the water shell (717×H_2O within *ca*. 5-6 Å around Cu / DNA) were modeled at MM (Amber FF) level. The M06-2x density functional [32] was used in combination with Stuttgart / Dresden effective core potential basis set for Cu [33] and 6-31G* basis set for the rest of the system. The optimization of bare L-Cu(II)-R and the estimation of the deformation energy were performed without or with taking into account the water solvent at PCM level [34,35]. Frequency calculations were performed to confirm local minima.

Although the actual size of the systems was kept the same for all QM/MM simulations, the direct comparison of the stability among different structures is not possible, as the conformation of the water + counter cation shell differs in every systems. Therefore, we made an estimation of the Cu / DNA interaction energy, $Eint^{QM/MM}$, between the pure DNA structure (excluding the water + Na^+ shell) and the copper complex, as well as the deformation energies of the copper complex and DNA, using the following formulas.

DNA^{def}@MM is the MM energy required to deform the DNA structure (without water or counter cations) from the most stable gapped DNA structure, DNA_56_*R*1 complex, to that in the ONIOM-optimized supramolecular complex:

$$DNA^{def}@MM = Esp[DNA@MM] - Esp[DNA_{DNA\_56\_R1}@MM] \quad (1)$$

Cu^{def}@QM is the QM deformation energy of the metal complex from its most stable (optimized) structure, L-Cu(II)-R(*S*1), to that in the ONIOM-optimized supramolecular complex. When the PCM solvent is included, the notation Cuw is used.

$$Cu^{def}@QM = Esp[Cu@QM] - Eopt[Cu@QM] \quad (2)$$

$Eint^{QM/MM}$ is the QM/MM interaction energy between DNA (i.e. without water and counter cations) and the copper complex in the ONIOM-optimized supramolecular structure, evaluated by the following equation:

$$Eint^{QM/MM} = Esp[Cu/DNA@ONIOM] - Esp[Cu@QM] - Esp[DNA@MM] \quad (3)$$

Although the sum

$$\Delta E' = Eint^{QM/MM} + DNA^{def} + Cu^{def}$$

has some of the reference energies cancelled, it is *not* the energy that can be used to compare directly all the supramolecular complexes in Fig. 2. However, because the difference in energy between different conformations of the L-Cu(II)-R complexes is small, as will be shown in Table 1, individual energy components may be used to compare the relative stability of different conformations of the supramolecular complexes.

3. Results and discussion

3.1 Reaction mechanism and conformations of the bare L-Cu(II)-R complex

Schematic representation of the metal complex, numbering of the atoms as applied in the following discussion and the reaction scheme are shown on Fig. 1a, while the conformations used for construction of Cu / DNA complexes for MD and QM/MM calculations are presented on Fig. 1b. The proposed reaction mechanism consists of two stages: C2'-C3 bond formation (stage I) and H transfer from C2' to C2 (stage II). The enantioselectivity should be defined at stage I by the site of attack of C2' to C3. The second stage, H transfer, is not expected to change the configuration of the formed C3 asymmetric center. The site of attack of C2' to C3 depends on the initial indole position with respect to the C2=C3 double bond, and, as mentioned in Section 2, four initial conformations of L-Cu(II)-R were considered: two pro-*S* structures, L-Cu(II)-R(*S*1) and L-Cu(II)-R(*S*2), and the corresponding pro-*R* structures, L-Cu(II)-R(*R*1) and L-Cu(II)-R(*R*2). *S* and *R* are defined on the position of the indole moiety with respect to the plane defined by Cu(II), L and the imidazole part of R, and C2'-C3 interaction would result in formation of *S* and *R* product, respectively. The difference between *S*1 and *S*2 vs *R*1 and *R*2 originates from rotation around the single C6-C3' bond and is reflected only on the stereochemistry of the C2' atom (*S* and *R*, respectively) in the intermediate formed after the C3-C2' bond formation. The most stable structure both in gas phase and in solvent is L-Cu(II)-R(*S*1). The *S*1 and *R*1 types of conformers are slightly more stable than the corresponding *S*2 and *R*2 structures by 1.2 (0.6) and 1.9 (1.7) kcal.mol^{-1} (values in brackets correspond to energies after full geometry optimization in implicit solvent), respectively (Table 1). All conformations of L-Cu(II)-R are very close in stability, and no preference of pro-*S* or pro-*R* conformation is observed in the absence of DNA. This is expected for free complex, suggesting that the free complex cannot provide the reaction enantioselectivity. In the absence of asymmetric environment none of the initial conformations prevails and the transition state structures, formed from these conformation, will have similar stability too. The distances between the plane defined by the indole moiety and the Cu(II) center and the imidazole part of the reactant molecule show interaction with the Cu(II) center (the indole part takes the role of an axial ligand located at distance ca. 2.5 ~ 2.7 Å) and π-π stacking between the indole and imidazole parts of the reactant molecule.

3.2 Results from MD and QM/MM simulations of Cu / DNA supramolecular complexes

As mentioned above, as shown on Fig. 2, we performed a systematic study of L-Cu(II)-R intercalation in all symmetrically unique BP gaps, assuming orientation of R to be on the side of the DNA minor groove. Our expectation was that the indole moiety of R might show preferences

to interaction (via H-bond formation and distant electrostatic interactions) with the DNA phosphate units as well as with the amino groups and carbonyl oxygen atoms of DNA BP, if such an interaction provides sufficient energy to override stacking within the metal complex.

Two RMSD values, the $RMSD^{MIN}$ of the most stable MD structure as well as the time-averaged $RMSD^{AV}$, both with reference to the time-averaged structure are provided in Table 2. The structures of both DNA and L-Cu(II)-R are stable with time as the average RMSD are within the ranges 1.1 to 1.7 Å for DNA and 0.4 to 2.0 Å for L-Cu(II)-R. The upper limit for the copper complex corresponds to a structure with L-Cu(II)-R located close to the edge of the DNA, DNA_12_*R*1. On the base of the $RMSD^{MIN}$ values the conclusion could be made that the most stable structures differ significantly from the time average ones in most of the cases both with respect to DNA and the complex itself. Quite high deviations are observed for the edge structures with L-Cu(II)-R intercalated between the first and the second BP layers. The RMSD as a function of time in the different trajectories are presented on Figs. S1-2. These results show that the trajectories are stable despite the relatively short time of the MD simulations.

3.3. Binding energy of L-Cu(II)-R to AT-rich DNA dodecamer

The dependence of the BE^{MD} and $Eint^{QM/MM}$ values on the intercalation position of the complex are presented on Fig. 3 as well as in Table 2. According to BE^{MD} values, evaluated by MM-MD in the presence of water shell and metal countercations, metal complexes with pro-*S* conformation bind to DNA stronger than the corresponding pro-*R* structures, and in some cases the energy differences exceed 10 $kcal.mol^{-1}$. On the other hand, $Eint^{QM/MM}$ energies, evaluated by Eq. (3), represent the direct interaction energies between DNA (excluding solvation shell consisting of water molecules and cations) and the copper complex at the ONIOM optimized geometries and are overestimated because of the neglect of ionic interactions. $Eint^{QM/MM}$, nevertheless, reproduces well the stability trends observed in BE^{MD}, suggesting that the results from MD and QM/MM calculations can be used as complementary sets of data in predicting the behavior of the supramolecular catalytic system. Both BE^{MD} and $Eint^{QM/MM}$ (with one exception for *nm*=67 case) indicate that the supramolecular complexes formed from pro-*S* structures show higher binding energy to DNA than the pro-*R* structures and, taking into account that there was very little energy difference between pro-*R* and pro-*S* structure for copper complexes only, one can conclude that the formation of pro-*S* Cu / DNA supramolecular reactant complexes is preferred to the formation of the pro-*R* reactant complexes. The present reaction is expected to take place in two stages (Fig. 1a) and the enantio-defining is the first stage, the C-C bond formation stage. Although we have not determined transition state (TS) structures and their stability for this first stage, one can suggest that the relative stability of the TS and the

enantioselectivity is determined by the relative stability of the reactant complex, and therefore the *S*-selectivity is expected for this reaction, which is consistent with the experimental findings [12].

The binding energies are also found to depend on the intercalation position, as seen in Fig. 3. Both in BE^{MD} and $Eint^{QM/MM}$, intercalation at the gap 45 is the most preferred. Structures DNA_45_*S*2, DNA_56_*S*2 and DNA_67_*S*2 are among those with the highest BE^{MD} values and also among the highest $Eint^{QM/MM}$ values. Among the DNA_67maj structures with major groove orientation of R, only one shows a significant binding energy: DNA_67maj_*S*2 with BE^{MD} = -64.3 kcal.mol^{-1} which is the strongest among all modeled complexes.

A similar trend of the intercalation dependence is also seen in the MM deformation energy of the DNA strand, DNA^{def}@MM, Eq. (1), relative to the most stable conformation in the whole series, DNA_56_*R*1. The deformation energy of the DNA strand increases by moving the gap from the central part of the dodecamer to its edges (see also Fig. S7); as the whole set of data could be roughly fit to linear equation: DNA^{def}@MM = -63 × *n* + 511 (R^2 = 0.78) kcal.mol^{-1} with *n* corresponding to the intercalation position, as we defined *nm*, starting from the edge of the strand. The most obvious reason for the higher stability of the central gap is that the DNA structure could distribute the strain resulting from gap opening by deforming the neighboring 3-4 BP layers and the central gap has most neighbors. No clear trend is found for the dependence of DNA deformation among *S*1, *S*2, *R*1 and *R*2 conformations of L-Cu(II)-R.

The QM deformation energy of the metal L-Cu(II)-R complex in the supramolecular complexes, as shown in Eq. (2), is calculated relative to the most stable bare L-Cu(II)-R conformation, *S*1. The obtained values without and with solvent effect are provided in Table 2 (see also Fig. S7). The deformation energy is independent of the intercalation position. The deformation energy is low in cases where the intramolecular π-π stacking between indole and the Cu ligands is more or less preserved: 7.7 (11.2) ~ 19.9 (16.9) kcal.mol^{-1} (the values in brackets correspond to solvent calculations). In cases where this intramolecular interaction is destroyed to facilitate interaction between the indole and the DNA minor groove, Cu^{def}@QM (Cuw^{def}@QM) is large, reaching 51.9 (59.0) kcal.mol^{-1} for DNA_45_*R*2. The solvent effect is more pronounced in the latter cases (loss of stacking), such as DNA_12_*R*1 and DNA_34_*R*2, as the highly polar water solvent destabilizes the structures additionally with up to 48 kcal.mol^{-1}.

3.4. Structural response of the DNA strand and the metal complex to the formation of the supramolecular complex

The detailed analysis of the DNA structural characteristics was inspired, on one hand, by the idea of gaining understanding of the influence of the metal coordination position on the local

and helix structure of the DNA double strand, and on the other hand, to check whether the formation of the supramolecular complex may result in partial change of the DNA conformation from B type to A type, as discussed in some experimental studies based on CD spectral data [1]. The analysis we performed was based on the standard base pair parameters, X displacement and inclination (for the definition of these parameters see Section SIII) [36]. The helix morphology of the whole DNA duplex can be described in terms of two essential characteristics: the average axis bending of the DNA strand and the minor groove width, MGW. The average values of the parameters over the whole DNA strand as functions of the intercalation position are shown on Fig. 4 (The values for each BP layer are shown on Figs. S3-6). Due to the relatively low simulation time, fraying of the DNA structure was not observed during the MD simulation (see SI Fig. S7 and comment therein). Additional structural analysis of the most stable structures of the supramolecular Cu / DNA complexes shows for some of the intercalation modes slightly increased values of the stretch and opening parameters from the expected values but still they are more incidental and can hardly be considered as symptoms for fraying of the DNA double strand.

As a common trend, the distortion of the DNA strand from a standard B type DNA is more pronounced for intercalation structures at more central positions. The average characteristics of the four DNA_12 conformations are very close to B DNA. Shifting the intercalation position towards the center of the DNA strand is accompanied by increasing the absolute values of the X displacement and the inclination of the base pairs, coming closer to the values typical for A DNA. Inspection of the individual values for each base pair shows that X displacement deviations usually affect the pairs that are closest to the intercalation position. Intercalation positions *nm*=45, 56 and 67 were found to result in more pronounced axis bending most probably as a result of the DNA attempting to stabilize the gap between the layers and to accommodate the metal complex by electrostatic interaction and H-bonding. The maximal values of MGW are also closer to A DNA values for most of the conformations, with only DNA_12 close to B DNA. Another structural parameter defining the DNA type is the pucker distribution of the deoxyribose part of the sugar-phosphate backbone. Interestingly, the calculated phase angles (Fig. S7) fall predominantly in the range 120-180 and correspond to C1'-exo and C2'-endo conformation of the deoxyribose observed in B DNA. For species with metal complex intercalated between 45 and 56 BP layers (DNA_45 and DNA_56) there is also contribution of C4'-exo and C4'-endo conformations. Contribution of C3'-endo typical for A DNA is observed in DNA_67.

As can be expected, the intercalation of the metal complex results in increased rise for the BP layers between which the metal complex is located. In most of the supramolecular complexes for a given intercalation position, the formed gap is slightly narrower for the pro-*S* species than

for the pro-*R* species, which could be a result from the rather shallow intercalation of the pro-*S* structures (see Section SIV). A narrower BP layer distance may mean better π-π stacking between L and DNA BP, which supports the higher stability of the pro-*S* complexes.

Concerning the analysis of the L-Cu(II)-R deformation a special attention was paid to the C3-C2' distance in the complexes at the different intercalation positions of the metal complex after QM/MM optimization, as it would show whether the complex is close to TS for C3-C2' bond formation or further conformational changes are required in order to reach a state good for C3-C2' bond formation. (Table 2 as well as the right bottom panel of Fig. 4. More detail description of the metal complexes local structure is provided in Section SIV.) The C3-C2' distances, as well as the deformation energy of the metal complexes (as discussed in Section 3.3), are not very sensitive to the intercalation position. However, the pro-*S* conformations in the supramolecular complex in general exhibit shorter C3-C2' distances than the pro-*R* structures at the same intercalation positions for the whole series of data. The QM/MM optimization resulted in shortening of the distance in most of the structures with 0.055 ~ 1.829 Å in comparison with the values in the most stable MD structures, especially in structures in which the stacking conformation of the complex is preserved, suggesting that strong attractive interaction is taking place. In two of the structures, DNA_34_*S*2 and DNA_45_*S*2, a spontaneous formation of C3-C2' bond occurs during the QM/MM optimization procedure. Considering that in the pro-*S*2 conformation of the metal complex, the H atom at C2' and the C2 atom, to which proton should be transferred, are in *s-cis* position with respect to the newly formed C3-C2' bond, the proton transfer may proceed even in the absence of proton mediator. The detailed analysis of the local structure of the supramolecular complexes (see Section SIV) shows also that the pro-*S* complexes are usually only shallowly intercalated into the DNA strand, and the formation of TS and the release of the obtained product would not require additional energy to drag the reactant complex out of the gap in order to provide space for the chemical process to occur. Such a statement may sound inconsistent with the higher stability of the reactant pro-*S* structures. However, how deeply L-Cu(II)-R would be intercalated into DNA is not a sole result of π-π stacking stabilization, but also depends also on the position of R. Whether it would prefer keeping the stacking between the indole and imidazole (meaning that L-Cu(II) cannot intercalate too deep) or would prefer opening the structure so the indole moiety may be located in the minor groove or find other ways to interact with the DNA (in which case the complex may be deeper intercalated), etc. On the other hand, in discussing possible reaction mechanism, we have to keep in mind that the reactant complex still have to be in proper conformation for the reaction to occur, i.e. C2'-C3 should preferably be within interacting distances. In this case no additional energy has to be inserted in order to find a proper initial conformation of L-Cu(II)-R / DNA.

Another interesting question of course is whether the discussed trend in the C3-C2' distances with respect to the conformation of the complex is observed also in the time average MD structures. In general indeed the pro-*S* structures show shorter C3-C2' distances in the average MD structures by 0.272 ~ 1.518 Å (the values correspond to the subtraction of the minimal C3-C2' distance for pro-*S* and *R* conformation for each intercalation position).

4. Conclusions

By means of MD and QM/MM simulations, we modeled and analyzed the stabilities and structures for 28 different conformations of the supramolecular Cu / DNA complex formed between AT-rich DNA dodecamer and L-Cu(II)-R complex. This is the system which was applied successfully in experimental studies of intramolecular enantioselective Friedel-Crafts reaction catalyzed by Cu / DNA catalyst. The model conformations differ in the conformation mode of the metal complex, pro-*S* or pro-*R*, and its intercalation position in the DNA strand. Intercalation of the metal complex at the edge of the DNA strand, BP layers C1/A2, is least favorable and hardly induces any distortion of the DNA structure from the standard B type of DNA. Structures with intercalation of L-Cu(II)-R in the central part of the DNA dodecamer (BP layers A4/A5, A5/A6, A6/T7) show higher binding energy BE^{MD} as well as higher energy of interaction $Eint^{QM/MM}$. These complexes show DNA distortion towards A DNA and more pronounced axis bending. As a result of the interaction with the DNA strand, the metal complex itself is also deformed as loss of the indole-imidazol π-π stacking of the reactant and the electrostatic interaction between the π-electrons and the Cu(II) center takes place. Some of the complexes exhibit conformation with the indole moiety enveloped in the DNA minor groove.

On the base of the obtained results and energetic and structural analyses, the preferred formation of the *S* product in the experimental studies can be explained on the base of several characteristics of the system.

1. Stability of the supramolecular complexes: pro-*S* structures show higher binding energy to the DNA strand, and the formation of pro-*S* Cu / DNA complexes should be preferred to the formation of pro-*R* reactant complexes. Although we have not determined transition state structures and their stability for this first enantio-deciding C-C formation stage, one can suggest that the relative stability of the TS and the enantioselectivity is determined by the relative stability of the reactant complex, and therefore the *S*-selectivity is expected for this reaction, which explains the experimental findings.

2. Short C3-C2' distance in the supramolecular complexes: the C3-C2' distances in the pro-*S* complexes are shorter in comparison to the pro-*R* structures and closer to the expected TS structures. This feature also suggests the reaction to occur more easily for the pro-*S* structure.

The pro-*S* structures are usually only shallowly intercalated in the DNA strand, which would also facilitate TS formation decreasing the energy required for conformational changes.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge via the Internet at http://pubs.ac.org.

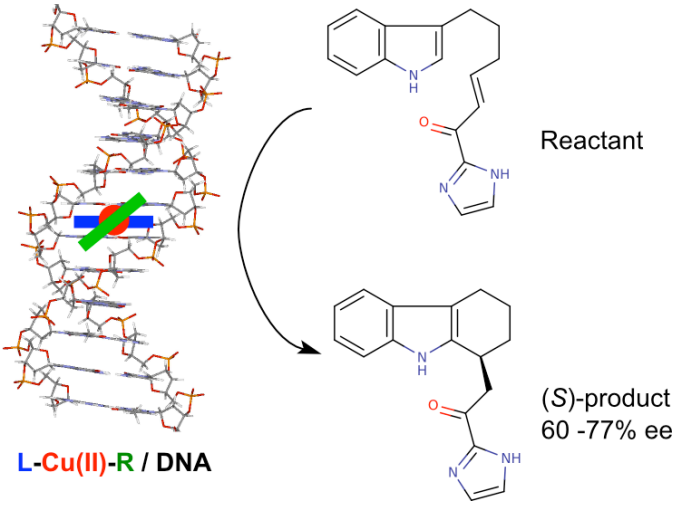
AUTHOR INFORMATION

Corresponding Author morokuma@fukui.kyoto-u.ac.jp

Notes The authors declare no competing financial interest.

ACKNOWLDEGEMENTS

This work was in part supported by a CREST (Core Research for Evolutional Science and Technology) grant in the Area of High Performance Computing for Multiscale and Multiphysics Phenomena from the Japanese Science and Technology Agency (JST) and in part by grants from Japan Society for the Promotion of Science (Grants-in-Aid for Scientific Research <KAKENHI> No. 24245005 and 25109525) at Kyoto University. The computer resources at the Academic Center for Computing and Media Studies (ACCMS) at Kyoto University and Research Center of Computer Science (RCCS) at the Institute for Molecular Science are also acknowledged.



Scheme 1. L-Cu(II)-R / DNA catalyzed unimolecular FC reaction [12].

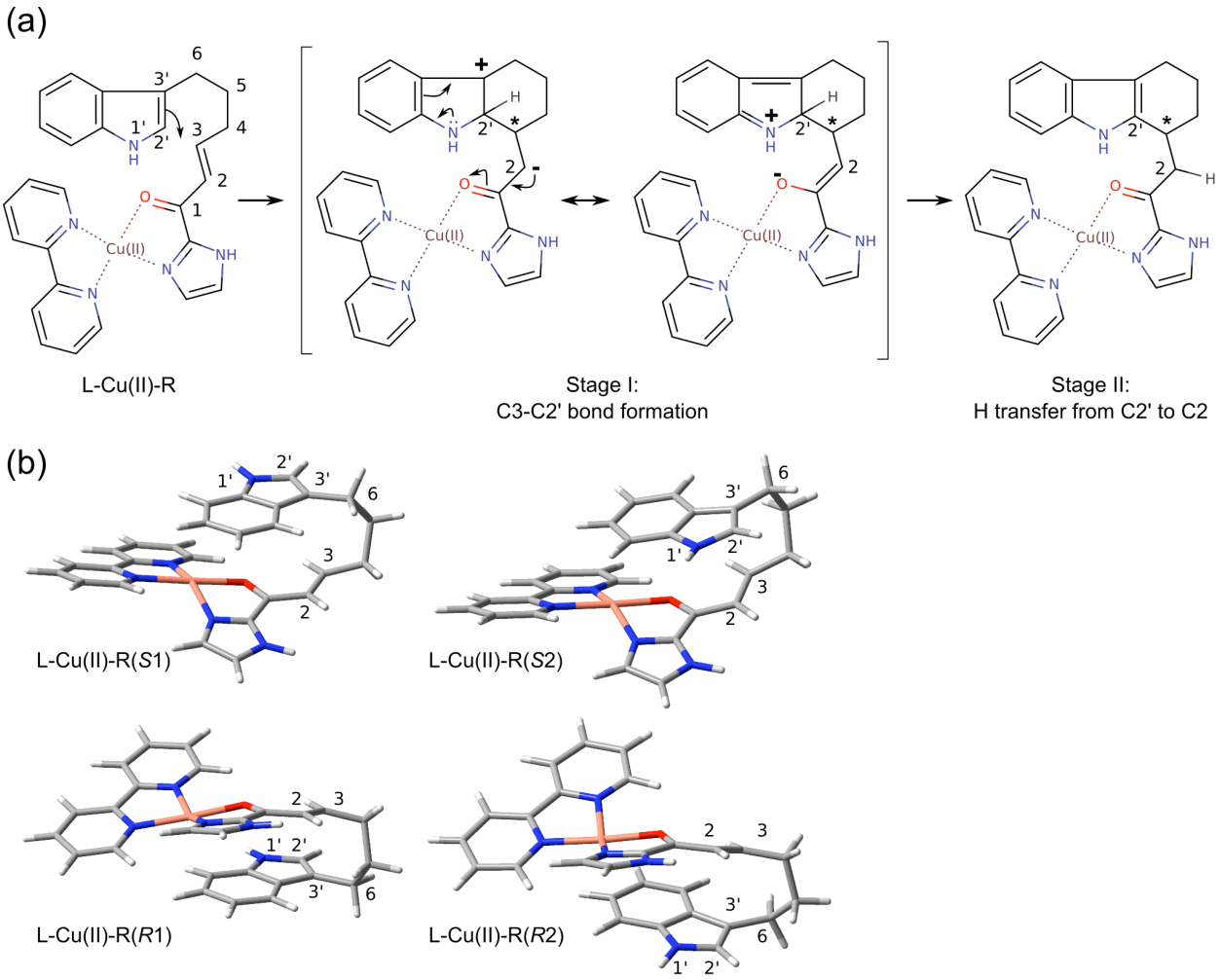


Figure 1. (a) Schematic representation of the intramolecular Friedel-Crafts reaction mechanism and notation of the atoms at the active part as used in the discussion; (b) Possible conformations of the bare L-Cu(II)-R complex.

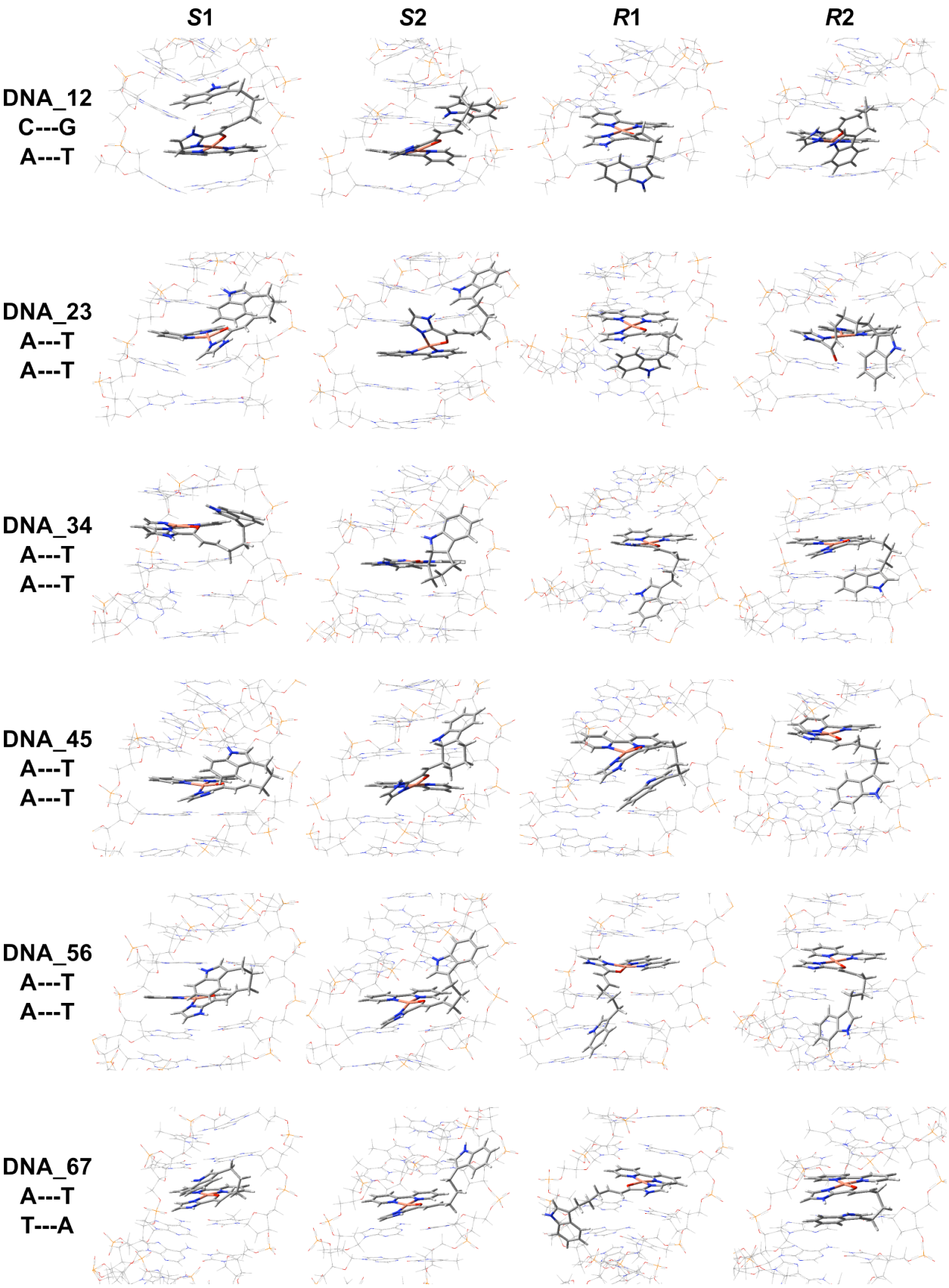


Figure 2. Close view of the metal complex intercalation in the Cu / DNA structures at ONIOM(M06-2x/6-31G*:AMBER) level (See Figs. S1-3 for DNA_67maj structures and the whole Cu / DNA supramolecular structures).

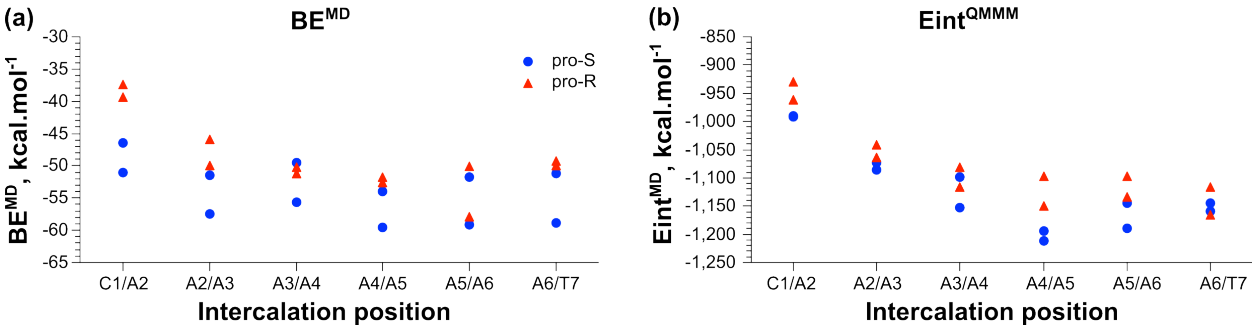


Figure 3. Binding energy at MD level, BE^{MD} (a), and interaction energy at ONIOM(M06-2x/6-31G*:AMBER) level, $Eint^{QM/MM}$ (b), as a function of the intercalation position *nm* of the supramolecular complex (blue for DNA_*nm*_*S*1 and _*S*2; red for DNA_*nm*_*R*1 and _*R*2).

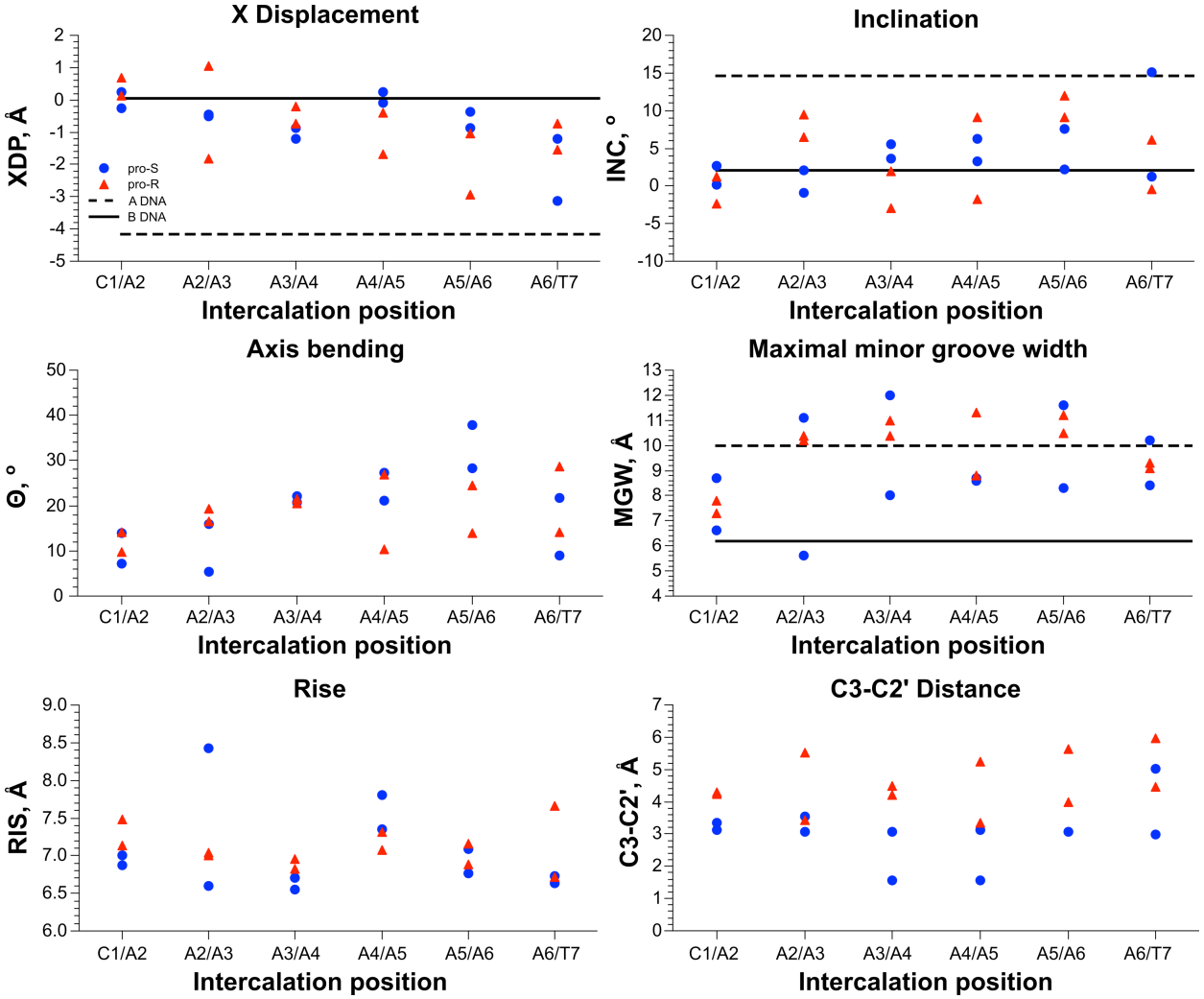


Figure 4. Average values of some DNA structural parameters for the most stable structure within each trajectory as functions of the metal intercalation position (blue for DNA_*nm*_*S*1 and *S*2; red for DNA_*nm*_*R*1 and *R*2). Typical values of the parameters for standard A and B type of DNA are shown with dashed and solid lines, respectively. For simplicity error bars showing deviation of the individual values within each structure from the average values are not shown.

Table 1. Relative energy (in kcal.mol^{-1}, with respect to the most stable conformation of the metal complex) and some distances (in Å) of L-Cu(II)-R at M06-2x/6-31G* in gas phase and in PCM water.

Structure	Gas phase				Water			
	Erel	C3-C2'	Ind-Cu[a]	Ind-Im[b]	Erel	C3-C2'	Ind-Cu[a]	Ind-Im[b]
L-Cu(II)-R(*S*1)	0.0	3.601	2.637	3.021	0.0	3.581	2.734	3.211
L-Cu(II)-R(*S*2)	1.2	3.376	2.549	3.152	0.6	3.367	2.676	2.800
L-Cu(II)-R(*R*1)	1.2	3.520	2.602	3.172	1.3	3.533	2.692	3.173
L-Cu(II)-R(*R*2)	3.1	3.385	2.568	3.237	3.0	3.546	2.663	2.814

[a] Distance between the plane of the indole and the Cu(II) center, and [b] between the plane of the indole and the imidazole part.

Table 2. Some energetic in kcal.mol^{-1} and structural characteristics of the model Cu / DNA supramolecular complexes: binding energy, BE^{MD}, estimated on the base of the MD results; deformation energy of DNA, DNA^{def}@MM, and that of the copper complex w/o, Cu^{def}@QM, and w/ solvent effect, Cuw^{def}@QM; interaction energy between the pure DNA structure (i.e. no water shell) and L-Cu(II)-R, $Eint^{QM/MM}$; C3-C2' distance (in Å) in the time average MD structure (Ave), the most stable MD structure (Min) and in the QM/MM optimized (QM/MM) structures; $RMSD^{AV}$ (in Å) is the time-averaged RMSD (DNA part and the Cu complex part) of the supramolecular complex with respect to the time-averaged structure; $RMSD^{MIN}$ is the RMSD (DNA part and the whole supramolecular complex) of the most stable MD structure of the supramolecular complex with respect to the time-averaged structure.

Complex	BE^{MD}	DNA^{def} @MM Eq. 1	Cu^{def} @QM Eq. 2	Cuw^{def} @QM Eq. 2	$Eint^{QM/MM}$ Eq. 3	C3-C2'			$RMSD^{AV}$		$RMSD^{MIN}$	
						Ave	Min	QM/MM	DNA	Cu	DNA	Supramolecular Complex
DNA_12_*S*1	-46.4	375.9	10.0	11.8	-992.4	3.455	4.061	3.346	1.399	0.559	2.208	1.440
DNA_12_*S*2	-51.1	352.6	27.0	32.6	-990.8	3.112	3.276	3.124	1.334	0.472	2.396	0.490
DNA_23_*S*1	-51.5	426.9	11.4	12.9	-1073.0	2.820	3.721	3.073	1.288	1.138	1.468	1.037
DNA_23_*S*2	-57.5	295.4	29.8	59.4	-1086.3	4.128	5.120	3.529	1.364	1.026	1.548	0.960
DNA_34_*S*1	-49.5	237.5	34.3	39.7	-1098.8	3.049	4.349	3.071	1.353	1.421	1.452	2.105
DNA_34_*S*2	-55.7	339.3	10.3	19.1	-1153.4	3.436	3.096	1.561	1.240	0.631	1.184	0.580
DNA_45_*S*1	-54.0	350.5	8.5	10.3	-1193.8	3.370	3.459	3.113	1.387	0.638	1.198	0.438
DNA_45_*S*2	-59.6	330.7	5.7	15.9	-1211.1	3.279	3.395	1.566	1.393	0.651	1.238	0.438
DNA_56_*S*1	-51.8	150.3	9.4	8.6	-1145.6	3.374	3.515	3.057	1.751	0.487	1.896	0.582
DNA_56_*S*2	-59.1	200.3	25.5	59.1	-1190.0	3.515	3.503	3.079	1.424	0.498	1.524	0.290
DNA_67_*S*1	-51.2	120.6	19.9	16.9	-1145.5	3.111	3.347	2.990	1.581	0.703	1.652	0.744
DNA_67_*S*2	-58.9	137.8	24.1	55.6	-1158.4	4.626	5.279	5.031	1.263	0.603	1.348	0.730
DNA_67maj_*S*1	-45.2	58.0	9.1	11.0	-1103.5	3.350	3.066	3.328	1.320	0.602	1.349	0.473
DNA_67maj_*S*2	-64.3	97.3	15.8	13.3	-1123.6	3.295	3.412	3.512	1.226	0.717	0.906	0.620
DNA_12_*R*1	-39.4	473.9	14.1	61.8	-930.3	4.631	5.211	4.285	1.395	1.340	2.496	1.668
DNA_12_*R*2	-37.4	460.2	16.7	15.2	-961.2	3.392	4.625	4.238	1.447	2.019	2.273	1.716

DNA_45_*R*1	-51.8	202.7	15.0	12.4	-1149.4	3.119	3.090	3.348	1.289	0.388	0.958	0.268
DNA_45_*R*2	-52.6	152.6	51.9	59.0	-1097.6	4.909	5.289	5.234	1.278	0.484	1.031	0.303
DNA_56_*R*1	-57.9	0.0	25.4	60.3	-1097.2	4.892	4.767	3.993	1.360	0.558	1.778	0.883
DNA_56_*R*2	-50.1	125.2	25.6	19.0	-1133.2	5.249	5.844	5.627	1.290	1.074	1.139	1.436
DNA_67_*R*1	-49.3	147.0	22.9	15.9	-1165.0	3.886	5.597	5.957	1.434	1.663	1.388	1.508
DNA_67_*R*2	-50.0	113.6	9.8	11.9	-1116.0	4.624	5.105	4.469	1.414	0.683	1.369	0.714
DNA_67maj_*R*1	-44.4	90.1	7.7	11.2	-1131.7	3.258	3.761	3.979	1.534	0.453	2.340	0.693
DNA_67maj_*R*2	-50.3	66.7	22.5	58.6	-1134.5	5.501	5.892	5.508	1.317	0.418	1.230	0.348

References

[1] A.J. Boersma, R.P. Megens, B.L. Feringa, G. Roelfes, Chem. Soc. Rev. 39 (2010) 2083-2092.
[2] A.J. Boersma, B. De Bruin, B.L. Feringa, G. Roelfes, Chem. Commun. 48 (2012) 2394-2396.
[3] A.J. Boersma, B.L. Feringa, G. Roelfes, Angew. Chem. Int. Ed. 48 (2009) 3346-3348.
[4] G. Roelfes, A.J. Boersma, B.L. Feringa, Chem. Comm. 6 (2006) 635-637.
[5] S. Park, H. Sugiyama, Angew. Chem. Int. Ed. 49 (2010) 3870-3878.
[6] S. Park, H. Sugiyama, Molecules 17 (2012) 12792-12803.
[7] G. Roelfes, B.L. Feringa, Angew. Chem. Int. Ed. 44 (2005) 3230-3232.
[8] G. Roelfes, B.L. Feringa, Angew. Chem. Int. Ed. 117 (2005) 3294–3296.
[9] S. Park, K. Ikehata, H. Sugiyama, Biomat. Sci. 1 (2013) 1034-1036.
[10] D. Coquière, B.L. Feringa, G. Roelfes, Angew. Chem. Int. Ed. 46 (2007) 9308-9311.
[11] A.J. Boersma, B.L. Feringa, G. Roelfes, Angew. Chem. Int. Ed. 48 (2009) 3346-3348.
[12] S. Park, K. Ikehata, R. Watabe, Y. Hidaka, A. Rajendran, H. Sugiyama, Chem. Commun. 48 (2012) 10398-10400.
[13] N. Shibata, H. Yasui, S. Nakamura, T. Toru, Synlett 7 (2007) 1153-1157.
[14] A.K. Patra, T. Bhowmick, S. Ramakumar, M. Nethaji, A.R. Chakravarty, Dalton Trans. 48 (2008) 6966-6976.
[15] H. Masuda, T. Sugimori, A. Odani, O. Yamauchi, Inorg. Chim. Acta 180 (1991) 73-79.
[16] S. Ramakrishnan, V Rajendiran, M. Palaniandavar, V.S. Periasamy, B.S. Srinag, H. Krishnamurthy, M.A. Akbarsha, Inorg. Chem. 48 (2009) 1309-1322.
[17] S. Park, L. Zheng, S. Kumakiri, H. Otomo, K. Ikehata, H. Sugiyama, submitted.
[18] J. Åqvist, J. Phys. Chem. 94 (1990) 8021–8024
[19] T. Darden, D. York, L. J. Pedersen, Chem. Phys. 98 (1993) 10089-10092.
[20] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, J. Comp. Phys. 23 (1977) 327-341.
[21] D.A. Case, T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, J. Comp. Chem. 26 (2005) 1668-1688.
[22] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, Comp. Phys. Commun. 91 (1995) 1-41.
[23] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Protein. Struct. Funct. Genet. 65 (2006) 712-725.
[24] Y. Zhu, Y. Wang, G. Chen, C.-G. Zhan, Theo. Chem. Acc. 122 (2009) 167-178.
[25] F. Wiesemann, S. Teipel, B. Krebs, U. Höweler, Inorg. Chem. 33 (1994) 1891-1898.
[26] A. Robertazzi, A.V.Vargiu, A. Magistrato, P. Ruggerone, P. Carloni, P. De Hoog, J. Reedijk, J. Phys. Chem. B 113 (2009) 10881-10890.
[27] J. Srinivasan, T.E. Cheatham III, P. Cieplak, P.A. Kollman, D.A. Case, J. Am. Chem. Soc. 120 (1998) 9401-9409.
[28] R. Lavery, M. Moakher, J.H. Maddocks, D. Petkeviciute, K. Zakrzewska, Nucleic Acids Res. 37 (2009) 5917-5929.
[29] http://gbio-pbil.ibcp.fr/cgi/Curves_plus
[30] J.S. Binkley, J.A. Pople, W.J. Hehre, J. Am. Chem. Soc. 102 (1980) 939-947.
[31] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K.Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S.Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma,

V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT, 2009.
[32] Zhao, Y., Truhlar, D.G., Theo. Chem. Acc. 2008, 120, 215-241.
[33] P. Fuentealba, H. Preuss, H. Stoll, L. Von Szentpály, Chem. Phys. Lett. 89 (1982) 418-422.
[34] J. Tomasi, R. Cammi, B. Mennucci, C. Cappelli, S. Corni, Phys. Chem. Chem. Phys. 4 (2002) 5697-5712.
[35] J. Tomasi, B. Mennucci, R. Cammi, Chem. Rev. 105 (2005) 2999-3093.
[36] S. J. Diekmann, Mol. Bio. 205 (1989) 787-791.