

## On The Effect of Methylation on DNA Structure and Dynamics

Sarkar A.K., Lahiri A.\*

University of Calcutta, Kolkata, India

\*[albmbg@caluniv.ac.in](mailto:albmbg@caluniv.ac.in)

DNA methylation is one of the major classes of epigenetic modifications and has a profound effect on gene expression. One of the mechanisms of epigenetic regulation of gene expression is the methylation of a promoter region which affects the binding of transcription factors. To study the types and extent of conformational changes in methylated DNA which may play a role in its function, we have carried out molecular dynamics simulations of two types of sequences, the Drew Dickerson dodecamer (DDD) and the cAMP DNA responsive element (CRE), with methylated and unmethylated sequences. We first tested the ability of the recent AMBER force fields in reproducing the conformational and dynamical propensities observed in various methylated forms of the DDD DNA in an NMR study and found that the OL15 force field gave results closest to the experimental observations. Simulations of the methylated and unmethylated counterparts of the CRE sequence with the OL15 force field revealed an overall increase in the end-to-end bending angle of the methylated CRE sequence as well as a decrease in the base-pair step flexibility at the hexameric core region. We discuss how these conformational and structural features may affect transcription factor binding to the CRE site.

*Key words:* Epigenetic modification, CpG methylation, CRE, DNA force fields, DNA flexibility.

### 1. Introduction

Force fields for simulating DNA have been under continuous refinement for quite some time [1–6]. Availability of a vast amount of experimental data and a significant increase in hardware and software capabilities have enriched the reference data set as well as allowed access to longer simulation time scales for more realistic systems. This not only unravelled errors in the force-field but also facilitated its refinement to a great extent. The AMBER [7] family of force-fields for nucleic acids is an excellent example where we observed a continuous effort to refine the parameters. The first correction tried to resolve the twist problem in sub-nanosecond time scale simulations with the parm94 set of parameters [8] leading to the development of the parm99 force field [9], which was further refined with parmbsc0 parameters [10, 11] to overcome the artifactual  $\alpha/\gamma$  transitions over multi-nanosecond time scales. The parmbsc0 force field along with the parm99 parameters were widely used for the simulation of nucleic acids until microsecond scale simulations revealed the existence of other errors which required further recalibration of the force-field (1, 2), leading to the development of the parmbsc1 set of parameters [12] and the OL15 family of force-fields [13–15].

In the present work, we have validated the two most recent AMBER force field modifications for DNA, the parmbsc1 [12] and the OL15 [13–15] force fields for double-stranded DNA having methylated residues, over a 100 ns simulation time scale. DNA methylation is one of the key processes in the epigenetic regulation of gene expression in complex organisms. In this process the cytosine residue of the d(CpG) step undergoes non-mutagenic chemical modifications resulting in

methylated cytosine [16]. In mammalian genomes, this process is carried out by DNA methyltransferases (DNMTs) [17]. DNA methylation is a potent suppressor of gene expression when it occurs in the promoter region [18]. One possible mechanism for gene repression is that the transcription factors can not bind to the methylated DNA [19]. One such example is provided by the binding of the transcription factors ATF/CREB to the cAMP DNA responsive element [CRE: d(TGACGTCA)<sub>2</sub>]. This binding gets inhibited by the methylation of the CpG motif located at the center of the CRE between the binding half-sites [20]. Inhibition occurs in spite of the fact that the modified cytosines in the CpG motif are not directly contacted by the proteins [20]. A short molecular dynamics (MD) simulation study on the methylated and the unmethylated CRE (referred to as CREmet and CRE respectively) reported insignificant differences in average structures [21]. However, as regards local flexibility at the base-pair level, methylation seems to decrease the flexibility as compared to the CRE. Methylation also changed the propensity for BII conformation of the base-pair steps around the CpG site though it does not affect the CpG step itself [21].

In another study using solid-state deuterium NMR on a DNA sequence having the EcoRI restriction binding site -GAATTC-, it was found that the cytidine methylation reduced the amplitudes of motion of the phosphate-sugar backbone [22]. They correlated this effect with the inhibition of restriction enzyme cleavage due to the methylation [22]. In a related study based on <sup>31</sup>P NMR on the dynamics of the Drew Dickerson Dodecamer (DDD) DNA, it has been shown that methylation changes the BI/BII propensity of the sugar phosphate backbone when methylation occurred at the

cytosine in the 3<sup>rd</sup> and the 9<sup>th</sup> positions [23]. However, this effect seems to be localized to only nearest neighbors and base-pairing partners [23].

Recent developments related to AMBER force-fields for nucleic acids aimed to improve the BI/BII distribution in the molecular dynamics simulation of B-DNA. It is also noteworthy that the force fields have been benchmarked several times on various forms of DNA, but, to the best of our knowledge, have never been tested on B-DNA having modifications like methylation. Here we present comparative results of MD simulations of the DDD-DNA in its two methylated states using two recent versions of AMBER force-fields (OL15 [13-15] and parmbsc1 [12]). We have also analyzed the effect of methylation on the CRE site as it is known to be involved in the recognition of the ATF/CREB family of proteins [20]. This study, to the best of our knowledge, is the first attempt to compare the ability of recent AMBER force fields to predict the effect of modifications, i.e., methylation on the structure and dynamics of DNA.

## 2. Materials and method

### 2.1. System preparation

The crystal structure of the canonical Drew Dickerson Dodecamer (DDD) DNA (PDB ID 1BNA) [24] was retrieved from RCSB [25]. The methylated variants were built using a suitable utility of AMBERTools14 [26]. Methyl groups were added to the cytosines at position 3 (referred to as 3CM) and at position 9 (referred to as 9CM) in both the strands of B-DNA duplex.

### 2.2. Force fields

The C5-methylated derivative of the deoxy cytosine (DMC) was modeled using the parameter developed by Lankas et al. [27] and deposited at the AMBER parameter database maintained by the Bryce group at the University of Manchester. On the other hand, standard residues were modeled using the FF99 [9] force-field in combination with one of the three corrections namely (1) bsc0 [11]; (2) bsc0 + parmbsc1 [12] (referred as bsc1) and (3) bsc0 + OL15 [13-15] (referred as OL15).

### 2.3. MD simulations

The canonical DDD DNA, henceforth referred to as the wild-type, and its methylated variants, henceforth referred to as 3CM (methylated at the 3<sup>rd</sup> position) and 9CM (methylated at the 9<sup>th</sup> position), were simulated separately with the three force fields mentioned above. In total, nine simulations, three for each DNA system, were carried out and in all the cases, the DNA was solvated in a truncated octahedral box of TIP3P water [28] in a manner such that the minimum distance between the box edge and the solute's surface was 10 Å. All molecular dynamics (MD) simulations were carried out using the MD simulation suite AMBER12 [29].

Neutralization was carried out by using Na<sup>+</sup> ions [30]. Prior to heating, two step energy minimizations were carried out. The first step of minimization was comprised of 500 steps of steepest descent followed by 1500 steps of conjugate gradient minimization holding the nucleic acid systems restrained with a force of 500 kcal/mol·Å<sup>2</sup>. In the second step of minimization, 1000 steps of steepest descent followed by 3000 steps of conjugate gradient were carried out without any restraint. After minimization, the systems were heated from 0 to 300 K in 200 ps with 2 fs time step using constant volume dynamics holding the solute with a restraint force of 25 kcal/mol·Å<sup>2</sup>. Langevin dynamics with random velocity scaling with a collision frequency of 1 ps<sup>-1</sup> was used. SHAKE [31] was turned on to constrain bonds involving hydrogen atoms and the particle mesh Ewald (PME) [32] method with 9 Å real space cutoff was used to calculate the electrostatic interactions. The van der Waals forces were truncated beyond the cutoff of 9 Å. Density equilibration was carried out by six consecutive constant pressure dynamics each of 100 ps long where the restraining force over the solute was released gradually from 25 kcal/mol·Å<sup>2</sup> to 0 kcal/mol·Å<sup>2</sup>. The reference pressure was 1 atm with a barostat relaxation time of 2 ps [33]. Production simulations were carried out under NPT condition maintaining 300 K temperature and 1 atm pressure using 2 fs time steps and the simulations were continued upto 100 ns.

### 2.4. Conformational analysis

End-to-end bend angle and base-pair step parameters were calculated using suitable utilities of Curves+ [34]. The base-pair step parameters of a particular base-pair step were combined to calculate the relative rigidity of that base-pair step in the DNA as

$$\begin{aligned} \text{Relative\_Rigidity} &= \\ &= \sum_{j=1}^{j=6} \left( \text{Variance}(j)_{\text{wildtype}} - \text{Variance}(j)_{\text{mutant}} \right), \end{aligned}$$

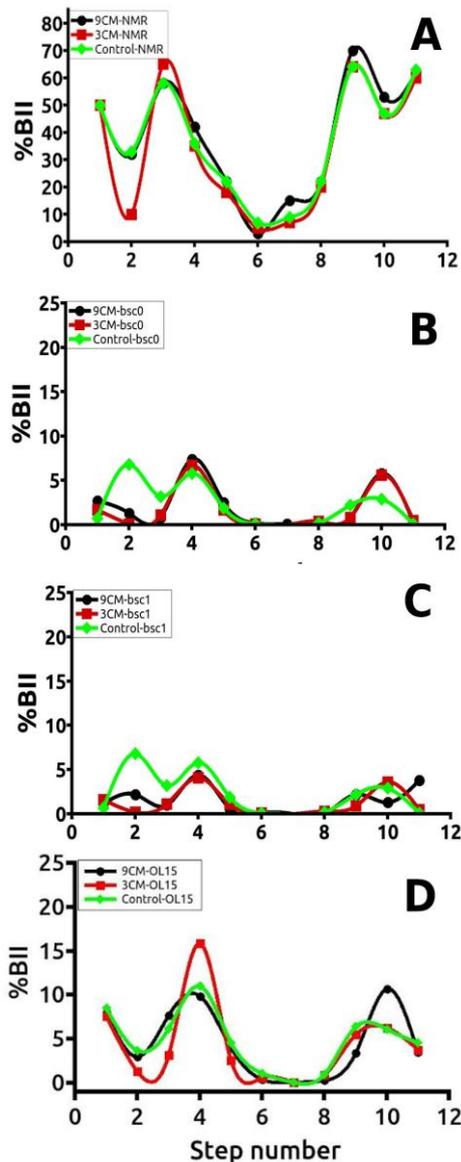
where step parameters (shift, slide, roll, twist, tilt, rise) were calculated and numerically referred to as  $j = 1 \dots 6$ . The DNA without methylation is defined as the wild type and the methylated counterpart is defined as the mutant. The sugar phosphate backbone conformation of the DNA was calculated in terms of the percentage of BI and BII conformations. The BI conformation is defined as having  $(\epsilon - \zeta) \sim -90^\circ$  and BII as having  $(\epsilon - \zeta) > 0^\circ$  [23].

## 3. Results and Discussion

### 3.1. Propensity for BII conformation: NMR vs simulations

The change in the backbone conformation in terms of BII (%) has been reported for the DDD DNA and its two methylated variants 3CM (symmetrically methylated at the 3<sup>rd</sup> position) and 9CM (symmetrically methylated at the 9<sup>th</sup> position) [23].

Based on their  $^{31}\text{P}$  chemical shift study, they have shown that the BII conformation gets affected at the neighbourhood of the methylated bases (Fig. 1,A). The distribution of BII conformation for the same system was calculated from the trajectories derived from bsc0, bsc1 and OL15 force fields (Fig. 1, B-D).



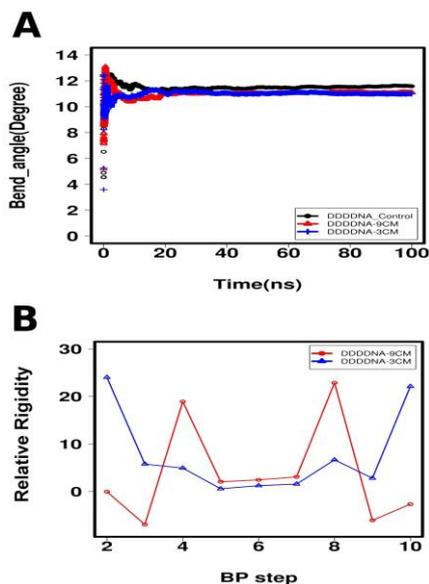
**Fig. 1.** The change in BII conformation as an effect of methylation. Green color is DDD DNA without methylation, red color represents 3CM and black represents 9CM. **A.** NMR data, **B.** bsc0, **C.** bsc1 and **D.** OL15 force fields.

From Fig. 1, it seems that the propensity for the BII conformation is quite similar to that observed in the NMR experiment (Fig. 1,A) for the case of OL15 force field (Fig. 1,D). Though the relative percentage of BII conformation is underrepresented but in comparison with bsc0 and bsc1, the OL15 seemed to perform well in this set of simulations.

### 3.2. Effect of methylation on the dynamics of DDD DNA: Analysis of the OL15 trajectory

In order to analyze the effect of methylation on the DDD DNA we have used two parameters, end-to-end bend angle and relative change in rigidity. The end-to-end bend angle was found to be less affected by the introduction of the 5-methylcytosine as depicted in Fig. 2,A.

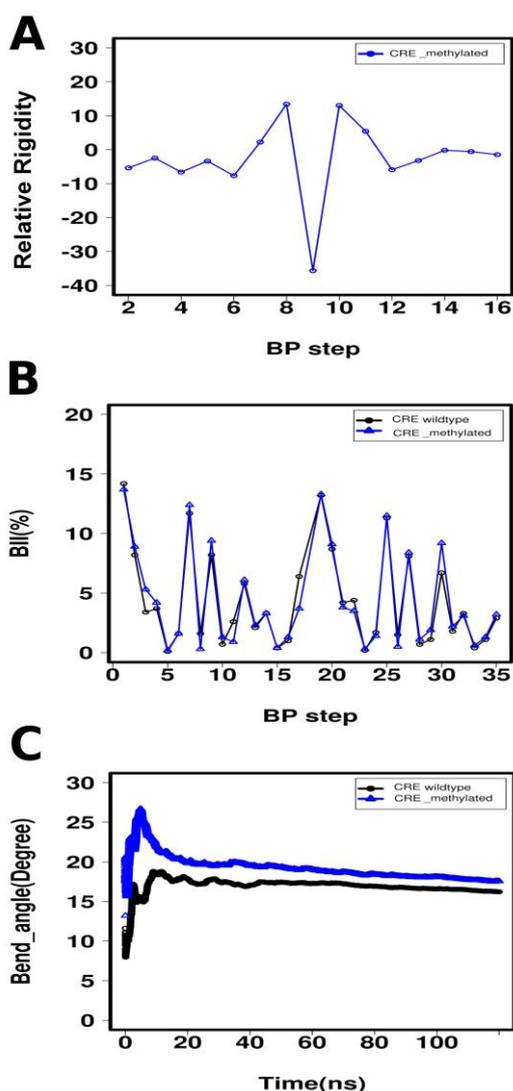
However, we observed a relative change in rigidity at the base pair step level in the two methylated variants of the DDD DNA (Fig. 2,B). Such a change in DNA flexibility due to methylation has been reported in DDD DNA based on the solid-state deuterium NMR studies [22].



**Fig. 2.** **A.** Change in the end-to-end bend angle and **B.** relative rigidity at the base pair step level due to methylation. Black color represents DDD DNA, blue is 3CM and red is 9CM.

### 3.3. Effect of methylation on the dynamics of CRE DNA

The effect of methylation at the CpG dinucleotide step in CRE has been analyzed using the OL15 force field. Similar to the DDD DNA simulation, we observed a relative change in rigidity at the base-pair step level in the methylated DNA (Fig. 3,A). However, no significant change in the BII conformational state was observed in the CRE sequence due to methylation (Fig. 3,B).



**Fig. 3.** A. Relative change in the rigidity in CRE DNA due to methylation (relative\_rigidity), B. occurrence of BII conformation in CRE (black) and CRE methylated (blue color) and C. time evolution of the cumulative average of bend angle, CRE (black) and methylated CRE (blue).

Surprisingly, we observed an increase in the end-to-end bend angle in the case of CRE having methylation at the CpG region as compared to the unmethylated CRE site (wild type). This increased bend in methylated CRE could be a reason for its inability to bind ATF/CREB protein. It has been suggested that the binding of ATF/CREB to the CRE site induces a bending in the CRE site towards the minor groove [35]. We hypothesize that ATF/CREB possibly would require comparatively more energy to bend the methylated CRE toward the minor groove because it is relatively more bent towards the major groove in comparison to the unmethylated CRE site. The requirement of this excess energy could be a reason behind the inability for complex formation.

## 4. Conclusions

Experimental studies on the B-DNA carrying the 5-methylcytosine (dmc) modification reported a relative increase in the BII conformational state and decrease in local flexibility at the base-pair step level [21–23]. In this study, one of our objectives was to test the ability of the recent AMBER force fields in reproducing the trend observed in various methylated forms of the DDD DNA [23]. Simulation of the DDD DNA and its two methylated variants using the OL15 force field showed a better agreement with the NMR observations in terms of the relative propensity for the BII conformational state. The exact values were, however, somewhat smaller than in the NMR data which can be due to different environmental conditions.

Further analysis of the trajectories simulated with the OL15 force field revealed an overall decrease in the base-pair step flexibility for the methylated DDD DNA in comparison to its native form. Experimental studies on the EcoRI restriction binding site carrying methylation at cytosine residue also reported the decrease in flexibility in terms of reduced amplitude of motion of the sugar-phosphate backbone. We also observed that the extent of reduction of base-pair step flexibility is dependent on the position of methylation. We extended this study further by investigating the dynamical behaviour of the CRE site in the presence and absence of the methylation using the validated OL15 force-field. We observed an overall increase in the end-to-end bending angle of the CRE site under methylation as well as a decrease in base-pair step flexibility at the hexameric core region (GACGTC). We speculate that this could affect the binding of the CRE site to its cognate transcription factor(s) which requires bending towards the minor groove.

## 5. Acknowledgements

The authors acknowledge support from the SERB project (EMR/2016/007753) and from the departmental DST-FIST and UGC-DSA programs. The UGC RFSMS and CSIR SRF programs are also acknowledged for providing fellowship to one of the authors (A.K.S.).

## 6. References

1. Dans P.D., Walther J., Gomez H., Orozco M. *Curr. Opin. Struct. Biol.* 2016. V. 37. P. 29–45.
2. Perez A., Luque F.J., Orozco M. *Acc. Chem. Res.* 2012. V. 45. P. 196–205.
3. Cheatham T.E., Case D.A. *Biopolymers.* 2013. V. 99. P. 969–977.
4. Orozco M., Noy A., Perez A. *Curr. Opin. Struct. Biol.* 2008. V. 18. P. 185–193.
5. Orozco M., Perez A., Noy A., Luque F.J. *Chem. Soc. Rev.* 2003. V. 32. P. 350–364.
6. Levitt M. *Cold Spring Harb. Symp. Quant. Biol.* 1983. V. 47. P. 251–262
7. Cheatham III T.E., Case D.A. *Biopolymers.* 2013. V. 99. P. 969–977.

8. Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A. *J. Am. Chem. Soc.* 1995. V. 117. P. 5179–5197. doi: [10.1021/ja00124a002](https://doi.org/10.1021/ja00124a002).
9. Wang J., Cieplak P., Kollman P.A. *J. Comput. Chem.* 2000. V. 21. P. 1049–1074. doi: [10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F).
10. Beveridge D.L., Barreiro G., Byun K.S., Case D.A., Cheatham T.E. 3rd, Dixit S.B., Giudice E., Lankas F., Lavery R., Maddocks J.H. et al. *Biophys. J.* 2004. V. 87. P. 3799–3813. doi: [10.1529/biophysj.104.045252](https://doi.org/10.1529/biophysj.104.045252)
11. Perez A., Marchán I., Svozil D., Sponer J., Cheatham III T.E., Laughton C.A., Orozco M. *Biophys. J.* 2007. V. 92. P. 3817–3829. doi: [10.1529/biophysj.106.097782](https://doi.org/10.1529/biophysj.106.097782).
12. Ivani I., Dans P.D., Noy A., Pérez A., Faustino I., Hospital A., Walther J., Andrio P., Goñi R., Balaceanu A. et al. *Nat. Methods.* 2016. V. 13. P. 55–58. doi: [10.1038/nmeth.3658](https://doi.org/10.1038/nmeth.3658).
13. Krepl M., Zgarbová M., Stadlbauer P., Otyepka M., Banáš P., Koča J., Cheatham III T.E., Jurečka P., Šponer J. *J. Chem. Theory Comput.* 2012. V. 8. P. 2506–2520. doi: [10.1021/ct300275s](https://doi.org/10.1021/ct300275s).
14. Zgarbova M., Luque F.J., Šponer J., Cheatham III T.E., Otyepka M., Jurečka P. *J. Chem. Theory Comput.* 2013. V. 9. P. 2339–2354. doi: [10.1021/ct400154j](https://doi.org/10.1021/ct400154j).
15. Zgarbova M., Šponer J., Otyepka M., Cheatham III T.E., Galindo-Murillo R., Jurečka P. *J. Chem. Theory Comput.* 2015. V. 11. P. 5723–5736. doi: [10.1021/acs.jctc.5b00716](https://doi.org/10.1021/acs.jctc.5b00716).
16. Doerfler W. *Annu. Rev. Bio-chem.* 1983. V. 52. P. 93–124.
17. Goll M.G., Bestor T.H. *Annu. Rev. Biochem.* 2005. V. 74. P. 481–514.
18. Jones P.A., Laird P.W. *Nat. Genet.* 1999. V. 21. P. 163–167.
19. Tate P.H., Bird A.P. *Curr. Opin. Genet. Dev.* 1993. V. 3. P. 226–231.
20. Iguchi-Arriga S.M.M. Schaffner W. *Genes Dev.* 1989. V. 3. P. 612–619.
21. Derreumaux S., Chaoui M., Tevanian G., Femandjian S. *Nucleic Acids Res.* 2001. V. 29. P. 2314–2326.
22. Geahigan K.B., Meints G.A., Hatcher M.E., Orban J., Drobny G.P. *Biochemistry.* 2000. V. 39. P. 4939–4946. doi: [10.1021/bi9917636](https://doi.org/10.1021/bi9917636).
23. Tian Y., Kayatta M., Shultis K., Gonzalez A., Mueller L.J., Hatcher M.E. *J. Phys. Chem. B.* 2009. V. 113. No. 9. P. 2596–2603. doi: [10.1021/jp711203m](https://doi.org/10.1021/jp711203m).
24. Drew H.R., Howard L.N. *Proc. Natl. Acad. Sci. USA* 1981. V. 78. No. 4. P. 2179–2183. doi: [10.1073/pnas.78.4.1981](https://doi.org/10.1073/pnas.78.4.1981).
25. Berman H.M. The Protein Data Bank. *Nucleic Acids Res.* 2000. V. 28. P. 235–242.
26. Case D.A. et al. AMBER 14. University of California, San Francisco. 2014
27. Lankas F., Cheatham III T.E., Spacková N., Hobza P., Langowski J., Sponer J. *Biophys. J.* 2002. V. 82. No. 5. P. 2592–2609. doi: [10.1016/s0006-3495\(02\)75601-4](https://doi.org/10.1016/s0006-3495(02)75601-4).
28. Jorgensen W.L., Chandrasekhar J., Madura J.D. *J. Chem. Phys.* 1983. V. 79. P. 926–935. doi: [10.1063/1.445869](https://doi.org/10.1063/1.445869).
29. Case D.A. et al. AMBER 12. San Francisco: University of California, 2012 .
30. Joung I.S., Cheatham III T.E. *J. Phys. Chem. B.* 2008. V. 112. No. 30. P. 9020–9041.
31. Ryckaert J.P., Ciccotti G., Berendsen H.J.C. *J. Comput. Phys.* 1977. V. 23. P. 327–341.
32. Darden T., York D., Pedersen L. *J. Chem. Phys.* 1993. V. 98. P. 10089–10092.
33. Berendsen H.J., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. *J. Chem. Phys.* 1984. V. 81. No. 8. P. 3684–3690. doi: [10.1063/1.448118](https://doi.org/10.1063/1.448118).
34. Lavery R., Moakher M., Maddocks J.H., Petkeviciute D., Zakrzewska K. *Nucleic Acids Res.* 2009. V. 37. No. 17. P. 5917–5929.
35. Paoletta D.N., Palmer C.R., Schepartz A. *Science.* 1994. V. 264. No. 5162. P. 1130–1133.