

## DNA 염기의 토토머화에 의한 염기쌍 변이 연구:

### 염기쌍의 전이 돌연변이를 중심으로

이연희, 이민준, 신석민

서울특별시 관악구 관악로 1, 서울대학교 자연과학대학 화학부, 151-742

E-mail: yoenhee722@hanmail.net, ak3220@snu.ac.kr

DNA 복제는 굉장한 정확도를 가지고 이루어진다. 하지만 내, 외부적인 여러 요인으로 돌연변이가 일어나기도 한다. 그 중 토토머화에 의해 치환 돌연변이가 일어난다는 가설은 오래전부터 그 가능성이 논의되어 왔고, 직접적인 증거를 찾으려는 노력도 있어 왔다. 토토머화에 의하여 DNA염기가 다른 형으로 변하면 왓슨-크릭 염기쌍 (아데닌-타이민, 시토신-구아닌이 수소결합한 염기쌍)이 아닌 다른 염기쌍이 생성될 수 있다. 이 염기쌍이 기준이 되어 DNA가 복제되기 때문에 결과적으로 전이 돌연변이(transition), 혹은 전환 돌연변이(transversion)된 염기쌍이 생성될 수 있다. 우리는 이런 사례 중에 염기쌍의 전이 돌연변이를 중심으로 연구하고자 한다. A-T염기쌍중 하나가 지배적인 아민형(amine form)이 아니라 이민형(imine form)으로 존재할 때 아민형과는 다른 수소결합이 가능해지며 보통 잘 생성되지 않는 A-C, G-T결합이 생성될 수 있다. 이후 그 염기쌍이 기준이 되어 DNA가 복제될 때에는 왓슨-크릭 염기쌍이 주로 생성되어 A-T염기쌍이 G-C 염기쌍으로 치환되는 DNA가닥이 생기게 된다. 우리는 이러한 과정을 에너지와 반응속도 측면에 집중하여 분석해보았다. 계산 결과, A-C\*, A\*-C, G-C\*, G\*-C의 염기쌍 생성이 왓슨-크릭 염기쌍과 비슷하거나 더 큰 정도로 에너지 면에서 유리하였으며 대부분의 돌연변이가 에너지 측면에서 보았을 때 A-C\* 염기쌍을 통하여 생김을 알 수 있었다. 계산된 값은 약  $10^{-6}$  정도의 빈도를 보였다.

#### 서론

생물 내에서의 DNA 복제는 개체의 유지와 생식에 중요한 의미를 가진다. 이러한 DNA 복제에는 여러 교정 메커니즘을 가진 중합효소들이 관여하며 굉장한 정확도를 가지고 이루어진다. 하지만 복제과정에서는 내부적, 외부적인 여러 요인으로 돌연변이가 종종 일어나며, 돌연변이의 내부적 요인으로는 토토머화(tautomerism), 효소의 에러 등이 존재하며, 외부적 요인으로는 화학물질, 방사능, 바이러스 등이 대표적으로 논의된다. 이러한 돌연변이는 때에 따라 유전자 발현에 큰 영향을 미칠 수

있다. 돌연변이에 의한 유전자의 변화는 진화의 중요한 조건이며, 진화방향에 영향을 미칠 수도 있다. 예를 들면 돌연변이가 진화에 영향을 미치는 메커니즘 중 하나로 biased mutation이라는 개념이 제시된다. Biased mutation이란 각각의 돌연변이가 유형에 따라 일어날 확률이 다르기 때문에 그에 따른 편향이 일어난다는 것이다.<sup>1</sup> 이런 경우 돌연변이의 유형을 분석하고 그 각각의 가능성을 알아보는 것은 진화과정을 이해하는데 도움을 줄 것이다.

토토머화(tautomerism)란 동시에 양성자와

이중결합 shift가 일어나 두 이성질체간에 전환이 일어나는 것을 말한다. 아데닌(A)과 시토신(C)의 경우 imine과 amine 형태, 티민(T)과 구아닌(G)의 경우 lactam과 lactim 형태의 토토머가 존재한다. 이 중 더 적은 비율로 존재하는 토토머일때 염기에 \*을 붙여서 표기하기로 한다.<sup>2</sup>

토토머화에 의해 치환 돌연변이가 일어난다는 가설은 왓슨과 크릭에 의해 DNA 염기쌍의 구조가 막 밝혀지기 시작했을 때부터 고려되어 오래 전부터 그 가능성이 논의되어 왔다.<sup>3</sup> 실제로 X-ray crystallography를 이용해 A-C의 비왓슨-크릭염기쌍이 관측되었고 이는 A-T 염기쌍과 유사한 모양으로, A나 C가 토토머화하여 결합된 것으로 추정되기도 하였다.<sup>4</sup>

Topal과 Fresco가 제시했듯이 토토머에 의한 돌연변이가 일어나는 메커니즘은 주로 2가지이다.<sup>5</sup> 하나는 퓨린과 피리미딘간의 결합에 의해 일어나는 전이 돌연변이(transition)이고 다른 하나는 퓨린사이의 결합으로 일어나는 전환 돌연변이(transversion)이다. 이 중 전환돌연변이의 경우 염기쌍의 크기가 통상적인 염기쌍보다 크기 때문에 중합효소의 검토에 걸려 제거될 가능성도 더 크고 DNA strand에서 strain이 걸려 불안정해져 염기쌍의 결합 에너지만으로 안정성을 비교하기 힘들 것이다. 따라서 우리의 논문에서는 비교적 간단한 계산으로 분석할 수 있는 전이 돌연변이(transition)의 발생에 대해 분석해보려고 한다.

전이 돌연변이(transition)의 경우 먼저 DNA가 복제되어 새로운 염기쌍이 생길 때 토토머화된 염기와 수소결합을 형성해 일반적이지 않은 결합이 생성된다. 예를 들어 A-C는 비왓슨-크릭염기쌍(NWCP)으로 통상적으로 왓슨-크릭염기쌍(WCP) 만큼 강하게 결합하기 힘들다.

하지만 A-C\*의 경우 2개의 수소결합이 생성될 수 있어 안정된 결합을 이룰 수 있다. 이후 분열을 거쳐 결국 왓슨-크릭염기쌍이 생성되며 A-T 염기쌍이 G-C로, 혹은 그 반대로 변하는 결과를 낳을 수 있다.

### 이론 및 계산방법

염기, 염기쌍의 최적화된 구조, transition state 모두 GAMESS에서 Restricted Hartree Fock Approximation(RHF)으로 6-31G Basis를 사용하여 계산되었다. Hatree fock method에서는 각각의 전자가 핵과 나머지  $N_e-1$ 개의 전자에 의해 생긴 전기장을 이동한다고 취급된다. 가장 낮은 에너지의 파동함수를 주는 스핀 오비탈을 variation theory를 이용해 찾을 수 있다. Open shell 시스템일 경우 Restricted open shell Hartree Fock method (ROHF)나 Unrestricted Hartree Fock method를 쓸 수 있으나 우리가 계산하는 시스템은 모두 closed shell이므로 RHF를 사용했다. RHF에서 각각의 오비탈 파동함수를 구하는 식은 다음과 같다.<sup>6,7</sup>

$$\hat{F}u_i = \varepsilon_i u_i \quad i = 1, 2, \dots, n$$

여기서  $\hat{F}$ 는 fock operator,  $u_i$ 는  $i$ 번째 spin orbital이며 eigenvalue  $\varepsilon_i$ 는 spin orbital  $i$ 의 오비탈 에너지이다.

또한 실제 생물체내에서는 분자들이 물에 의해 solvation되어있으므로 Polarizable Continuum Model(PCM)을 이용하여 계산하였다. PCM에서는 각각의 원자핵이 그 원자의 반데르발스 반지름의 1.2배를 가지는 구로 둘러싸여 있다. Polarized dielectric continuum에 의해 만들어진 electric potential은 이러한 구 표면에 분포된 apparent surface charge와 같고, 이러한 charge를 iterative한 방법으로 계산해 낸다.

표 1. 각각의 염기의 형태에 따른 에너지와 두 형태간의 평형상수 계산

base	E(Hartree)	base	E(Hartree)	$\Delta E(\text{kcal/mol})$	평형상수
A	-503.3278	A*	-503.3075	12.7	$2.20 \times 10^9$
C	-431.4797	C*	-431.4687	6.9	$1.15 \times 10^5$
T	-490.3388	T*	-490.3086	19.0	$7.91 \times 10^{13}$
G	-578.1866	G*	-578.1684	11.4	$2.37 \times 10^8$

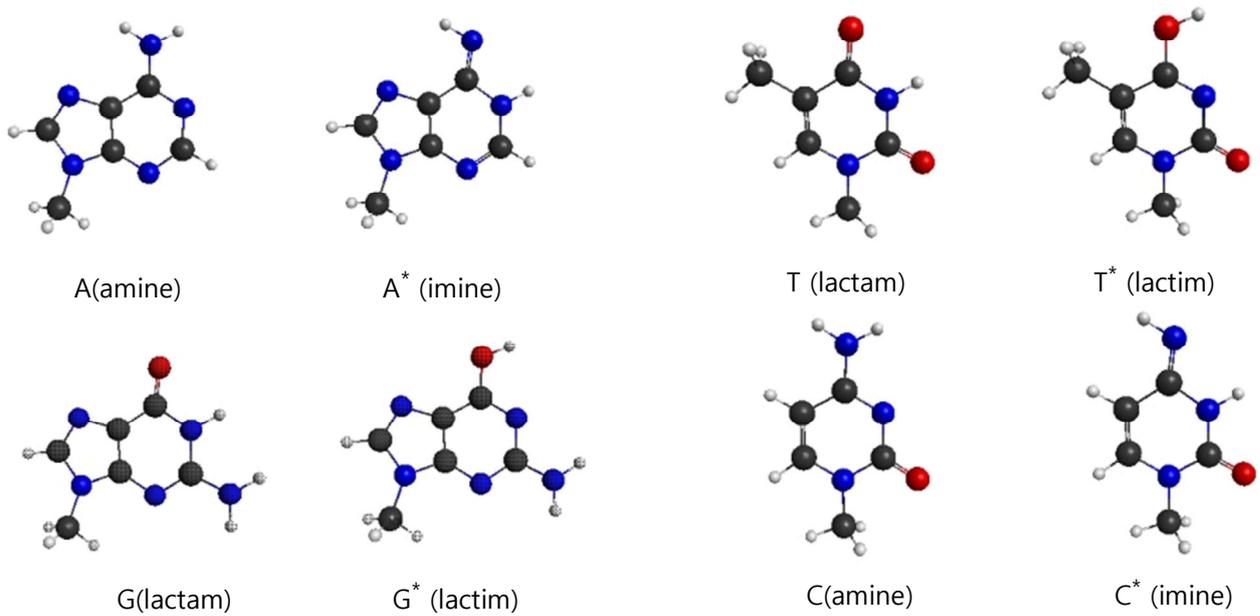


그림 9. DNA 염기의 토토머 형태들

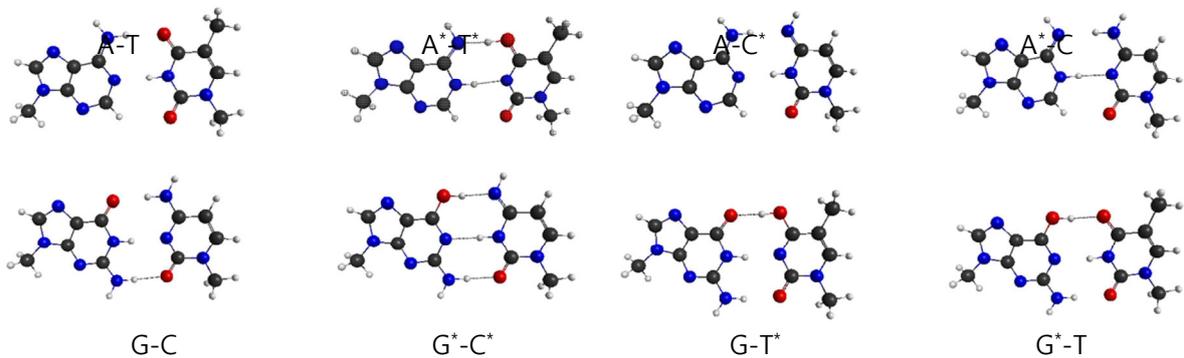


그림 2. DNA 염기쌍들

표 2. 가능한 염기쌍의 에너지와 C1-C1간의 거리 측정

Base Pair	Energy(Hartree)	$\Delta E(\text{kcal/mol})$	Length(Å)
A-T	-993.6783	7.34	10.59
G-C	-1009.6899	14.8	10.75
A*-T*	-934.8091	7.91	10.74
G*-C*	-934.8034	10.2	10.69
A-C*	-1068.5258	11.7	10.6
A*-C	-1068.5233	17.6	10.37
T-G*	-993.6413	15.8	10.97
T*-G	-1009.6588	13.6	10.77

\* $\Delta E$ 는 계산된 두 base monomer의 에너지 합계에서 형성된 염기쌍의 에너지를 뺀 값이다.

결론 및 토의

각각의 base에 대해 tautomer 구조를 만들어 계산해본 것이 표 1과 그림 1에 나타나 있다. tautomer사이의 평형상수는  $1.15 \times 10^5$ 에서  $7.91 \times 10^{13}$ 까지 각각의 base에 따라 많은 차이가 났다. 가장 rare tautomer 형태로 많이 존재하는 것은 C로, 이는 아민기가 이민기가 되면서 conjugated된 부분이 생겨 안정화 되기 때문으로 추정된다.

그림 2에서 가능한 염기쌍의 구조를 보여준다. 표 2에서 보듯이 염기쌍을 이룰 경우 토토머화된 염기가 이루는 쌍이 생성되며 얻어지는 에너지는 왓슨-크릭염기쌍과 비슷하거나 오히려 높았다. 따라서 토토머화 된 염기와 무리없이 염기쌍을 생성할 수 있을 것이다. 또한 당의 1번 탄소대신 붙인 메틸기의 탄소 사이 거리를 척도 삼아 켜 크기도 왓슨-크릭염기쌍과 토토머가 포함된 염기쌍이 비슷한 값을 보였다. 따라서 정상적인 DNA구조에서 크게 벗어나지 않고, polymerase의 active site에서 형성될 수 있을 것이라 예상되는 조건을 갖추고 있다.

또한 주목할 점은 A-C\*와 A\*-C, T-G\*와 T\*-G

의 경우 서로 바뀔 수 있다는 것이다. A-C\*와 A\*-C의 경우 단계적으로 수소를 주고 받아 변할 수 있고 T-G\*와 T\*-G의 경우 concerted되어 하나의 transition state를 통해 변이한다. 둘 다 약 15 kcal/mol의 activation barrier를 가지고 있어 상온에서도 충분히 반응이 일어난다. 따라서 충분한 시간이 있다면 A와 C 염기쌍의 경우 대부분 더 에너지적으로 안정한 A-C\*로, G와 T 염기쌍의 경우 T-G\*로 변화할 것이다.

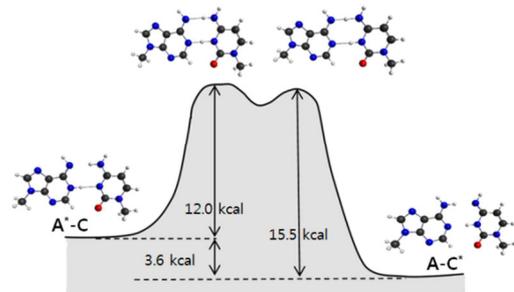


그림 3. A\*-C와 A-C\*간의 transition state

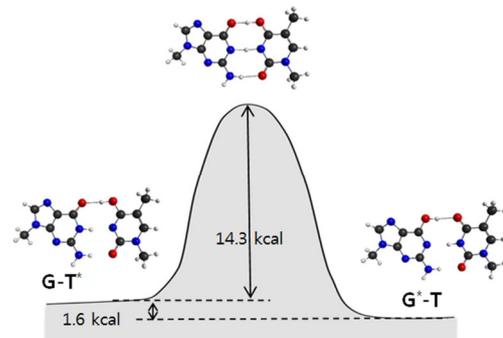


그림 4. G-T\*와 G\*-T간의 transition state

토토머화에 의한 전이 돌연변이의 경우 우선 중합되는 한쪽의 DNA가 토토머화 되어 있어야 한다. 상황을 간단하게 하기 위해 우선 처음 template이 되는 가닥은 주로 토토머화 되지 않은 그대로라고 가정해보자. 중합과정에서 template에 맞는 염기가 중합효소에 의해 염기쌍을 이룬다. 예를 들어 A의 경우 T나 C 토토머가 수소결합에 의한 염기쌍을 이룰 수 있다. 여기서 토토머로 존재하는 비율이 다르기 때문에 각각의 비왓슨-크릭염기쌍의 형성비는 다를 것이다. 예를 들어 A-C\*가 G-T\*의 비율보다 클 것이라고 예측 할 수 있다.

실제로는 template에서 제공되는 염기가 토토머화된 경우도 생각해 볼 수 있다. 왓슨-크릭 염기쌍으로 결합된 상태에서 한쪽만 다른 형태로 바뀌면 합이 매우 불안정해 지기 때문에 토토머화가 될 경우 염기 두 개가 동시에 토토머화 될 것으로 예상할 수 있다. 이 또한 계산해 볼 수 있는데 표 2에 나와있는 계산된 에너지 값을 이용하면 A-T와 A\*-T\*간의 평형상수는  $1.07 \times 10^{17}$ , G-C와 G\*-C\*간의 평형상수는  $2.05 \times 10^{14}$  가 나온다

계산한 값들을 이용하면 돌연변이의 빈도수를 대략적으로 계산 해볼 수 있다. 또한 중합 과정에서 주어진 템플릿이 그대로 유지된다고 가정한다. 따라서 모든 A\*, T\*, G\*, C\*는 A\*-T\*, G\*-C\* 염기쌍에서 공급된다고 가정한다. 템플릿에 존재하는 하나의 염기에 대하여 각각 가능한 두 가지 경우를 고려해 볼 수 있다. 예를 들어 A의 경우는 T혹은 C\*와 결합을 형성할 수 있고 전체 형성되는 결합에서 C\*와 결합하는 빈도수가 얼마나 되는지 고려해 볼 수 있을 것이다. 실제로는 염기쌍이 결합하는 에너지가 염기쌍의 종류에 따라 다르기 때문에 이 또한 형성 확률에 영향을 미칠 것이나, 표 2에서 구하였듯이 에너지 차이는 크지 않다고

생각되어 빈도가 용액 속의 해당 화학종 농도에 단순히 비례한다고 가정하였다.

구하는 방법은 아래와 같다. 템플릿의 염기X가 염기Y와 염기Z\*랑 염기쌍을 형성할 때, 각각 존재할 확률을 x, y, z라 하자. 여기서 x는 A-T, G-C간의 비율과 A-T와 A\*-T\*간의 평형상수와 G-C와 G\*-C\*간의 평형상수에서 구할 수 있으며, y와 z는 생물체내에서 중합에 공급되는 염기들의 비율과 각 토토머의 존재비로 구할 수 있다. 전체 염기쌍에서 G-C염기쌍의 비율은 생물마다 다르며 인간에서는 약 40%정도이다. 여기서는 A-T와 G-C비율이 6대 4라고 가정하고 계산해보았다.

$$\text{템플릿의 X에 의한 돌연변이 빈도} = x \times \frac{z}{y+z}$$

총 돌연변이의 빈도수는  $10^{-6}$  정도이며 그 대부분을 A-C\* 염기쌍이 차지한다. 이는 C\*가 다른 염기들의 드문 토토머에 비해 안정되어 있어 농도가 비교적 높기 때문이다. 따라서 토토머에 의한 돌연변이의 경우 템플릿이었던A에 C가 결합하며 결국 GC형태로 변하는 돌연변이가 대부분을 차지할 것이다. 즉, A-T염기쌍이 G-C로 변하는 것이 주가 될 것이다. 자연적인 돌연변이 비율은  $10^{-4} \sim 10^{-11}$  로 생물 종에 따라 다를 수 있으며 인간의 경우  $10^{-8}$  정도로 추정되고 있다. 인간의 돌연변이 비율을 생각하면 실제보다 크게 나온 것이라고 볼 수 있는데, 이는 polymerase가 가지는 여러 오류 수정기능이 영향을 미쳤을 것으로 생각된다. 예를 들면, polymerase는 active site에서 왓슨-크릭 염기쌍을 분별해내는 여러 기능을 가지고 있다. 우리가 구한 염기쌍이 비록 왓슨-크릭 염기쌍과 매우 유사한 구조를 가지고 있지만, 계산할 때 가정했듯이 염기쌍이 형성될 확률이 같다는 보장은 없다.

표 3. 각 가능한 염기쌍의 빈도수 계산

Template	NWC Pair를 이루는 base	NWC pair/생성되는 pair	전체 중합에서의 빈도수
A	C*	$5.78 \times 10^{-6}$	$1.73 \times 10^{-6}$
T	G*	$2.81 \times 10^{-9}$	$8.42 \times 10^{-10}$
G	T*	$1.90 \times 10^{-14}$	$3.79 \times 10^{-15}$
C	A*	$6.82 \times 10^{-10}$	$1.36 \times 10^{-10}$
A*	C	1	$2.82 \times 10^{-18}$
T*	G	1	$2.82 \times 10^{-18}$
G*	T	1	$9.75 \times 10^{-16}$
C*	A	1	$9.75 \times 10^{-16}$
총 돌연변이 빈도수			$1.73 \times 10^{-6}$

또한 DNA 중합효소에는 proofreading 메커니즘이 존재하기 때문에, 대부분의 돌연변이는 수정되어 없어진다. 따라서 실제로 나타나는 돌연변이의 빈도는 훨씬 낮을 것이다.

#### 결론

염기의 토토머화는 분자를 매개로 반응이 진행되고, 주변 환경에 의해 영향을 많이 받을 수 있다는 측면 때문에 이론적으로 접근하기 어려웠다. Molecular dynamics 계산을 통해 접근한 논문도 있었지만, 양자 역학으로 에너지 계산을 통해 돌연변이의 빈도를 예상해보려는 시도는 없었다. 우리는 RHF 계산을 통해 토토머에 의한 전이 돌연변이의 가능성을 에너지와 반응속도 측면에서 분석해 보았다. 토토머로 인해 생성된 일반적이지 않은 염기쌍들은 왓슨 크릭 염기쌍과 비슷한 크기와 비슷한 구조를 가지며, 염기쌍 형성으로 얻어지는 에너지면에서 보았을 때에도 비슷하거나 조금 더 유리했다. 간단한 계산을 통해  $10^{-6}$  정도의 빈도로 돌연변이가 나타남을 계산할 수 있었다.

#### 감사의 글

본 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단 첨단 사이언스·교육 허브 개

발 사업의 지원을 받아 수행된 연구임(No. NRF-2012-M3C1A6035357)

#### 참고문헌

1. Yampolsky, L. Y. and Stoltzfus, A. 2008. Mutational Biases. eLS. .
2. Peter Vollhardt and Neil Schore., Organic Chemistry 6th ed., Freeman, p. 584, Freeman (2011)
3. Watson, J. D., and Crick, F. H. C., Nature 171, 737-738 (1953)
4. Weina Wang, Homme W. Hellinga, and Lorena S. Beese. Proc Natl Acad Sci U S A. 2011, 108(43): 17644-17648.
5. MICHAEL D. TOPAL AND JACQUES R. FRESCO, Nature 263, 285 - 289, 1976
6. Peter Atkins and Ronald Friedman, Molecular Quantum Mechanics 5th ed., pp. 297~298, Oxford, 2011
7. IRA N. Levine, Quantum Chemistry 6th ed., pp. 312~316, Pearson
8. Michael W. Nachman and Susan L. Crowell, Genetics September 1, 2000 vol. 156 no. 1 297-304