

## Deterministic intermediate Sampling과 Yen's algorithm을 이용한 Urey-Miller 반응 경로 제시

김진우,<sup>1</sup> 김연준,<sup>1</sup> 김우연<sup>1\*</sup>

대전시 유성구 대학로 291 KAIST 화학과, 305-701.

E-mail: starhouse@kaist.ac.kr

이 연구는 초기 지구의 대기 환경에서 유기화합물(glycine)이 합성되는 실험(Urey-Miller 실험)에서의 반응 경로를 Deterministic한 방법의 중간체 sampling 방법으로 반응 네트워크를 구성하고 Yen's 알고리즘으로 네트워크 내의 최단경로를 제시함으로써 반응물과 생성물이 결정되어 있을 때 최소한의 화학적 직관만을 이용하여 제시하는 것이 목표이다. 이 연구 결과는 2014년도 Nature Chemistry에 발표된 다른 방법론을 적용하여 제시된 Urey-Miller reaction path와 비교해 어떠한 반응이 상대적으로 더 타당한 경로를 제시했는지 알아보았다. 이 연구에서 나온 reaction path에서의 중간체들에 대해 GAMESS를 이용한 B3LYP/6-31g(d,p) DFT계산을 수행하였다. 결과를 분석해보면서 어떤 부분이 부족하며 이 연구에 적용한 방법론을 어떻게 발전시켜나가야 더 나은 결과를 얻을 수 있을지를 함께 고려해 보았다.

핵심어 : Urey-Miller reaction, reaction path, DFT

### 서론

반응 경로를 예측하는 것은 화학을 연구하는 목표 중의 하나이다. 화학적 경향성을 토대로 알려진 분자내의 reactive atom들을 선별하고 유기화학의 기본적인 SN1, SN2, E1, E2 반응들과 같은 각 step을 구성하는 기초적 메커니즘들을 조합하여 reaction path를 알아낸 후, 이 때 나온 중간체들에 대하여 optimized geometry에서의 에너지 계산을 통해 제시된 반응 경로의

타당성을 고려해보는 것이 일반적인 반응 경로 예측 방법의 하나이다. 그러나 이 경우 미리 정해진 화학적 직관을 바탕으로 제시된 반응 경로이기 때문에 실제 반응에서 일어날 수 있는 수 많은 화학 결합의 생성과 분해를 대부분 고려할 수 없고, 충분히 가능성이 있는 반응경로를 제시 하지 못 할 수도 있다. 특히, 여러 reactive atom들이 포함된 복잡한 시스템을 다룰 수록 그런 예측의 불확실성은 더욱 커질 수 밖에 없다. 따라서 화학적인 직관을 최소한으로 하고, 기본적인

variable들만을 미리 정해주면 예상되는 반응 경로를 예측해보도록 하는 계산 방법들이 연구 되어지고 있다. 그러한 연구들 중 반응물들의 구조를 특정 기준에 따라 fragmentation한 후 deterministic sampling으로 가능한 모든 중간체들을 만들어 중간체들끼리의 네트워크를 구성하고 Yen's algorithm을 이용하여 반응 경로 예측을 하는 코드가 개발되어지고 있는데<sup>1</sup> 현재까지 개발된 코드를 이용하여 Urey-Miller 반응의 경로를 예측해보는 작업을 진행하였다. Urey-Miller 반응을 선택한 것은 이 반응이 초기 지구의 대기 환경으로부터 어떻게 유기화합물이 합성되어졌는지를 알 수 있다는 반응 자체로서의 중요도 뿐만이 아니라 C,H,O,N을 포함하고 있기 때문에 이 후 연구 되어질 많은 반응들을 적용해보기 전에 코드의 결과나 실행능력을 테스트 해 볼 수 있는 가장 기본적인 example이 될 수 있기 때문이다. 2014년 Nature Chemistry에서 AIMD(ab initio molecular dynamics) nano reactor를 이용해 반응 경로를 예측하는 논문<sup>2</sup>에서 분석한 Urey-Miller 반응 경로와 앞에서 말한 코드를 이용하여 제시된 Urey-Miller 반응 경로를 비교 분석해 보기로 하였다. GAMESS를 이용해 각 각의 방법으로 제시된 반응 경로 후보들의 중간체들에 대한 에너지 계산을 하여 energy diagram을 만들어 보았고, 그 diagram분석을 통해서 경로 예측의 타당성을 비교하도록 하였다. 새로 만들어지고 있는 code에 깔린 화학적인 원리들을 통해 그러한 결과가 나오게 된 원인을 분석해 보았다. 또한 code에서 fragmentation과 postscreening하는 기준에서 적용되는 학부 과정의 화학과 관련성과 그 기준들이 결과에 어떠한 영향을 미칠 수 있었는지에 대해서도 논해 본다.

## 이론 및 계산방법

### 1. Code description

다음과 같은 4 가지 과정에 의해 reaction path 들의 후보를 제시할 수 있다.

- Fragmentation(Input)
- Sampling Intermediates
- Postscreening
- Generating Edges and Suggesting reaction path(Output)

Fragmentation 은 주어진 반응물들을 특정 기준에 따라 Sampling 의 재료들로서 특정 결합들을 쪼개어 두는 것이다. 이 과정에서는 반응물이 주어졌을 때 어떤 fragmentation 을 할 것인지는 체계적인 방법이 정해져 있지 않고 사용자의 화학적인 직관이나 실험적인 경험이 적용되게 된다. 예를 들면 전이금속 화합물이 포함된 반응에서 리간드와 다른 화합물 간의 반응이 일어나는 경우가 없다고 판단 될 때 리간드를 fragmentation 을 하지는 않는다. 마찬가지로 saturated hydrogencarbon 의 chain 의 경우 backbone 을 이루는 공유결합의 경우는 fragmentation 하지 않을 수 있다. 이렇게 화학반응에 직접적으로 관여하지 않는 원자들의 결합들은 쪼개지 않고, 나머지 결합들은 모두 쪼개어 둔다. 모든 결합을 쪼개어 input 으로 설정해도 이론적으로 reaction path 가 나올 수는 있지만 fragment 가 많아질 수록 실행시간이 크게 늘어나기 때문에 이런 과정이 필요하다.

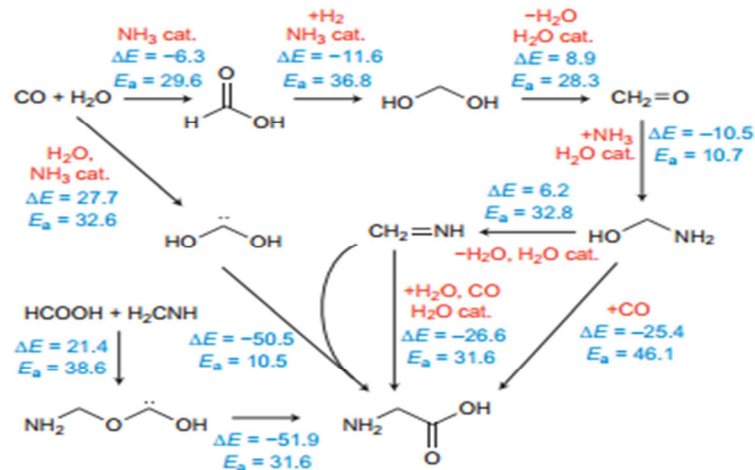
중간체 Sampling 을 할 때 deterministic sampling 방법을 적용한다. 간단히 설명하면

Fragmentation 과정에서 만든 반응물들의 fragment 들 사이에서 일어 날 수 있는 화학 결합의 조합을 모두 고려하여 중간체들의 구조(3D geometry)를 얻는 방법이다. 이 과정에서는 화학적인 직관이나 실험적인 경험이 전혀 고려 되지 않고 오직 조합론적으로 반응에서 생길 수 있는 모든 중간체들을 고려할 수 있게 한다.

Postscreening 과정에서는 일반화학에서 배운 가장 기본적인 규칙에 따라 Sampling 과정에서 형성된 중간체들의 구조들 중 화학적으로 구조의 안정화가 불가능한 중간체들을 걸러내게 된다. 예를 들면 Sampling 과정에서는 모든 결합 생성의 경우를 고려 했기 때문에 탄소 원자가 5 개 이상의 결합 개수를 갖거나 산소가 세개 이상, 수소가 두개 이상, 질소가 5 개 이상의 결합을 할 경우가 존재한다. 이런 경우는 화학 반응에 포함되지 않을 가능성이 매우 높으므로 다음 과정 전에 미리 제거해 둔다. 또한 하나의 원자가 분자내에서 여러개의 삼각 고리 등을 형성하고 있는 불안정한 경우도 제거할 수 있다. carbene 의 경우는 화합물 자체는 불안정한 것으로 알려져 있으나 실제로 reaction path 에 등장하는 경우가 있기 때문에 중간체들 중 carbene 까지 포함하여 생각해보고 싶으면 포함을 해도 되고 만약 그렇지 않으면 걸러낼 수 있다.

마지막 과정에서 해야할 것은 postscreening process 가 끝난 후 생긴 중간체들과 반응물들 각 각의 vertex 으로 취급하는 vertex 와 edge 들로 이루어진 그래프를 그리고, 해당 그래프에서 Yen's algorithm 을 적용하여 반응물로부터 생성물로 진행되는 경로를 구해보는 것이다. Yen's algorithm 은

일반적으로 edge 과 vertices 로 이루어진 그래프에서 점과 점들 사이의 distance 가 정의 되어 있고 시작점과 끝점이 정해져 있을 때 distance 를 고려하여 시작점으로부터 끝점까지 이어지는 모든 경로 중에서 가장 짧은 경로를 제시하는 알고리즘이다. 여기서 고려해야 할 것은 edge 들이 만들어 지는 기준과 Yen's algorithm 에서 적용되는 distance 의 정의 방법이다. 먼저 distance 의 경우 이 코드에서는 bond dissociation energy 를 고려하였다. Hammond 가설에 따르면 transition state 는 이 전이상태와 에너지 차이가 적은 중간체와 가까운 구조를 갖는다. 따라서 bond dissociation energy 가 작게 형성되는 중간체들 사이의 edge 의 경우 작은 distance 를 할당 받고 그렇지 않은 경우는 상대적으로 큰 distance 를 할당 받는다. edge 들이 만들어지는 기준은 기본적으로 어떤 중간체에서 다른 중간체로 반응이 진행 될 때 끊어지는 결합과 생성되는 결합 수의 합이다. 만약 이 수가 제한 되어 있지 않고 반응물과 생성물의 free energy 차이가 작다면 생성물에서 바로 반응물이 생기는 반응 경로가 제시되어 질 것이다. 이러한 기준들을 가지고 Yen's algorithm 이 적용되었을 때 나오는 reaction path 들의 후보들의 개수는 사용자가 임의로 지정할 수 있다. 한편 Edge 생성과정에서 끊어지는 결합을 규명하면서 그 결합들에 대한 bond dissociation energy(BDE)는 사용자가 미리 작성해둔 BDE list 를 참조하여 가져온다. 만약 사용자가 만든 list 에서 빠진 경우가 있을 때 프로그램이 쓰는 로그파일에 더 필요한 BDE 값이 나타나고 그 값들에 대해 다시 list 에 더해주면 된다.



[그림 1] Ab initio nano reactor 를 이용해 예측한 Urey-Miller 반응 경로 예측 2

## 2. Urey-Miller Reaction and energy diagram

Urey-Miller reaction 은 초기 지구의 대기환경을 고려하여 글라이신이 형성되는 반응으로 실험적으로 gas phase 상태의 water(H<sub>2</sub>O), methane(CH<sub>4</sub>), ammonia(NH<sub>3</sub>), Hydrogen(H<sub>2</sub>), carbon monoxide(CO)들에 energy 를 가해 주었을 때 글라이신이 나오게 되는 반응이다.

ref[1]의 nature 논문은 ab initio MD nanoreactor 에서 가상의 piston 이 nano reactor 의 중심으로 분자들을 밀어주어 분자들의 충돌 횟수를 증가시켜 반응을 향상 시키는 작용을 하게 한다. 이 때 나온 결과는 다음과 같다.

### 3. 코드의 input(Fragment file, BDE data)

Fragment는 다음과 같다. [CH<sub>4</sub>], [CO], [C], [O], [NH], [OH],[H], [H], [H], [H]. NH<sub>3</sub>와 H<sub>2</sub>O의 경우 center 원자와 hydrogen을 모두 쪼개서 넣어야 더 화학적인 직관이 포함되지

않은 경우가 되겠지만 모두 쪼갤 경우 실행시간의 증가가 있기 때문에 그렇게 하지는 않았다. 코드를 실행 시킬 때 edge 생성시 결합에 직접적으로 관여하는 원자들이 동일 하더라도 해당 원자들에 대한 first nearest atom들이 서로 다르다면 bde 값들은 달라진다. Urey-Miller reaction에 대해 반응 경로를 알기 위해 필요한 BDE 값의 개수는 총 400개 였고, 이 값들을 구분하여 얻기 위해 second nearest atom들은 모두 수소 원자들로 통일된 분자들에 대해 Gaussian 09를 이용하여 DFT 계산을 하여 BDE list를 만들었다. 각 각의 molecule에 대한 functional은 B3LYP이고 6-31g(d,p) basis set 을 사용하였다. ref[1]에서 제시된 결과와 이 연구 결과에서 나온 결과로 energy diagram을 그려보면 어떤 반응 경로가 더 타당한지 판단할 수 있을 것이다.

### 결과 및 토의

#### 1. Urey-Miller reaction 을 코드에 적용시킨 결과

Fragmentation                      직후                      Deterministic

Sampling 방법을 이용하여 얻은 중간체들의 개수는 9143 개이고 postscreening process 를 거친 후 최종적으로 얻어진 중간체들의 개수는 293 개이다. 따라서 이 293 개의 중간체들간에 생성될 수 있는 총 Edge 들의 개수는  $293 \times 292 = 85,556$  개이지만 위에서 말한 두 중간체들 사이의 Edge 들을 생성하려는 기준을 정했을 때 그 중에서 17,176 개의 Edge 들만이 생성되었다. 이렇게 생성된 그래프를 이용하여 bond dissociation energy 에 따라 정의된 distance 를 Yen's algorithm 에 적용해서 얻어지는 path 들의 후보는 다음과 같다.

Path 1: R->6542->5085->5083->P

Path 2: R->6542->5085->5081->P

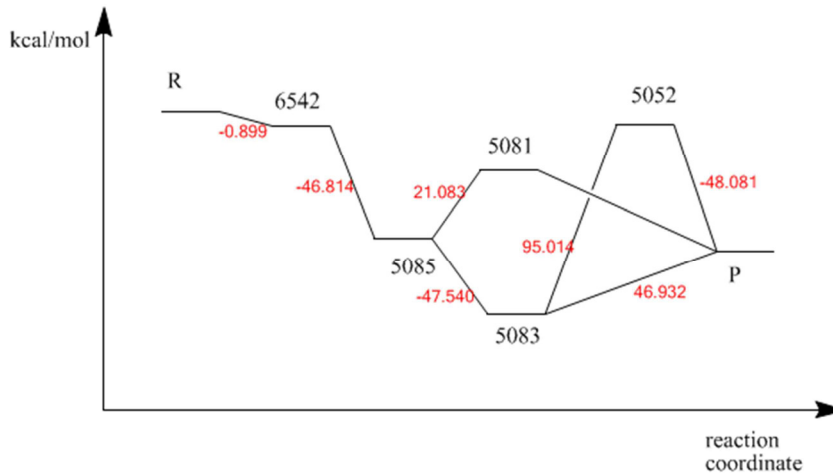
Path 3: R->6542->5085->5083->5052->P

이 때 얻어지는 energy profile을 제시된 path3까지에 대해 나타내어보면 다음과 같다.

가장 눈에 띄는 특징은 3개의 후보들 모두 반응을 R->6542->5085로 시작을 한다는 것이다. 6542의 경우 반응물에서의 일산화탄소의 탄소 원자가 질소와 결합을 형성하여 화합물을 생성하였다. 일산화탄소의 탄소원자가 음전하를 띠고 있으며 전자쌍을 가지고 있어, nucleophilic 하기는 하지만 암모니아의 질소원자 또한 nucleophilic 하기 때문에 결합하기가 실제로는 어려울 것으로 예상된다. 실제 논문 조사에서도 그러한 결합을 형성하는 경우는 찾기 어려웠고, DFT 계산을 통해서 나온 Energy profile을 보아도 reactant 와 중간체 6542의 에너지 차이는 0.899kcal/mol로 이러한 중간체가 생기기에는 어려울 것으로 판단된다. 중간체 5085 또한 중간체 6542에 일산화탄소가 결합하여 생성된 것이다. 이 경로가 모든 path 후보들에 대해 공통적으로 생성된 것은 다른 가능한 path 들에 비해 이 부분이 path의 전체적인 distance를 줄여주는 역할을 하기 때문이라고 생각한다.

Reactants		6542	5085	5083
CO	H <sub>2</sub> O			
NH <sub>3</sub>	H <sub>2</sub>	H <sub>2</sub> O CO H <sub>2</sub>	H <sub>2</sub> O	<sup>+</sup> H <sub>2</sub> N=CH-C(=O)O <sup>-</sup>
CH <sub>4</sub>			H <sub>2</sub>	H <sub>2</sub> O H <sub>2</sub>
5081		5052	Product	
O=C=O	H <sub>2</sub> O		H <sub>2</sub> O	
H <sub>2</sub> O			Glycine	
<sup>+</sup> H <sub>2</sub> N=CH <sup>-</sup>				

[그림 2] 반응 경로 후보(Path1~Path3)에 나타난 중간체들의 번호와 구조



[그림 3] Path1~Path3을 나타낸 energy profile

즉, 이 코드에서는 distance를 BDE를 가지고만 정의를 하였기 때문에 R->6542->5085 로 진행되는 동안 해리되는 결합이 없으므로 distance는 0가 되고 이 경로가 가장 선호되는 부분으로 나왔을 것이라고 생각된다. 따라서 BDE만으로 distance 를 정의를 한다면 Urey-Miller reaction 뿐만이 아니라 다른 반응들에 대해서도 이와 같이 결합이 해리되기 보다는 생성되기만 하는 반응이 선호되어질 수 있다. 코드를 개선하기 위해서 distance를 정의 할 때 BDE뿐만이 아니라 다른 variable 또한 추가적으로 고려해 볼 필요가 있다. 한편, 6542->5085의 경우에는 열역학적으로 선호되며 5085->5083의 경우에도 분자내에서 생기는 nucleophilic reaction으로 통해 C-C bond가 형성되고 C-O bond가 해리 되면서 나타날 수 있다. 그 다음 과정은 수소가 해리되어 glycine 을 형성하는 과정으로 흡열 반응이기는 하지만 일반적으로 energy barrier가 50 kcal/mol 이내일 경우 일어날 가능성이 있는 반응으로 보기 때문에 가능성이 있는 반응으로 생각 할 수 있다. path3의 경우에는 5083->5052로

진행되는 부분에서의 energy barrier가 95.01kcal/mol 이상이기 때문에 일반적으로 화학 반응에서 energy barrier가 50kcal/mol 이상이 되면 반응이 일어나기가 힘들다는 점을 감안하면 path1,2 보다는 덜 선호되는 path로 생각할 수 있다. 이 반응경로들을 다음의 nature에서 제시한 반응경로와 비교한다면 열역학적인 분석만으로는 해당 논문의 반응경로가 더 타당하다고 볼 수 있다. 비교 논문의 결과를 살펴보면 대부분의 중간체 사이의 반응의 Energy barrier가 50kcal/mol을 넘지 않고, free energy에 대해서도 대부분의 step이 발열 반응으로 나타나고 흡열 반응이 일어난다 하더라도 그 에너지 폭이 크게 높지 않아 동역학적으로나 열역학적으로도 이 연구에서 제시된 path보다는 favorable하다고 생각할 수 있다.

## 2. code에서 반응 경로를 결정하는 기준들의 결정 방식과 결과의 영향

(1) 끊어지고 생성되는 결합들의 수의 합

어떤 중간체에서 다음 중간체로 반응이 진행될 때 일반적으로 끊어지고 생기는 결합의 수의 합은 상한선이 있다. 그러한 상한선이 없다는 것은 반응물에서 생성물로도 반응이 바로 진행 될 수 있음을 의미하기도 한다. 이 계산에서 중간체들 사이의 반응 진행에 있어서 끊어지고 생기는 결합의 수의 합은 5개로 제한 시켰다. 명확하게 이러한 방식으로 결정된 구체적인 계산법은 없지만 먼저 nature 논문에서 제시된 path가 나오는지 알아보기엔 적어도 그 수가 4개 이상이 되어야한다. 또한 화학 반응에서 일반적으로 termolecular reaction 이상의 분자들의 수가 관여하는 반응은 잘 고려하지 않음을 생각하면 5개의 상한선이 적당하다고 생각하였다. 계산 결과를 고려하면 이러한 제한은 적절하였다. 반응 경로 상에서 메커니즘을 그려볼 때 분자의 결합이 동시에 여러 개(3개 이상) 끊어지면서 동시에 여러 개가 생성되는 반응은 나타나지 않았음을 알 수 있다.

(2) 분자내에서 가질 수 있는 가능한 고리의 개수

화학반응에서는 분자 내 고리형성 반응에 의해서 고리를 가진 화합물이 형성될 수 있음을 고려해야 하는데, 특히 삼각 고리의 경우 3차원적인 구조를 고려 했을 때 분자의 불안정성을 증가시키는 역할을 하기도 한다. 그러나 반응 중간체로서 cyclic epoxide가 종종 등장하고 반응에서 electrophilic한 화합물로서의 역할을 하게되므로 분자 내에서 최대 1개까지의 고리를 허용시키는 것이 타당하다고 생각했으나 이러한 조건을 추가했을 경우 가능한 중간체의 개수가 크게 늘어나고 이 때 Edge 생성을 하면서 프로그램의 실행 시간이  $N^2$  배로 상승하게

되므로 정해진 시간내에 계산 결과를 얻기 위해 고리화합물은 모두 걸러내기로 하였다. 결과적으로 intermediate에서는 고리화합물은 전혀 나오지 않았고, 이런 부분들이 이번에 제시된 path의 불완전함의 원인을 제공하고 있음을 알 수 있다.

(3) intermediate에 carbene의 존재 여부

Carbene은 불안정한 화합물이기 때문에 화학 반응에서 중간체로서 쉽게 생성되지는 않지만 비교 논문에서 제시된 결과의 중간체들에 carbene이 포함되어 있었으므로 비교를 위해 반응 중간체들 중에서 carbene이 형성될 수 있는 조건을 허용하였다. 그러나 계산 결과의 path 후보에서 carbene은 등장하지는 않았는데 이것은 이 화합물의 높은 에너지가 distance 값을 높인 결과라고 생각 할 수 있다.

(4) 화학반응이 일어날 수 있는 BDE값의 제한 여부

이 제한 조건은 Edge 생성여부에 대한 결정을 한다. 화학 반응에서 중간체들 사이에 energy barrier가 50kcal/mol 이상이 되게 되면 반응이 잘 일어나지 않는다<sup>3</sup>. 따라서 이 barrier를 정해두면 모든 Edge를 생성하지 않기 때문에 프로그램 실행의 시간이 줄어들게 된다. 다만 이번 계산에서는 이러한 제한을 두지 않았는데 비교 논문과 열역학적인 측면에서 path를 비교하기 위해서였고, Urey-Miller 실험에서 전기적 자극과 같은 에너지를 추가로 가해준다는 것을 고려하였다. 실제로 path3에서 energy barrier가 95.01kcal/mol이상인 경우가 등장함을 볼 수 있었다.

## 결론

이 연구의 목표는 화학적으로 유기화합물의

기본 구성 요소라고 할 수 있는 C, H, O, N을 포함한 중요한 Urey-Miller reaction path를 제안해보고 다른 논문에서 제시된 path와 비교했을 때 어떤 path가 더 화학적으로 타당지를 알아보기 위한 것이었다. 결과적으로는 이 연구에서 제시한 세 개의 path 후보들의 경우 첫번째 step에서 unfavorable한 반응이 일어났기 때문에 그 뒤에 일어나는 반응에 대해서 타당성을 논하기 어려운 부분이 있었다. 또한 nature에서 제시된 논문의 결과는 열역학적으로나 동역학적으로 가능성이 있는 경로를 제시하였기 때문에 아직 이 코드가 더 보완되어야 할 점이 있다는 것을 알 수 있었다. 한편, 이 코드의 장점은 비교 논문이나 다른 reaction path의 제시 방법과 달리 화학 반응에 따라 중간체들에 대한 DFT계산을 하지 않아도 된다는 것이다. 경로를 결정하는 요소가 BDE뿐이므로 중간체들 하나 하나에 대한 single point energy 계산을 할 필요가 없고 해당 BDE database로부터 필요한 값들을 가져오기만 하면 되기 때문이다. BDE의 계산을 할 때 DFT 계산을 해야 하지만 계속 database를 쌓아가다보면 추가적으로 해야하는 DFT계산의 양이 줄어 들게 된다. 보완 할 점은 distance의 정의 방식의 변화이다 BDE로만 결정할 경우에는 분자들이 서로 결합을 형성하는 경우가 선호되기 때문에 적절한 reaction path 를 얻기 어렵다. 다만 앞에서 말했듯이 이 코드는 중간체 각각에 대한 single point energy의 계산을 할 필요가 없다는 장점이 있으므로 중간체들에 대한 single point energy 계산을 하지 않는 범위 내에서 distance를 좀더 개선시킬 수 있는 variable과 식이 있는지 찾아볼 예정이다. 이 반응은 system의 크기가 상대적으로 작은 화학 반응에 속하기 때문에 fragment들의

개수도 작았으나 중간체(vertice)의 개수가 증가함에 따라 생성될 수 있는 Edge의 수도  $n^2$ 배로 증가하기 때문에 코드에서의 bottle neck을 더 개선시킬 필요가 있다. 한편, 이번 연구에서 시간의 제약상 fragment의 수를 제한시키거나 중간체들 중 고리가 형성된 화합물을 고려하지 못했다는 점이 계산 결과에서의 아쉬운 부분이었고, 이런 부분을 개선하여 연구를 더 진행시킬 것이다.

### 감사의 말. (ACKNOWLEDGEMENT)

본 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단 첨단 사이언스·교육 허브 개발 사업의 지원을 받아 수행된 연구임(No. NRF-2012-M3C1A6035357)

### 참고문헌

1. Kim, Y., Choi, S., & Kim, W. Y. (2014). Efficient basin-hopping sampling of reaction intermediates through molecular fragmentation and graph theory. *Journal of Chemical Theory and Computation*. doi:10.1021/ct500136x
2. Lee-Ping Wang, Alexey Titov, Robert McGibbon, Fang Liu, Vijay S. Pande, T. J. M. (2014). Discovering chemistry with an ab initio nanoreactor. Submitted to *Nature Chemistry*, 6(December), 2-7. doi:10.1038/nchem.2099
3. Ufimtsev, I. S. & Martinez, T. J. Quantum chemistry on graphical processing units.3. Analytical energy gradients, geometry optimization and first principles molecular dynamics. *J.Chem.Theory Comput.* 5, 2619-2628(2009)