

일반연구논문

RNA 타이 클럽의 유전암호 해독 연구: 다학제 협동연구와 공동의 연구의제에 관한 고찰¹⁾

김봉국*

■ 유익한 논평으로 논문을 다듬는 데 도움을 준 세 분의 심사위원회께 감사드린다.

* 서울대학교 과학사 및 과학철학 협동과정 박사과정

전자우편: nomadian94@hotmail.com

1953년에 이론물리학자 조지 가모프는 “DNA의 염기 서열에 의해 단백질의 아미노산 서열이 암호화된다”는 가설로 단백질 합성 현상을 설명했고, “다이아몬드 코드”라는 핵산-단백질 정보전달 모형을 제안했다. 왓슨, 크릭, 브레너, 스텐트 같은 당대의 생물학자들은 이런 대담한 제안에 관심을 보이며 가모프와 토론했고, 이에 가모프는 이들 간의 의견교환을 활성화하기 위해 RNA 타이 클럽이라는 비공식 연구자 모임을 결성했다. 이후 생물학자들뿐만 아니라 물리학자, 수학자, 컴퓨터엔지니어들이 RNA 타이 클럽에 동참했고, 이들의 다학제적 유전암호 해독 연구는 1950년대 후반까지 활발히 이루어졌다. 본고는 RNA 타이 클럽의 형성, 성장, 와해의 과정을 살피면서 다음과 같은 논제를 다루려 한다. 첫째, 정통 생물학과 거리가 먼 ‘문자열 간 번역원리를 찾는 가모프의 수학적 접근’은 어떻게 생물학자들의 공동의 연구의제로 채택될 수 있었을까? 둘째, RNA 타이 클럽의 연구의제는 어떻게 다양한 학제의 연구주제들로 풍부하게 확장될 수 있었을까? 셋째, RNA 타이 클럽의 쇠락과 와해를 가져온 요인은 무엇이 있었을까? 이런 분석을 통해, 본고는 타이 클럽 연구자들의 근본 가정인 “아미노산 서열에 배열 패턴이 존재한다”는 가정이 다양한 학제적 접근방식을 매개하는 연결고리 역할을 했고, 또 이를 통해 당대 생물학의 실험 데이터를 활용할 수 있게 되면서 이들의 번역코드 연구는 유의미한 생물학 연구방법으로 자리 잡을 수 있게 되었음 주장할 것이다. 아울러 위의 논의의 연장선상에서 타이 클럽의 와해 요인을 해명함으로써, 다양한 학제의 연구자들을 성공적으로 매개할 생산적 연구의제가 갖춰야 할 요건은 무엇인지 고찰해 볼 것이다.

주제어 | RNA 타이 클럽, 조지 가모프, 다학제 연구, 연구의제, 유전암호

1. 서론

1961년, 첫 번째 유전암호가 니렌버그(M. W. Nirenberg)와 마테이(H. Matthaei)의 실험에 의해 밝혀졌다. 이들은 무세포계 시험관에 우라실(uracil) 염기뿐만 이뤄진 인공 RNA를 첨가해 페닐알라닌(phenylalanine)만으로 연결된 단백질을 합성해냄으로써 삼중코돈 UUU가 페닐알라닌 아미노산으로 번역된다는 사실을 규명했다(Nirenberg & Matthaei, 1961). 이를 계기로 나머지 63개 유전암호를 마저 규명하려는 실험이 경쟁적으로 잇달았고, 1966년에 이르러 64개 코돈과 20개 아미노산의 대응관계가 모두 밝혀졌다.

1966년에 열린 콜드 스프링 하버 심포지엄은 유전암호 해독의 완수를 자축하는 자리가 되었다. 300여명의 연구자들이 참석한 가운데 그간의 성과와 향후 연구 전망에 대한 논의가 이뤄졌다. 심포지엄의 첫 발표를 맡은 프랜시스 크릭(F. H. C. Crick)은 “유전암호-어제, 오늘, 내일”이라는 강연에서 유전암호 연구의 역사를 개괄했는데, 생화학자들의 해독 실험이 본격화되기 전인 1950년대의 연구 경향을 소개하면서 이런 말을 했다.

1953년에 발표된 DNA 구조에 관한 이론은 암호화(coding) 개념의 등장을 조장했다. 많은 사람들이 DNA 구조의 단순성에 매료되었는데, 우주론학자 조지 가모프도 그중 하나였다. [...] 당시 가모프는 RNA 타이 클럽(RNA Tie Club)이라는 기묘한 단체를 결성했다. 이 단체는 암호화 문제에 관심을 보인 사람들로 구성되었고, 회원 수는 (아미노산당 한 명씩) 스무 명으로 제한되었다. [...] 각 회원은 가모프의 디자인대로 로스앤젤레스 의류점에서 특별 제작한 넥타

이와 각자의 아미노산 약자가 새겨진 벡타이핀을 받았다(Crick, 1966: 4).

1953년에 핵물리학자 가모프(George Gamow)는 단백질 합성 현상을 ‘번역에 의한 정보 전달 문제’로 간주하는 당시로서는 매우 독특한 견해를 제시했다. 크릭, 왓슨(J. D. Watson), 브레너(S. Brenner), 델브뤼크(M. Delbrück)을 비롯한 생물학자들은 핵물리학자의 이 같은 대담한 제안에 관심을 보였고, 이에 가모프는 이들 간의 의견 교환을 활성화하기 위해 RNA 타이 클럽이라는 비공식 연구자 모임을 결성했다. 이후 생물학자들뿐만 아니라 물리학자, 수학자, 컴퓨터엔지니어 등 다양한 학제에 속한 연구자들이 모임에 동참했고, 분과의 경계를 넘나드는 이들의 협동연구는 1950년대 말까지 이루어졌다.

본고는 타이 클럽의 연구 활동을 분석하면서, 이 독특한 모임 내에서 향후 유전암호가 해독되는 데 기여한 여러 개념과 분석틀이 고안되었음을 보일 것이다. 또한 RNA 타이 클럽의 결성, 성장, 와해의 과정을 다루면서 다음과 같은 논제에 주목할 것이다. 결성의 과정을 다루는 2절에서는 정통 생물학과는 거리가 먼 ‘문자열 간 번역원리를 찾는 가모프의 수학적 접근’이 몇몇 생물학자의 관심을 끌며 ‘공동의 연구의제’로 채택된 과정을 살펴볼 것이다. 성장의 국면을 다루는 3절에서는 RNA 타이 클럽의 연구의제가 다양한 학제의 관련 연구주체들로 풍성하게 확장될 수 있었던 이유를 분석할 것이다. 특히 타이 클럽의 연구의제가 정교해지는 과정에서 가미된 한 가지 가정(“코드는 중첩된다”)이 다양한 학제적 접근을 매개하는 연결고리 역할을 했음을 보일 것이다. 4절에서는 3절 논의의 연장선상에서 모임 와해의 원인을 해명할 것이다. “코드는 중첩된다”는 가정을 더 이상 고수할 수 없게 되었을 때, 즉 상이한 학제적 접근들을 매개함으로써 공동의 연구의제에 생산적으로 기여할 수 있게 해준 기반이 무너졌을 때, 타이 클럽의 연구 활동이 얼마나 극적인 변화를 겪게 되었는지 고찰할 것이다.

이상의 분석은 학제간 협력과 융합을 다양한 층위에서 다뤄온 STS의

연구 동향에 일정 부분 보탬이 될 수 있을 것이다. 이런 STS 연구에 따르면, 학문의 경계는 종종 새로운 아이디어의 출처가 되며, 한 학제 내에서는 해결되지 않던 난제의 실마리가 여러 학제를 넘나드는 가운데 발견되기도 한다(홍성욱, 2012; Simonton, 1984, 2004). 융합이 야기하는 혁신 효과나 문제해결력에 대한 이 같은 인식은 다학제 연구(multidisciplinary research), 학제간 연구(interdisciplinary research), 초학제 연구(transdisciplinary research)에 대한 사회적, 제도적 차원의 지원을 정당화하는 근거로 종종 사용된다(Rossini & Porter, 1984). 또 이 같은 사회적 지원이 이뤄지는 가운데, 진짜 융합과 가짜 융합을 구분하고 실질적 성과를 산출할 가망이 있는 생산적 융합을 사전에 선별해야 할 필요성도 대두하는 바, 이에 대한 나름의 기준을 마련하는 방안은 이미 STS의 주요한 주제가 되었다(Klein, 1996; Klein et al., 2001). 이런 맥락에서 학제간 융합을 성사시키는 데 필요한 외재적, 제도적 조건은 무엇인지 고찰이 이뤄지고 있으며(Cummings & Kiesler, 2005), 또 성공적인 다학제 연구의 사례를 상세히 분석함으로써 상이한 분야 간의 생산적 융합이 어떤 내재적 조건들이 갖춰졌을 때 가능한지 따져보는 연구들도 이뤄지고 있다(이상욱, 2015; Jeffrey, 2003; Park, 2010).

RNA 타이 클럽은 제도적 차원의 지원이 결여된 비공식 연구자 모임이라는 점에서 STS에서 관심을 기울여온 학제간 연구의 전형과는 거리가 멀다. 그럼에도 연구의제의 특성과 그것이 상이한 학제의 학자들의 상호작용에 미치는 구체적 효과에 천착하는 본고의 분석은 시사하는 바가 있다. 다학제적 연구 그룹의 성패에서 생산적인 연구대상과 생산적인 연구의제를 설정하는 작업은 갈수록 중요해지고 있거니와(Saari & Miettinen, 2001), 이것 없이는 존속조차 불가능했던 타이 클럽 같은 비공식 모임에 대한 분석은 상이한 학제의 연구자들을 결속시킬 수 있는 연구의제가 어떤 요건을 갖춰야 하는지 생생하게 보여줄 수 있을 것이다.

2. 새로운 연구의제의 등장: RNA 타이 클럽의 결성

단백질의 구조와 기능은 1930-40년대 생물학의 가장 중요한 주제였다. 단백질은 세포의 구성물질일 뿐만 아니라 화학반응의 촉매 역할을 하는 분자였다. 많은 생화학자들이 이 다재다능한 분자의 화학 조성을 연구했고, 그 결과 단백질이 여러 아미노산들이 선형적으로 결합된 폴리펩티드 사슬로 이루어져 있음이 밝혀졌다. 또한 분자생물학자들은 각종 단백질의 기능이 폴리펩티드 사슬이 접혀져 형성되는 3차원 구조를 통해 발현된다고 보면서 그 구조를 파악하기 위한 결정학 실험에 열을 올렸다. 그렇다면 이런 단백질은 어떻게 만들어지는가? 1941년, 스탠퍼드 대학의 비들(G. Beadle)과 테이텀(E. Tatum)은 단백질의 화학적 특이성이 유전자에 의해 결정된다는 점을 붉은빵곰팡이 실험으로 보여줬다(Beadle & Tatum, 1941). 하지만 유전자에 대한 이해가 불완전했던 당시에 유전자에 의한 단백질 합성이 어떤 과정으로 일어나는지 밝혀내기란 불가능했다(Keller, 2000: 51-52).

이러한 점에서 DNA 이중나선 구조는 중요한 전환점이었다. 왓슨과 크릭은 DNA의 상보적 구조를 통해 유전자의 세대 간 안정성과 복제 메커니즘이 해명되었다고 말했지만, 사실 이중나선 구조는 단백질 합성에 대한 새로운 이해를 가능하게 한 점에서도 의의가 있었다. DNA 분자 구조에 유전정보가 저장된다고 해석됨에 따라, 유전자에 의한 단백질 합성은 핵산 분자와 단백질 분자의 구조적 관계를 규명하는 문제로 변형될 수 있었다. 생체 내 핵심 기능을 하는 두 분자의 특별한 관계는 향후 생물학이 규명해야 할 연구과제로 부상했는데, 흥미롭게도 이 문제를 처음 다룬 이는 생물학자가 아니라 핵물리학자 조지 가모프였다.

러시아에서 태어난 가모프는 레닌그라드대 물리학부를 졸업한 후 핵물리학자로서 경력을 쌓았고, 1934년 미국으로 망명해 조지워싱턴대 물리학 교

수가 되었다. 젊은 시절 그는 터널효과 개념에 입각한 알파붕괴이론을 제안해 학계의 주목을 받았고, 1936년에는 베타붕괴이론을 발표하며 핵물리학 권위자가 되었다. 하지만 그의 활동이 물리학에 국한되지는 않았다. 과학대중화에도 앞장섰던 그는 대중서 저술에서도 남다른 재능을 보였다. 또한 그는 미해군 고성능폭약부 고문과 군정보국 암호해독 자문을 맡았고, 로스알라모스의 수폭 개발에 참여하는 등 군사연구 방면에서도 두드러진 활동을 보였다.

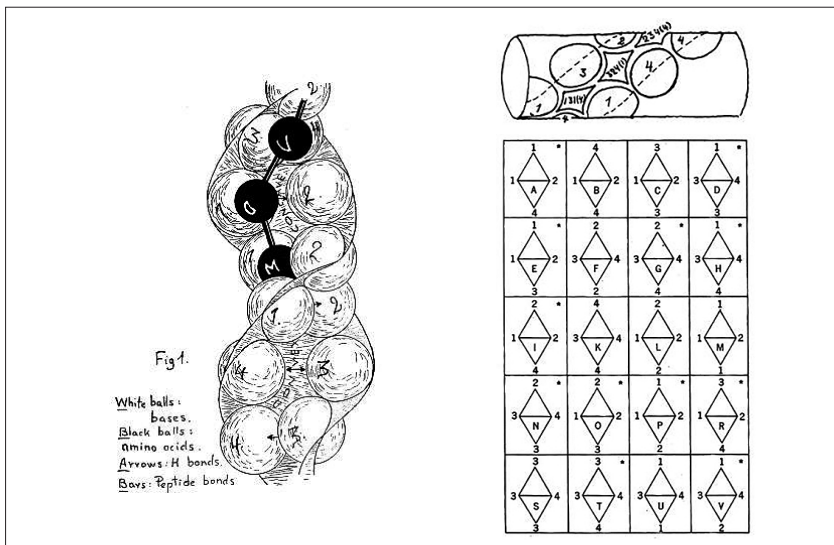
이중나선 모형은 생물학에도 관심이 많았던 가모프가 유전암호 연구에 빠지는 계기가 되었다. 특히 가모프는 왓슨과 크릭의 논문(Watson & Crick, 1953b)을 읽은 후, 원저자들조차 간과한 이중나선의 함의를 자신이 발견했다고 생각했다. 1953년 7월, 왓슨과 크릭에게 보낸 편지에서 가모프는 “당-인산염 골격에 부착된 염기들이 유전정보를 전달하는 부호”라면, 개별 생명체의 유전적 특성을 “4종의 숫자로 쓴 큰 수”로 표현할 수 있다고 주장했다. DNA 사슬에 배열된 4종의 염기(아데닌, 구아닌, 시토신, 티민)를 1, 2, 3, 4로 표기해 “4비트의 숫자열”로 나타낼 수 있을 터, 이 “짐승의 숫자(number of the beast)”를 정수론과 조합론을 활용해 분석하면 생물학도 정밀과학의 반열에 오를 것이라 기대했다.¹⁾ 또한 가모프는 자신의 첫 생물학 논문에서 복잡한 3차원 구조의 단백질도 결국 아미노산들이 일직선으로 연결된 폴리펩티드 사슬일 뿐이며 그 구성 아미노산이 20종이므로 “단백질은 20가지 알파벳으로 쓴 긴 ‘단어’로 볼 수 있다”고 주장했다.²⁾ DNA와 단백질은 화학 조성, 성질, 구조가 전혀 다른 이질적 분자였지만, 가모프의 단순화된 도식에서 두 분자는 동일한

1) 왓슨의 책에 첨부된 서신을 참고했다(Watson, 2001: 32-33, 262-265).

2) 이 논문에서 가모프는 단백질의 구성 아미노산이 20종이라고 언급만 했고, 두 번째 논문에서는 20가지 아미노산 목록을 제시했지만 어떻게 그런 결론에 도달했는지 그 근거는 밝히지 않았다(Gamow, 1954b: 4).

방식으로 다뤄졌다. 이런 단순화 덕분에 가모프는 유전자에 의한 단백질 합성을 ‘핵산의 화학언어’와 ‘단백질의 화학언어’ 간의 번역의 문제, 다시 말해 “DNA의 염기 서열에 의해서 아미노산의 결합순서가 결정되는 현상”으로 개념화할 수 있었다(Gamow, 1954a).

가모프는 핵산-단백질 간 번역 메커니즘을 고안하는 작업을 ‘4종의 염기’와 ‘핵산과 단백질의 분자 구조’를 참조해 ‘20종의 아미노산’을 도출해야 하는 퍼즐로 취급했다. 일단 그는 이중나선상에 연속적으로 형성되는 마름모꼴 공동들이 구조적, 화학적 정합성에 따라 개별 아미노산들을 수용하게 되며, 이렇게 배열된 아미노산들이 결합해 폴리펩티드 사슬을 형성할 때까지 붙들



〈그림 1〉 다이아몬드 코드

마름모 꼭짓점의 1, 2, 3, 4는 아데닌, 티민, 구아닌, 시토신을 나타내며, 가로선은 (서로 반대편 사슬에 놓이는) 염기의 상보적 결합을 표현한 것으로 1-2(아데닌-티민)와 3-4(구아닌-시토신)의 결합만이 가능하다는 점을 나타내고 있다. 위의 다이어그램에 근거해서 가모프는 마름모꼴 공동의 종류가 20개라고 주장했다(Gamow, 1954b: 5).

어둔다고 가정했다. 이런 가모프의 설명은 마름모꼴 공동들의 간격과 폴리펩티드 사슬상의 아미노산 간 거리가 $3.4\text{\AA} \sim 3.6\text{\AA}$ 로 거의 같다는 점과 잘 부합했다. 그렇다면 아미노산과 열쇠-자물쇠 관계인 마름모꼴 공동은 몇 종류인가? 이런 마름모꼴 공동은 꼭짓점 염기의 배치 양상에 따라 분류될 수 있고, 또 가로 대각선상의 염기가 상보적 결합을 해야 한다는 점까지 고려하면, 공동의 종류는 아미노산의 개수와 정확히 일치하는 20개였다(그림1). 즉, 가모프는 DNA의 구조적 특징을 참조해서 4비트의 염기 서열을 20비트의 아미노산 서열로 번역해주는 다이아몬드 코드(diamond code)를 도출했던 것이다(Gamow, 1954a; Gamow, 1954b).

이런 착상을 담은 논문을 1953년 10월 『네이처』에 제출한 직후, 가모프는 전문가들에게 평가를 부탁했다. 생물학자들의 반응은 미온적이었는데, 단백질 연구 최고 권위자였던 폴링은 염기 서열이 아미노산 서열을 결정한다는 가설은 흥미롭지만, DNA의 공동이 아미노산의 수용체라는 주장은 세부 분자 구조를 고려하지 않은 지나친 단순화라고 충고했다.³⁾ 폴링을 비롯한 당대 주류 생물학자들은 가모프의 논문을 진지한 연구로 여기지 않았는데, 가모프가 사용한 개념과 접근법은 생물학자들에게 친숙한 것이 아니었고, 더욱이 이들에게 가모프는 생물학을 피상적으로만 알고 있는 외부인일 뿐이었다.

하지만 가모프의 주장을 물리학자의 지나친 상상력과 무분별한 단순화라고 무시한 생물학자들만 있었던 것은 아니었다. 1953년 말, 크릭을 방문한 가모프는 다이아몬드 코드의 세부 특징에 대해 토론했고, 이 때의 논의를 바탕으로 첫 논문의 내용을 보완할 수 있었다(Gamow, 1954b). 또한 가모프는 옛 친구 텔브릭의 도움으로 칼텍의 유전학자들과 번역코드에 관해 토론하는

3) <http://scarc.library.oregonstate.edu/coll/pauling/dna/corr/sci9.001.43-gamow-lp-19531022.html>. From G. Gamow to L. Pauling, October 22, 1953.

기회를 가졌고, 이를 계기로 델브릭 연구실의 왓슨, 폴링 연구실의 오겔(L. E. Orgel), 스탠리 바이러스 연구소의 스텐트(G. S. Stent), 국립보건원의 생화학자 리치(A. Rich)와 소중한 인연을 맺었다(Watson, 2001: 61-62).

이렇듯 가모프와 코드에 관해 의견을 나누던 연구자들은 RNA 타이 클럽의 주축 멤버가 되었다. 1954년 4월, 가모프, 왓슨, 오겔, 스텐트는 체계적 연구와 원활한 의견교환을 위해 정기적 모임이 필요하다고 의견을 모았고, 이에 RNA 타이 클럽이라 명명한 모임을 결성했다. 가모프는 위 멤버 외에도 리치, 크릭, 육군병참연구소의 이차스(M. Yčas)를 타이 클럽에 동참시켰고, 최종적으로 아미노산의 종류에 맞춰 회원 20명을 확보한다는 목표를 세웠다(Judson, 1979: 67-69). 타이 클럽 생물학자들은 생물학에 무지한 가모프가 초보적 실수를 범했지만, 이런 무지 때문에 가능했던 대담한 문제제기를 공동의 연구의제로 채택했다. 이들은 “염기 서열이 아미노산의 서열을 결정한다”는 가모프의 전제가 단백질 합성을 규명하는 데 유용하다고 보았고, 그의 미숙한 개념과 분석에 대해서도 가능한 접근으로 평가하며 보완하려 했다.

이 과정에서 새로운 연구주제들이 부상하기도 했는데, 아미노산의 종류에 관한 고찰이 대표적이다. 가모프의 원고를 받았을 때 왓슨과 크릭은 단백질을 구성하는 아미노산이 20종류라는 대목에서 의아했는데, 단백질 분석 실험으로 더 많은 아미노산이 이미 발견된 사실에 무지한 가모프가 착각했다고 생각했다. 새로운 아미노산이 속속 발견되던 상황에서 그 종류를 한정하려는 시도는 부적절해 보였으나, 개수를 확정하지 않고서는 가모프식의 번역코드 연구는 불가능했기에 이내 왓슨과 크릭은 이 문제에 대한 색다른 고찰을 시작했다(Crick, 1988: 91-92). 우선 그들은 전체 아미노산 중에서 일부만이 실제 합성에 관여한다는 것을 보면서, 단백질 합성에 사용되는 ‘표준 아미노산’과 이미 합성된 단백질이 변성해 생겨나는 ‘변종 아미노산’을 가정했다. 그리고 전체 아미노산 목록을 조사하여 대부분의 단백질에 포함되어 있으면 표준

아미노산으로, 특정 단백질에서만 선별적으로 나타나면 변종 아미노산으로 분류했다. 이를 통해 그들은 비록 가모프의 목록과는 달랐지만 단백질 합성에 이용되는 아미노산은 실제로 20종밖에 없다는 결론에 도달했다(Crick, 1963). 1957년 실험생물학회 심포지엄에서 크릭은 당시 어떤 생화학 교과서에도 표준 아미노산에 관한 고찰은 없으면서 “생화학자들은 마법의 20[표준 아미노산]과 나머지에 대한 구분 없이 가능한 많은 아미노산으로 목록을 채우는 일에만 열중했다”고 나무랐다. 하지만 새로운 아미노산을 발견하는 실험 자체에 열중했던 생화학자들에게 아미노산이 몇 종인지 규명하는 일은 중요한 문제가 될 수 없었다. 이와 달리 번역코드 연구에 빠져든 크릭은 당대의 생화학자들과는 다른 관점에서 이 문제를 해결할 수 있었다(Crick, 1958: 140).

RNA 분자모형도 새로 부상한 연구주제였다. 애초에 가모프는 아미노산이 DNA상에 배열된 후 그대로 중합된다고 주장했지만, 이는 단백질 합성 장소가 DNA가 전무한 세포질의 마이크로솜이라는 실험결과와 모순되었다. 타이 클럽 생물학자들은 DNA가 단백질 합성에 관여하고 마이크로솜에는 RNA가 풍부하다는 점을 고려할 때, DNA에 의해서 RNA가 만들어진 후 RNA를 주형으로 단백질이 합성된다고 보는 것이 더 적절하다고 지적했다(Crick, 1988: 93; Watson, 2001: 53-54). 이후 타이 클럽 연구자들은 RNA 분자 구조를 시급한 과제로 간주했는데, 다이아몬드 코드가 DNA 이중나선에 근거했듯이 번역코드를 고안하려면 정확한 RNA 모형이 필요했기 때문이다. 이에 왓슨과 리치는 회절사진을 분석하여 RNA는 DNA와 달리 단일나선 구조를 띄고 있다는 점을 규명했다. 비록 양질의 사진을 구하지 못해 명확한 모형까지는 제시하지 못했으나, 그들은 유전암호 해독 연구가 향후 RNA 구조 연구를 통해 보완되어야 한다고 강조했다(Rich & Watson, 1954: 760).

이렇듯 번역코드 연구는 관련 주제들로 점차 확장되었는데, 이 과정에서 더 많은 연구자들이 타이 클럽에 동참하게 되었다. 칼텍의 유전학자 델브

릭과 왓슨, 이론화학자 오겔, 케임브리지 분자생물학연구소의 크릭과 브레너, 육균병참연구소의 생화학자 이차스, 콜롬비아 대학의 생화학자 샤가프, 버클리 대학의 바이러스학자 스텐트 같은 생물학자들이 가모프와 유전암호 해독 연구를 함께 했다. 생물학자 외에도 칼텍의 물리학자 파인만(R. P. Feynman)과 버클리 대학의 물리학자 텔러, 존스홉킨스 대학의 수학자 레들리(R. S. Ledley), 로스알라모스의 컴퓨터엔지니어 메트로폴리스(N. Metropolis)도 타이 클럽에 참여했다. 번역코드에 흥미를 느끼며 타이 클럽에 동참한 연구자들은 자신의 전공 지식이 유전암호 해독에 기여하는 부분이 있다고 보았다. 학문적 배경이 서로 달랐던 연구자들은 상대방이 제시한 개념에 익숙하지 못한 경우가 많았고, 구체적인 내용을 이해하는 데 어려움을 겪었으며, 서로의 이질적인 접근 방식에 회의적인 태도를 보이기도 했다. 하지만 이들은 이와 같은 의견 교환과정이 필요하다고 생각했는데, 이 과정에서 ‘핵산-단백질 정보전달 문제’를 해결하는데 도움이 되는 새로운 아이디어들이 제기되었기 때문이다.

3. ‘퍼즐풀이’에서 ‘서열해독’으로: 중첩코드 연구, 1953-1956

RNA 타이 클럽의 결속력은 강하지 않았다. 상이한 대학과 연구소에 소속한 채 각자의 연구에 몰두했던 연구자들은 서신이나 토론용 원고를 회람하며 의견을 교환했다. 구성원의 결속력이 유지되고 개별 연구성과들이 교류되는 데는 유전암호 해독에 열의를 보였던 가모프의 노력이 결정적이었다. 클럽의 “합성장치”라 불렸던 가모프는 연구 방향을 제시하고 의견들을 종합하며 새로운 문제를 제기하는 등 클럽 연구 활동의 중심에 있었다. 비공식 모임으로서 타이 클럽은 운영상의 어려움을 겪었지만, 그럼에도 가모프를 중심으로 활발한 의견 교환이 이루어졌다. 이 과정에서 타이 클럽의 유전암호 해독 연구는

점차 체계를 갖추었고 관련 연구주제들로 풍부하게 확장되었다.

1) 번역코드를 검증하라: 중첩코드의 기호간 상관관계

다이아몬드 코드를 고안할 당시 가모프는 인접한 공동들이 염기를 두 개씩 공유한다고 가정했다. 아미노산의 수용체인 공동이 중첩된다고 보면 “아미노산 간 거리와 염기 간 거리가 거의 같다”는 점이 잘 설명되었기 때문이다. 하지만 크릭을 통해 중첩코드(overlapping code) 가정의 더 중요한 효과가 드러났다. 크릭은 인접한 코돈(공동)들이 염기를 공유하므로 20개의 코돈 간에 상관관계가 존재한다고 지적했다. 가령 세 개의 아데닌으로 이뤄진 코돈 뒤에는 아무 것이나 올 수 있는 게 아니라 두 개 이상의 아데닌을 포함한 코돈이 와야만 했다(Crick, 1966: 67). 가모프는 이러한 제약을 “기호간 상관관계(intersymbol correlation)”라고 불렀고 서로 연결 가능한 코돈을 조사하여 결합규칙 표를 작성했다(표11). 코드가 중첩된다는 가정은, 염기 서열은 무작위로 배열되더라도 그것을 통해 재현되는 코돈은 특정 조건을 만족시키는 형태로만 배열된다는 점을 시사했다(Gamow, Rich & Yčas, 1956: 39-41).

	B C A C D D A B A B D C
Overlapping code	<pre> B C A C A C A C D C D D </pre>
Partial overlapping code	<pre> B C A A C D D D A A B A </pre>
Nonoverlapping code	<pre> B C A C D D A B A B D C </pre>

〈그림 2〉 중첩/부분중첩/비중첩코드

중첩코드와 부분중첩코드는 인접코드가 각각 두 개와 한 개의 염기를 공유하지만, 비중첩코드는 코드 간 염기의 공유가 없다(Grick, Griffith & Orgel, 1957: 417).

Nearest Neighbor Combinations in Diamond Code

A, E, I, O
 associates with
A, D, E, F, G, H, I, L, M, N, O, P, U, V

D, G, H, N
 associates with
A, B, C, D, E, G, H, I, K, N, O, R, S, T

L, M, P
 associates with
A, E, I, O, L, M, P

K, S, T
 associates with
K, G, H, K, N, S, T

B, C, R
 associates with
D, F, G, H, N, U, V

F, U, V
 associates with
A, B, C, E, I, O, R

〈표 1〉 다이아몬드 코드의 결합규칙 표

가모프는 [그림1]로부터 위의 ‘코드 간 결합규칙’을 이끌어냈다(20개 알파벳은 [그림1]의 20개 코드를 나타낸다). 가령, 다이아몬드A는 D와 연결될 수 있지만 구성 염기상 K와는 연결될 수 없다(Gamow, 1954b: 8). 가모프가 다이아몬드 코드의 검증에 활용한 인슐린A, 인슐린B, 코르티코트로핀의 아미노산 서열은 다음과 같다.

Gly-Ileu-Val-Glu-Glun-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Tyr-Glun-Leu-Glu-Asn-Tyr-Cys-Asn(인슐린A);Phe-Val-Asn-Glun-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Tyr-Pro-Lys-Ala(인슐린B);Ser-Ser-Met-Glu-His-Phe-Arg-Try-Gly-Lys-Pro-Val-Gly-Lys-Arg-Arg-Pro-Val-Tyr-Pro-Asp-Gly-Ala-Glu-Asp-Glun-Leu-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe(코르티코트로핀).

중첩코드의 이 같은 특성은 번역코드의 타당성을 실제 아미노산 서열을 이용해 검토할 수 있게 해줬다. 코돈의 연결에 제약이 있다면 그 대응 아미노산들도 규칙을 만족시키는 형태로만 배열돼야 했다. 이에 크릭은 인슐린과 코르티코트로핀의 서열이 다이아몬드 코드의 배열규칙을 따르는지 분석했다. 하지만 실제 서열에서 8종 이상의 아미노산과 인접 가능한 아미노산은 10종이었던 반면, 다이아몬드 코드에서 8종 이상의 코돈과 인접 가능한 코돈은 [표1]처럼 8개에 불과했고, 따라서 다이아몬드 코드로는 인슐린과 코르티코

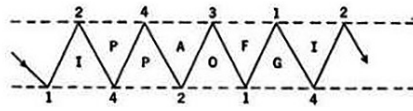
트로핀의 서열을 설명할 수 없었다(Crick, 1966: 4; Watson, 2001: 91, 119).⁴⁾

비록 다이아몬드 코드는 폐기되었지만, 이론과 실험의 연결고리가 마련되었다는 점에서 기호간 상관관계에 대한 인식은 중요한 전환점이었다. 1950년대에 활성화된 생화학자들의 서열분석 실험은 아미노산 조성을 파악하여 단백질의 기능을 이해하기 위한 것이었지만, 타이 클럽 연구자들은 이를 유전자의 통제 아래 진행되는 단백질 합성의 양상을 보여주는 자료로 취급했다. 염기 서열과 아미노산 서열 간 번역 원리의 흔적이 서열 데이터상에 남아 있기에, 그것과 번역코드의 기호간 상관관계를 체계적으로 분석한다면 이론적으로 제시된 정보전달 모형의 타당성을 검증할 수 있었기 때문이다.

2) 새로운 번역코드들

다이아몬드 코드의 대안이 될 만한 정보전달 모형을 모색했던 가모프는 이내 RNA를 주형분자로 가정하는 삼각형 코드(triangular code)를 제안했다. 단일 나선상의 염기 서열에 의해 형성되는 삼각형의 공동이 아미노산의 수용체 역할을 한다고 보았는데, 세 꼭짓점의 염기 배열에 따라 달라지는 삼각형의 개수가 아미노산의 종류와 일치하는 20개이므로 번역코드의 요건을 갖추었다(Gamow, Rich & Yčas, 1956: 48-49). 가모프는 아미노산의 배치 간격에 따라 “조밀한 삼각형 코드”와 “성긴 삼각형 코드”라는 두 종류의 코드를 고려했다(〔그림3〕). 인접한 코드가 두 개의 염기를 공유하는 조밀한 코드는 한 개의 염기를 공유하는 성긴 코드에 비해 배열상의 제약이 심했는데, 가모프는 각 코드의 기호간 상관관계를 분석함으로써 그것의 타당성을 판단할 수 있는 기준도 마련했다(Gamow, Rich & Yčas, 1956: 50-51).

4) 훗날 크릭은 이 방법을 자세히 소개하면서 “이미 죽은 말에 채찍질하려는 게 아니라 번역코드를 검증하는 간단한 방법을 보여주려고 했다”고 말했다.



〈그림 3〉 삼각형 코드

꼭지점에 4종의 염기가 배치되는 삼각형의 종류는 $20 ({}_4H_3=20)$ 개가 된다. 배열방식에 따라 두 종류의 코드가 고려되었는데, 조밀한 코드는 $-I-P-P-A-$ 의 형태로 아미노산이 배열되고 성긴 코드는 $-I-P-O-G-$ 의 형태로 배열된다. 삼각형 코드에서도 '기호간 상관관계'에 따른 결합규칙이 존재하는데, 조밀한 코드의 경우, 20개의 삼각형 중에서 4개는 10종류의 다른 삼각형과 연결될 수 있었고 12개는 7종류와, 그리고 나머지 4개는 4종류의 다른 삼각형과 연결될 수 있었다. 반면 성긴 코드의 경우, 20개의 삼각형 중에서 4개는 19종류의 삼각형과 연결될 수 있었고 12개는 16종류와 그리고 4개는 10종류의 삼각형과 연결될 수 있었다(Gamow, Rich & Yčas 1956: 48-49).

오겔과 파인만은 실험으로 축적된 아미노산 서열의 배열패턴을 직접 참조한 메이저-마이너 코드(major-minor code)를 제안했는데, 마법의 20을 도출하는 수학적 원리를 강조한 가모프와 달리 경험적 접근을 취했다. 옥시토신과 바소프레신처럼 상이한 단백질이더라도 서열 전체에서 한두 개의 아미

Major-Minor Grid for Amino Acids				
Major Determinants				
Minor Determinants	1	2	3	4
1 1			Tyr	
1 2			Tyr	Glu
1 3				
1 4				
2 1				
2 2				Glu
2 3				
2 4				
3 1				
3 2				
3 3				
3 4	Ileu	Phe		
4 1				
4 2				
4 3				
4 4				

〈표 2〉 메이저-마이너 코드표

1,2,3,4는 RNA를 구성하는 네 종류의 염기이며, 표의 가로항목과 세로항목은 각각 하나의 '주요한 결정인자 염기'와 두 개의 '부수적 결정인자 염기'를 나타낸다(Gamow, Rich & Yčas 1956: 53).

노산만 교체되는 경우가 많았다.⁵⁾ 이에 착안한 오겔과 파인만은 코드를 구성하는 세 염기 중 중앙의 염기가 주요한(major) 결정인자이고 양옆의 염기가 부수적(minor) 결정인자라고 보면서, 주요 염기가 대체되면 대응 아미노산이 반드시 바뀌지만 부수적 염기의 대체는 아미노산의 변화를 반드시 수반하지는 않는다고 가정했다. 아미노산 치환현상에 부합하는 이런 번역원리는 코드를 고안하는 데 서열 데이터를 직접 활용할 수 있게 해줬다. 오겔과 파인만은 실제 서열 데이터 상에서 X아미노산이 Y아미노산으로 치환되는 경우를 찾아내어 “X의 주요 염기는 Y의 주요 염기와 다르다”는 규칙을 적용했고, 이를 통

5) 가령 옥시토신의 서열은 Cys-Tyr-Ileu-Glu-Aspn-Cys-Pro-Leu-Gly이고, 바소프레신의 서열은 Cys-Tyr-Phe-Glu-Aspn-Cys-Pro-Arg-Gly이다.

해 64개 삼중코드와 20개 아미노산의 대응관계를 보여주는 ‘메이저-마이너 코드표’를 작성했다(표2). 서열 데이터의 양이 충분치 않아 표의 일부를 작성 하는데 그쳤지만, 단백질 서열분석 실험으로 아미노산 서열이 계속 축적되고 있으므로 코드표의 나머지 부분도 조만간 완성할 수 있을 거라 기대했다 (Gamow, Rich & Yčas, 1956: 51-53).

마지막으로 에드워드 텔러가 제안한 번역코드도 메이저-마이너 코드와 비슷한 유형이었다. 일단 텔러는 단백질이 합성될 때 아미노산들이 RNA의 두 염기 사이에 위치하며, 양옆의 염기에 의해서 대응 아미노산이 결정된다고 보았다. 두 개의 염기만으로는 20개의 아미노산과 대응관계를 갖기에 불충분하기에, 텔러는 염기 외에도 앞에 놓인 아미노산이 그 뒤를 따르는 아미노산이 정해지는 데 영향을 미친다고 가정했다. 선행하는 아미노산이 뒤따르는 아미노산들을 연쇄적으로 결정해주는 특징 때문에 이 번역코드에는 연쇄 코드(sequential code)라는 이름이 붙었다. 가모프는 텔러의 연쇄 코드가 실험 데이터에 대한 의존도가 가장 크기에 서열 데이터만 충분해지면 가장 빨리 성과를 거둘 번역코드라고 평가했다(Gamow, Rich & Yčas, 1956: 53-54).

3) 암호해독 메타포와 서열해독 기법의 활용

번역코드 연구의 난점은 ‘출력’에 해당하는 아미노산 서열은 약간 알려져 있지만 ‘입력’에 해당하는 염기 서열은 알 길이 없다는 점에 있었다. 1950년대 중반에 몇몇 단백질의 서열은 이미 밝혀져 있었고, 서열분석 실험을 통해 아미노산 데이터는 계속 축적되고 있었다. 반면 염기 서열은 당대의 실험으로는 접근 불가능한 미지의 영역이었다. 이런 측면에서 타이 클럽 연구자들은 단백질 합성을 암호화(coding) 과정에 비유했는데, “평균”인 염기 서열이 “암호문”인 아미노산 서열로 번역되는 것으로 개념화했다. 마치 암호문을 해독(decoding)할 때처럼, 이미 알려진 “아미노산 서열(암호문)”을 체계적으로 분석

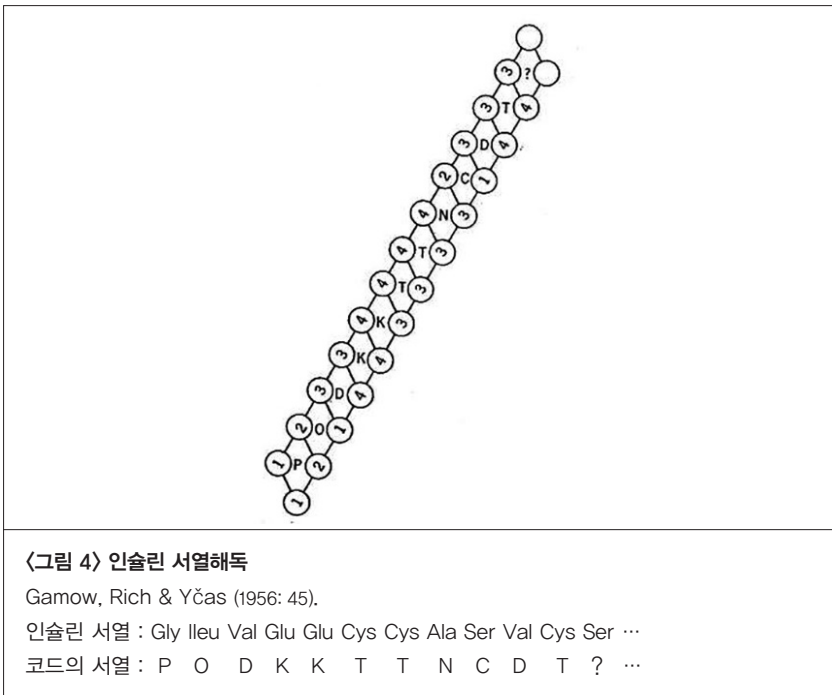
하는 것이 그것에 대응하는 “염기 서열(평문)”과 “번역 메커니즘(암호화의 원리)”을 알아내는 유일한 방법이라 여겼다(Gamow & Yčas, 1956: 68).

가모프는 실험으로 밝혀진 아미노산 서열을 번역코드를 사용해 나타냄으로써 그 대응 염기 서열을 추정하는 기법을 고안했다. 개별 번역코드들은 각각 고유한 기호간 상관관계를 지녔고, 20개의 삼중코드들은 이런 상관관계의 규칙을 만족시키는 형태로만 배열될 수 있었다. 따라서 기호간 상관관계의 제약조건을 만족시키면서 실제 아미노산 서열을 나타낼 수 있는 코드의 배열을 찾아냄으로써, 20개의 삼중코드들이 각각 어떤 아미노산에 대응하는지 조사할 수 있었다. “서열해독(decoding sequence)”이라 불린 이 기법은 이론적으로 제시된 번역코드의 타당성을 검증하고 코드와 아미노산의 대응관계를 밝혀내기 위한 방법으로 활용되었다(Gamow, Rich & Yčas, 1956: 43)⁶⁾.

이해를 돕기 위해 가모프의 서열해독 사례를 살펴보자. 인슐린A의 서열은 “Gly-Ileu-Val-Glu-Glu-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Ala-Gly-Val”이었는데, 가모프는 이것을 다이아몬드 코드로 나타낼 수 있는지 조사했다. 불과 몇 줄에 불과한 문장이 암호문 해독의 키 구절이 되듯이, 많은 서열 데이터 중에서도 특이하게 배열되는 부분이 중요했다. 가모프는 전체 서열 중에서도 -Glu-Glu-Cys-Cys-에 주목했는데, 다이아몬드 코드의 결합규칙에 의하면 이중으로 연속할 수 있는 코드 배열은 네 개(-A-A-N-N-, -L-L-M-M-, -K-K-T-T-, -T-T-S-S-)밖에 없었고, Cys가 Glu외에 다른 아미노산과 결합할 수 있다는 점까지 고려하면 -Glu-Glu-Cys-Cys-의 대응 코드는 단 한 가지(-K-K-T-T-)로 좁혀졌다([그림1], [그림4]). 가모프는 이 방법을 나며

6) 가모프는 “번역코드가 옳으면 단 하나의 대응 해(解)가 존재할 것”이라 말했는데, 특정 번역코드로 전체 아미노산 서열을 해독할 수 있으면 이를 실험결과에 부합하는 번역코드로 간주할 수 있으며, 더 나아가 코돈과 아미노산의 실제 대응관계까지 밝힐 수 있다고 보았던 것이다.

지 서열에도 차례로 적용해 11번째 아미노산까지 대응 코돈을 밝혀냈지만, 12번째 아미노산에서 문제가 생겼다. 11번째 아미노산까지의 해독 결과 Ser에는 코돈C가 할당되었는데, 코돈C는 코돈T(Cys)와 연결될 수 없었기에 11-12번째 서열인 -Cys-Ser-에서 모순이 발생했다(Gamow, Rich & Yčas, 1956: 44-46). 이러한 서열해독 작업은 인슐린A의 서열이 배열이 불가능한 '금지된 서열'이며, 따라서 다이아몬드 코드는 부적절한 정보전달 모형임을 알려줬다.



한편 위와는 다른 서열해독 시도도 있었다. '서열해독 기법'이 염기 서열을 추정하는 유일한 방법은 아니었는데, 사실 그것은 시행착오와 요령에 의존하는 방법으로 한계가 많았다. 가장 확실한 길은 20개의 코드에 아미노산들

을 먼저 대응시킨 후 그것으로 아미노산 서열을 나타낼 수 있는지 검토하고, 이 작업을 $20!(2.3 \times 10^{17})$ 개에 달하는 모든 코드-아미노산 대응 경우에 일일이 반복하여 이 중에서 실제 서열과 모순되지 않는 코드-아미노산의 대응관계를 찾아내는 것이었다. 하지만 코드-아미노산의 대응 가짓수만 해도 엄청나므로 “모든 가능한 경우를 일일이 검토하기란 컴퓨터를 이용하더라도 불가능”했다.⁷⁾ 즉, 직접적인 서열해독을 위해서는 하드웨어적 속도를 높이는 것만으로는 부족했고, 이 문제에 적합한 데이터 처리기법이 필요했다.

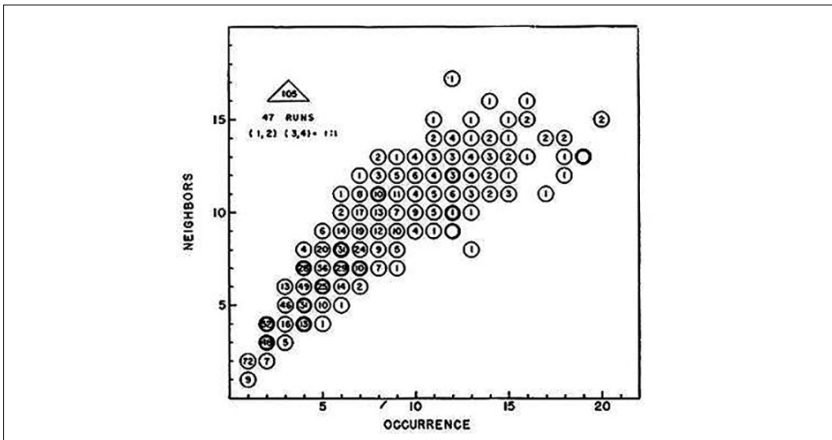
존스홉킨스의 컴퓨터엔지니어 레들리(R. S. Ledley)는 이런 직접적인 해독 기법에 관심을 보였다. 부울대수를 활용한 컴퓨터 계산기법을 개발했던 레들리는 단백질 서열해독에 자신의 데이터 처리기법이 유용하게 활용될 수 있음을 알아챘다(Ledley, 1987: 33-34). 이에 레들리는 막대한 데이터를 효율적으로 처리하는 컴퓨터 계산기법과 그 응용 사례를 소개한 논문을 발표했다. 여기서 레들리는 단백질 서열해독 문제를 예시하면서, 코드-아미노산 대응관계를 기호논리 방정식으로 표현한 후 이미 알려진 아미노산 서열의 제한조건으로부터 방정식 해를 구하는 기법을 제안했고, 이를 이용할 경우 계산시간을 수백 시간 정도로 단축시킬 수 있음을 보여줬다(Ledley, 1955: 508-511).

4) 예기치 않은 변화: 아미노산 서열에 대한 통계적 분석

RNA 타이 클럽 연구자들은 아미노산 서열을 문제해결의 실마리가 담긴 자료로 취급하며 번역코드 연구에 다양한 방식으로 활용했다. 메이저-마이너 코드와 연쇄 코드처럼 몇몇 번역코드는 실제 서열을 참고해 만들어졌는데, 이런

7) 가모프에 따르면 컴퓨터를 사용해 “1초에 100만개의 코드-아미노산 대응 경우를 조사하더라도, 20!의 경우를 모두 검토하기 위해서는 로마제국시기 이상의 시간이 소요된다(Gamow, Rich & Yčas 1956: 54-55).”

코드에는 축적된 서열 데이터에서 나타나는 배열상의 특징이 고스란히 반영되었다. 또한 서열해독의 경우 아미노산 서열을 이용해 이론적으로 제시된 번역코드의 타당성을 검토했고, 이를 통해 코드-아미노산의 대응관계까지 규명하고자 했다. 마지막으로 서열 데이터에 대한 통계분석도 이루어졌는데, 이 기법은 여러 번역코드 중 타당성이 더 높은 것을 선별하기 위해 고안된 기법이였다. 서열 통계분석은 “번역코드로 재현된 가상 서열(artificially constructed sequences)과 실제 아미노산 서열(observed sequences)의 배열 형태를 비교하는 기법”으로, 이 방법을 쓰려면 실험으로 밝혀진 전체 단백질 서열과 번역코드로 재현된 가상 서열에서 20종의 아미노산이 몇 번씩 나타나며 그것에 인접한 아미노산은 몇 종인지 조사해야 했다. 그리고 ‘아미노산 발생 빈도’와 ‘인접 아미노산 개수’가 표시된 통계 그래프를 작성하여 실제 서열과 가상 서열의 배열 형태가 얼마나 유사한지 비교해야 했다(그림5).



〈그림 5〉 가상 서열과 실제 서열의 통계적 특징 비교

굵은 원은 서열 데이터에 근거해 산출된 아미노산 ‘발생빈도’와 ‘인접 아미노산 개수’를 나타내고, 가는 원은 ‘조밀한 삼각형 코드’로 재현된 가상 서열의 ‘발생빈도’와 ‘인접 아미노산 개수’를 나타낸다(Gamow, Rich & Yčas 1956: 61).

이 같은 통계처리는 로스알라모스의 컴퓨터엔지니어 메트로폴리스의 도움이 있었기에 가능했다. 가상 서열을 만들려면 무작위로 배열된 염기 서열의 견본을 추출한 후 그것을 코드로 번역해야 했고, 이 작업을 여러 차례 반복해 통계적으로 유의미한 결과를 얻어야 했다. 가모프는 “매니악(MANIAC)을 활용한 몬테카를로 시뮬레이션으로 이 작업을 진행했다”고 언급했는데, 이 기법에 정통했고 매니악을 자유롭게 쓸 수 있었던 메트로폴리스는 이런 방향의 접근을 가능하게 해줬다(Gamow & Metropolis, 1954: 779-780; Metropolis, 1990: 247-248). 타이 클럽 연구자들은 번역코드A보다 번역코드B로 재현된 가상 서열이 실제 아미노산 서열과 유사하면 번역코드B를 더 적절한 정보전달 모형으로 간주할 수 있다고 보았고, 이 방법을 여러 가능한 번역코드 중에서 더 적합한 것을 판단하는 기준으로 이용했다(Gamow, Rich & Yčas, 1956: 59).

하지만 번역코드의 타당성을 검증하려는 의도와 달리 실제 서열 통계 분석은 예상 밖의 결과를 산출했다. 여러 번역코드로 재현한 가상 서열을 실제 아미노산 서열과 비교하니, 번역코드의 기호간 상관관계가 덜 제약적일 수록 실제 서열과 배열패턴이 유사해지는 결과가 나왔던 것이다. 실험으로 밝혀진 아미노산 서열은 어떤 제약이 있기보다는 특별한 규칙 없이 무작위로 배열되는 듯했다. 이에 타이 클럽 연구자들은 당황할 수밖에 없었는데, 이런 분석결과는 “코드가 중첩되어 그것들 사이에 상호 연관관계가 존재한다”는 그들의 근본 가정을 흔들었기 때문이다(Gamow, Rich & Yčas, 1956: 59-61).

따라서 이런 결과에 대한 추가적 검토 작업, 즉 아미노산이 실제로 무작위로 배열되는 것인지를 확인하는 작업이 시급해졌다. 이를 위해 가모프와 이차스는 전체 서열을 두 개의 아미노산이 결합된 디펩타이드(dipeptide) 단위로 분해하여 ‘인접 아미노산 분포도’([표3])로 정리했고, [표3]의 다수의 공

란(空欄)과 디펩타이드의 발생 빈도를 분석했다.⁸⁾ 아미노산이 무작위로 배열된다면 향후 서열이 충분히 축적되어 400(20×20)종의 디펩타이드가 모두 발견될 것이고, 따라서 분포도의 공란은 실험 데이터 부족에 기인한 우연한 누락으로 간주할 수 있었다. 반면 아미노산이 규칙에 따라 배열된다면, 분포도의 공란에는 통계적 누락 외에 실험 데이터가 축적되어도 절대 나타나지 않는 ‘금지된 디펩타이드’가 포함되어야 했다. 따라서 분포도의 공란은 전자보다 후자의 경우에 더 많을 것이었다. 또한 서열이 규칙에 따라 배열된다면 더 빈번하게 나타나는 아미노산 배열이 있을 것이고 이에 따라 무작위로 배열될 때보다 같은 디펩타이드가 여러 번 나타날 확률이 커질 것이었다.

First Amino Acid	Second Amino Acid																			
	Gly	Ala	Glu	Gln	Arg	Asp	Cys	Ser	Leu	Val	Pro	Thr	Lys	Ileu	Phe	Tyr	His	Met	Try	Aspn
Gly	—	—	1	—	1	1	—	1	—	2	—	—	2	1	2	—	—	1	—	—
Ala	2	1	1	—	—	—	—	1	3	—	—	—	1	1	—	—	1	1	—	—
Glu	—	2	—	1	1	—	—	—	—	4	—	—	—	1	1	—	1	—	—	1
Gln	—	4	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	1
Arg	2	—	—	—	1	—	1	—	—	—	1	—	—	—	—	—	—	—	—	1
Asp	1	1	—	2	—	1	—	1	—	1	—	—	—	1	—	—	—	—	—	—
Cys	2	2	—	1	1	—	1	1	—	—	1	1	—	—	1	1	—	—	—	1
Ser	—	—	2	—	—	1	—	—	1	1	—	—	—	2	1	1	1	1	—	—
Leu	2	2	2	—	—	—	1	—	—	3	—	—	—	1	—	—	—	—	—	—
Val	2	—	2	2	—	2	3	—	—	—	—	—	1	—	1	1	—	—	—	1
Pro	1	1	1	—	1	—	—	—	2	2	—	—	—	2	—	—	—	—	—	—
Thr	—	2	—	1	—	1	—	3	—	—	2	—	—	—	—	—	—	—	—	—
Lys	1	1	—	1	1	—	1	—	—	2	1	—	1	—	—	—	—	—	—	—
Ileu	—	1	—	2	—	—	1	—	1	1	1	—	—	1	—	—	—	—	—	—
Phe	1	—	—	2	1	1	—	—	1	1	1	—	—	—	1	1	—	—	—	—
Tyr	—	—	—	1	—	—	1	1	1	—	1	1	—	1	1	—	—	—	—	—
His	—	—	—	—	—	—	—	1	1	—	—	1	1	1	1	—	—	—	—	—
Met	—	—	1	—	—	—	1	—	—	—	—	1	1	—	—	—	—	—	—	—
Try	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Aspn	—	—	—	1	—	—	1	—	—	—	—	—	—	—	—	1	—	—	—	—

〈표 3〉 인접 아미노산 분포도

전체 단백질 서열 중에서 총 136종의 디펩타이드가 177번 나타났고, 이중에서 두 번 이상 나타난 디펩타이드는 33종이었으며, ‘Gly-Gly’처럼 한 번도 나타나지 않은 디펩타이드는 264종이었다(Gamow, Rich & Yčas 1956: 63).

8) 그들은 인접 아미노산 분포도를 작성하기 위해 Actin, Adrenocorticotropin, Carboxypeptidase, Cytochrome C, Globulin, Hemoglobin, Insulin, Lysozyme, Ovalbumin, Oxytocin, Papain, Pepsin, Ribonuclease, Serum Albumins, Silk Fibroin, Vasopressin의 아미노산 서열을 분석했다 (Gamow, Rich & Yčas 1956: 29-32).

위의 추론에 근거해 가모프와 이차스는 디펩타이드 발생빈도와 무작위 분포의 상관성을 조사했다. 우선 그들은 [표3]의 디펩타이드를 발생빈도에 따라 분류하여 [표4]의 A항목으로 정리했는데, 실제 서열에서 한 번도 나타나지 않은 디펩타이드는 264종이었고 한 번 이상 나타난 것은 각각 103종(1번), 27종(2번), 4종(3번), 2종(4번)이었다. 그리고 그들은 이런 발생빈도가 무작위 분포인지 살피기 위해 푸아송 분포공식을 사용해 아미노산이 무작위로 배열될 때의 디펩타이드의 개수를 이론적으로 계산했다. 그 결과는 [표4]의 E항목에서 확인할 수 있듯이 실제 디펩타이드의 발생 빈도와 거의 일치했다(Gamow, Rich & Yčas, 1956: 63). 실제 아미노산 서열의 배열패턴은 무작위 분포와 잘 일치했지만 번역코드로 재현된 가상 서열과는 큰 차이가 났다. 가모프와 이차스는 메트로폴리스의 도움을 받아 [표3]에 포함된 전체 디펩타이드와 동일한 정보량을 가지는 임의의 염기 서열을 추출한 후 그것을 성긴 삼각형 코드와 다이아몬드 코드로 번역했고, 이렇게 재현된 가상 서열에 포함된 디펩타이드를

No. of units per grid box	A Nearest neighbor, observed distribution	B Loose triangle code, observed distribution	C Diamond code, observed distribution	D English language, observed distribution	E Poisson distribution calculated
0	264	276	301	305	264.2
1	103	86	54	55	116
2	27	26	28	23	25.6
3	4	9	11	7	3.77
4	2	3	3	3	0.42
5	0	0	1	3	0.037
6	0	0	0	2	0.0027
7	0	0	2	0	1.7×10^{-4}
8	0	0	0	2	9.5×10^{-6}
χ^2 (obs.)	7.64	22.7	2.4×10^4	4.2×10^5	
χ^2 (calc.) for 5% significance	7.82	7.82	11.1×10^4	12.6×10^5	

〈표 4〉 디펩타이드의 발생 빈도 비교표

표의 A항목은 실제 단백질의 디펩타이드 분포패턴이고, B와 C항목은 각각 '조밀한 삼각형 코드'와 '다이아몬드 코드'로 재현된 가상 서열의 분포패턴이다(Gamow, Rich & Yčas 1956: 64).

위와 동일한 방식으로 분류했다. 하지만 [표4]의 A항목과 B.C항목에서 확인할 수 있듯이, 번역코드로 재현된 가상 서열은 실제 서열의 디펩타이드 분포 패턴과 무시할 수 없는 차이가 났다.

‘아미노산이 무작위로 배열된다’는 점을 보여준 통계분석 결과는 ‘코드가 중첩되어 그 배열에 제한이 가해진다’는 근본 가정에 위배되었기 때문에 곤혹스러운 것이었다. 타이 클럽의 몇 년간의 연구 성과를 정리한 논문의 결론에서 가모프는 위의 분석결과를 논하면서 “올바른 정보전달 모형은 저자들이 제안한 번역코드들보다 훨씬 덜 제약적”이어야만 하며, 애초의 가정과 달리 “인접 아미노산 간 염기 공유는 일어나지 않는 것으로 보인다”고 언급했다 (Gamow, Rich & Yčas, 1956: 66-67). 당시 타이 클럽 연구자들은 ‘코드가 중첩된다’는 가정을 번역코드 연구의 핵심 전제로 삼았는데, 다이아몬드 코드뿐만 아니라 그 대안으로 고안된 여러 코드들은 예외 없이 모두 중첩코드였다. 따라서 위의 통계분석에 어떤 착오가 없다면, 그 동안 연구된 번역코드들은 잘못된 모형으로 모두 폐기되어야 했다.

더 나아가 중첩코드 가설이 틀렸을 경우, 아미노산의 배열을 분석해 핵산-단백질 간 번역 원리를 규명한다는 접근방식 자체가 쓸모없는 것이 될 수 있었다. 중첩코드의 기호간 상관관계 분석은 타이 클럽 연구자들의 핵심 연구 방법이었다. 이를 활용해 아미노산 서열을 분석하면 번역코드의 타당성을 검토할 수 있었고 코드-아미노산의 대응관계도 규명할 수 있었다. 그들은 번역코드를 검증하기 위해 고안한 여러 방법과 그 과정에서 축적한 경험이 향후 유전암호 해독 연구의 기반이 되며, 더 많은 아미노산 서열 데이터가 축적되면 그들의 연구가 가시적 성과를 낼 것이라 기대하고 있었다. 하지만 위의 통계분석 결과는 이런 낙관적 전망이 착각에 불과하다고 암시하고 있었다. 아미노산이 무작위로 배열된다면, 그래서 사실 코드는 중첩되지 않고 기호간 상관관계 또한 없는 것이라면, 기존의 다양한 검증방법들과 서열 데이터를 활용한

연구들은 전혀 근거 없는 것이 되었기 때문이다.

가모프, 리치, 이차스는 “아미노산 서열이 푸아송 분포를 따른다는 사실이 해독방법의 발견 가능성이라는 측면에서 실망스러운 결과”라고 말했다. 델 브릭은 이들의 연구를 소개하면서 “인접 아미노산 간 제약은 DNA 염기 서열의 실마리를 제공할 것이란 기대와 달리 중첩코드를 폐기하는 결과를 가져왔다”고 평했다(Gamow, Rich & Yčas, 1956: 65). 타이 클럽 연구자들은 그동안 제기된 다양한 번역코드의 타당성을 비교하는 방법을 마련하고자 아미노산 서열의 통계적 특성을 분석했지만 역설적이게도 그것은 이들이 제시했던 ‘모든’ 번역코드와 ‘모든’ 검증방법을 의미 없게 만들었다.

4. ‘퍼즐풀이’로의 회귀와 타이 클럽의 와해, 1956-1959

1) 예기치 않은 변화: 비중첩 번역코드와 어댑터 가설

서열 통계분석은 아미노산이 무작위로 배열되며 중첩코드보다 비중첩코드(nonoverlapping code)가 더 적합한 정보전달 모형임을 암시하고 있었다. 하지만 타이 클럽 연구자들은 이런 결과를 다른 식으로 해석하며 중첩코드에 한 동안 집착했다.⁹¹ 인접 아미노산이 염기를 공유하지 않으면 “코드의 모든 배열이 허용되므로 아미노산 서열 데이터를 이용해 이를 검증할 수 없다”는 언급에서 드러나듯, 비중첩코드는 타당성을 검증하기 힘든 번역코드였기 때문이

91 예를 들어 가모프, 리치, 이차스는 통계분석 결과를 ‘코드가 중첩되지 않는다’는 결론지으려면, 분석에 이용된 서열 데이터가 전체 단백질 서열을 대표하는 것으로 신뢰성을 갖는지, 아미노산 서열이 푸아송 분포를 따르는 것이 정확히 무슨 의미인지 먼저 고찰해야 한다고 주장했다(Gamow, Rich & Yčas 1956: 65).

다(Gamow & Yčas, 1955: 1011). 또한 기호간 상관관계에 대한 다양한 분석을 통해 관련 연구주제들로 확장되었던 중첩코드에 비하여, 이것이 원천적으로 불가능했던 비중첩코드는 훨씬 흥미가 떨어지는 것일 수밖에 없었다.

하지만 서열분석 실험이 잇달았던 상황에서 중첩코드 가설을 고수하기는 갈수록 힘들어졌다. 더 많은 단백질 서열이 밝혀지고 디펩타이드 목록에 인접-아미노산 쌍들이 새로 추가되면서, 중첩코드 가설의 기반은 더욱 취약해졌다. 1956년경 타이 클럽 연구자들은 중첩코드 가설을 폐기하고 비중첩코드로 연구를 전환하게 되는데, 케임브리지의 브레너가 이를 확고히 했다. “모든 중첩 삼중코드의 불가능함에 대하여”에서 브레너가 보여준 간명한 증명은 이전의 통계분석과 달리 이론의 여지가 없었다. 이 논문에서 브레너는 “코드가 중첩되어 두 개의 염기를 공유할 경우, 지금까지 알려진 아미노산 서열을 나타내기에는 64개의 코돈만으로 불충분하다”고 주장했고, 이것을 “인접 아미노

Amino Acid	C-Neighbors	N-Neighbors	Minimum No. of Triplets Required	Amino Acid	C-Neighbors	N-Neighbors	Minimum No. of Triplets Required	
Lys	18	17	5	Pro	13	12	4	
Ser	17	13	5	Tyr	12	10	3	
Gly	15	15	4	Glu	11	11	3	
Leu	15	15	4	Gln	12	9	3	
Cys	15	14	4	Asp	10	11	3	
Arg	14	16	4	Asn	9	10	3	
Ala	14	15	4	Ileu	9	9	3	
Val	14	12	4	His	6	9	3	
Thr	13	14	4	Met	5	7	2	
Phe	13	14	4	Try	3	3	1	
							Total	70

〈표 5〉 아미노산 별 삼중코드의 필요 개수

C-Neighbors와 N-Neighbors는 각각 특정 아미노산의 오른쪽과 왼쪽에 인접하는 아미노산이 몇 종류인지를 나타낸다. 중첩코드에서 ‘삼중코드 개수’와 ‘인접 아미노산 개수’의 관계는 다음과 같다. 특정 아미노산에 0-4종류의 아미노산이 인접할 때 그것에 필요한 삼중코드의 최소 개수는 1개이고, 5-8종류는 2개, 9-12종류는 3개, 13-16종류일 때는 4개, 17-20종은 5개이다. 삼중코드가 두 개의 염기를 공유할 경우 양쪽에 인접한 아미노산이 각각 독립적으로 가지는 염기는 하나인데, 염기의 종류가 총 4개에 불과하기 때문에 인접 아미노산의 개수가 4개씩 추가될 때마다 별도의 삼중코드가 1개씩 더 필요해지는 것이다 (Brenner, 1957: 687-694, esp. 688).

산의 개수에 따라 아미노산을 나타내는 데 필요한 삼중코드의 최소개수를 헤아릴 수 있다”는 점을 이용해서 증명했다. [표5]에서 확인할 수 있듯이, 브레너는 20종의 아미노산들에 대해서 인접 아미노산의 개수를 조사한 후 각각에 필요한 삼중코드의 최소개수를 헤아렸는데, 그 총합은 70개로 코돈의 개수가 64개에 불과하다는 점과 모순이었다. 따라서 그 전제인 중첩코드 가설이 잘못된 것일 수밖에 없었다.¹⁰⁾

중첩코드 가설이 폐기된 후 타이 클럽 연구자들은 유전암호 해독 연구의 한계와 어려움을 자주 피력했는데, 이는 초기의 낙관적 태도와 대조적이다. 물론 4비트 염기 서열을 20비트 아미노산 서열로 번역해주는 메커니즘을 찾는다는 점에서는 중첩코드와 별 차이 없었기 때문에 비중첩코드를 고안하는 것 자체는 그리 어려운 일이 아니었다.¹¹⁾ 문제는 중첩코드와 달리 그렇듯한 가설 이상의 의미를 갖기 힘들다는 점에 있었다. “현재의 주된 난국은 알려진 현상을 설명해주는 암호화 원리[번역 메커니즘]를 제시하는 데 있는 것이 아니라, 오히려 가능한 수많은 암호화 원리들 중에서 적합한 것을 선택하는 데에 있다”는 이차스의 언급은 당시 타이클럽 연구자들의 고민 지점을 잘 드러내준다(Yčas, 1956: 96). ‘코드가 중첩된다’는 가설이 폐기되어 아미노산 서열을 분석하는 방법이 쓸모없어짐에 따라, 유전암호 해독 연구의 성격은 큰 변화를 겪을 수밖에 없었던 것이다.

이런 측면에서 볼 때, 크릭의 어댑터 가설(adaptor hypothesis)이 타이

10) 이런 브레너의 증명은 이전의 통계분석 때보다 훨씬 많은 서열 데이터가 밝혀져 있었기 때문에 가능했다(Brenner, 1957: 689-692).

11) 별다른 수정 없이 비중첩코드로 바로 전용할 수 있는 중첩코드도 있었는데, 예를 들어 ‘조합 코드’는 염기가 공유되지 않는 점만 제외하면 이전의 삼각형 코드와 동일했다(Gamow & Yčas 1955: 68-69).

클럽에 미친 영향도 비중첩코드로의 전환 못지않게 컸다. RNA와 아미노산 간 매개 분자의 필요성을 논한 이 가설은 1955년에 크릭이 클럽 내부토론용 논문으로 작성한 “겹친 주형과 어댑터 가설에 관하여”에 포함되어 있었고, 정식으로 발표되지 않았음에도 여러 논문들에 인용되는 등 큰 주목을 받았다. 특히, 1957년 자메크닉과 호글랜드가 수용성 RNA를 발견한 후 어댑터 가설에 대한 관심이 높아졌는데, 아미노산들이 단백질로 합성되기 전에 ‘작은 RNA 분자’에 먼저 부착된다는 그들의 실험결과가 어댑터라는 매개 분자의 도움으로 단백질이 합성된다는 크릭의 가설과 잘 부합했기 때문이다.¹²⁾

어댑터 가설이 제기된 맥락은 타이 클럽의 유전암호 해독 연구와 관련지어 파악해야 분명해지는데, 이를 위해 먼저 RNA 분자모형에 대한 왓슨과 크릭의 입장 차이에 주목할 필요가 있다. 비록 왓슨과 크릭 모두 번역코드 연구에 관심을 보였지만, ‘핵산 분자에 아미노산들이 구조적, 화학적 정합성에 따라 배열, 합성된다’는 가설에 대한 두 사람의 평가는 극명하게 달랐다. RNA 분자 구조의 중요성을 강조했던 왓슨은 핵산이 단백질의 주형이라는 가모프의 가설에 우호적이었고, 이중나선 모형을 통해 유전자의 안정성과 복제 메커니즘이 해명되었듯이, RNA 분자모형이 단백질 합성 현상을 규명하는 실마리가 될 것이라고 주장했다(Rich & Watson, 1954: 760). 반면 크릭은 아미노산과 RNA의 직접적 결합을 가능케 하는 구조적, 화학적 정합성을 발견하기 힘들다고 보았다. 크릭의 관점에서 주형 가설은 어떤 물리-화학적 근거도 없는 추측일 뿐이었고, 가모프의 번역코드는 핵산이 어떻게 아미노산을 수용하는지 물리, 화학적 용어로 설명할 수 없다는 점에서 약점이 있었다.

12) 호글랜드는 어댑터 가설이 수용성 RNA의 기능을 이해하는 이론적 기반이 되었고 이를 통해 단백질 합성 메커니즘을 명료하게 설명할 수 있게 되었다고 평가했다. 평가했다(Hoagland, 1959: 40-46).

크릭의 어댑터 가설은 이러한 주형 가설의 약점을 보완하기 위한 것이었다. RNA와 아미노산이 직접 결합할 수 없다고 본 크릭은 둘을 매개하는 별도의 분자를 가정했고 이를 통해 단백질 합성을 설명했다. 그는 개별 아미노산을 인식하는 20개의 어댑터 분자가 존재하며 이들이 효소의 작용으로 각각에 맞는 아미노산과 결합한다고 보았다. 그리고 이렇게 아미노산과 결합한 어댑터 분자가 RNA와 다시 결합하여 아미노산들을 배열시키고 그것들이 연결되어 단백질이 만들어질 때까지 붙들어둔다고 주장했다. 크릭은 ‘RNA-어댑터-아미노산’의 간접적 상호작용을 도입함으로써 지나치게 단순화된 가모프 모형과는 차별화되는, 생화학적 관점에서도 가능한 단백질 합성 메커니즘을 제시했다(Crick, 1958: 155-156; Crick, 1966: 4-5).

하지만 수용성 RNA가 발견되어 어댑터 가설이 재조명될 때까지, ‘아미노산과 RNA를 매개하는 분자가 존재한다’는 크릭의 주장은 타이 클럽 연구자들 사이에서 쉽게 받아들여지지 않았다. 크릭은 “이 가설을 통해서 유전암호가 ‘어떠한’ 구조로든 가능하다는 점을 강조하고 싶었다”고 말했는데, 바로 이런 함의 때문에 그에 대한 반응은 긍정적이지 않았다(Crick, 1988: 96). 가모프는 자신의 주형 가설이 단순화된 것이지만 향후 정확한 RNA 분자모형이 밝혀지면 정교하게 보완될 수 있다고 보았고, 다른 타이 클럽 연구자들도 RNA 분자 구조를 번역 메커니즘의 실마리가 담긴 연구주제로 취급했다. 하지만 크릭의 어댑터 가설은 이런 RNA와 아미노산의 구조적 정합성에 착안한 연구가 세부 내용에서 잘못되었을 뿐만 아니라, 그런 방향의 접근 자체가 아예 불필요함을 이야기하고 있었다. 매사추세츠공대의 분자생물학자 레빈탈(C. Levinthal)은 수용성 RNA가 실험을 통해 발견되기 전에 그에 대한 이론적 고찰이 있었다고 지적하면서 “어댑터 가설 덕분에 구조적 논의로부터 자유로워졌으며, 암호화 관점에서만 문제를 고려할 수 있게 되었다”고 호평했지만, 사실 연구주제의 축소 측면에서 볼 때 이 가설이 타이 클럽 연구자들에게 미친

영향은 부정적인 것이었다(Levinthal, 1959: 250).

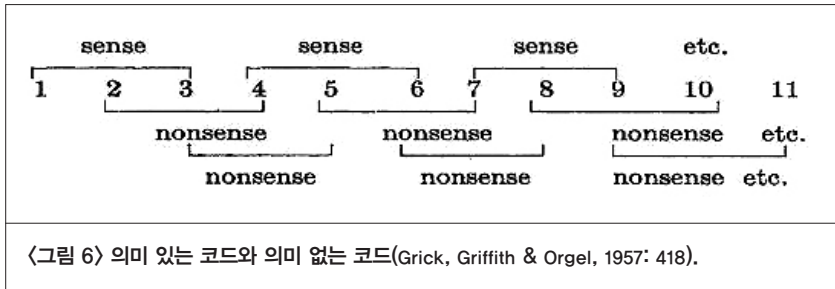
2) 다시 ‘퍼즐풀이’로: 크릭의 콤마 없는 코드

타이 클럽 연구자들은 단백질 합성을 ‘언어 간의 번역’에 비유하곤 했지만, 이들의 설명을 자세히 살펴보면 20종의 아미노산으로 번역되는 것은 선형적 염기 서열로 표현되는 ‘추상적 언어’라기보다는 핵산 분자의 ‘구조적 특성’이었다.¹³⁾ 하지만 어댑터 가설의 등장 이후 핵산 분자의 구조적 특징은 번역코드를 제시하는 데서 더 이상 중요한 정보가 될 수 없었고, 점차 핵산의 특이성은 염기들의 순서에 의해서만 결정되는 것으로 다루어졌다. 결국 비중첩코드 가설과 어댑터 가설이 수용되면서 번역코드 연구는 다양한 관련 연구들에서 점차 분리되었고 점점 더 추상적인 성격을 띠게 되었다. 이렇게 변화한 연구 성격은 1957년에 크릭이 케임브리지의 오겔과 그리피스스의 도움을 받아 고안했던 콤마 없는 코드(comma-free code)에서 잘 드러난다.

크릭은 합리적 정보전달 모형을 제시하려면 염기 서열이 아미노산 서열로 번역될 때 생기는 “구두점 문제”가 해결되어야 한다고 보았다. 이는 중첩코드 가설이 비중첩코드 가설로 대체되면서 대두한 문제로, 삼중코돈의 시작 지점과 끝 지점을 정하는 방법에 대한 것이었다. 가령 “·BCACDD ABABDC···”와 같은 염기 서열이 있을 때 “··BCA,CDD,ABA,BDC···”, “··B,CAC,DDA,BAB,DC···”, “··BC,ACD,DAB,ABD,C···” 중에서 무엇이 실제 아

13) 다이아몬드 코드의 경우 이중나선의 구조적 특징을 고려하지 않고서는 코드의 개수가 20개라는 결론을 이끌어 낼 수 없고, 삼각형 코드의 경우 수학적 논리만으로 코드의 개수가 20개라고 결론내릴 수도 있으나 그러한 설명은 선호되지 않았다. 예컨대 가모프는 삼각형 코드에 관해 ‘4종류의 염기에서 중복을 허용해 3개의 염기를 조합하는 가지수가 20개가 된다’고 설명하기보다는, “핵산 분자 상에 연속한 세 염기를 꼭짓점으로 하는 삼각형들이 형성되며 그것의 종류는 회전과 뒤집힘에는 무관하기 때문에 20개가 된다”고 설명했다.

미노산 서열에 대응하는지 모호하다는 점이 문제였다. 일단 그는 이 문제를 해결하기 위해 두 가지를 가정했다. 첫 번째는 64개의 삼중코돈 중 일부분은 아미노산에 일대일로 대응되는 ‘의미 있는 코드(sense code)’인데 반해 나머지는 대응 아미노산이 없는 ‘의미 없는 코드(nonsense code)’라는 가정이었다. 두 번째는 두 종류의 코드가 [그림6]과 같은 형태로, 즉 염기 서열의 모든 지점에서 ‘의미 있는 코드’가 중첩되지 않은 형태로 배열되며, 그것들 사이에 놓이는 두 개의 중첩코드들은 반드시 ‘의미 없는 코드’여야 한다는 가정이었다. 크릭은 ‘의미 있는 코드’와 ‘의미 없는 코드’의 배열패턴으로 가상의 콤팩타가 형성되며, 이 덕분에 정보전달 중에 생길 수 있는 오류를 피하게 된다고 보았던 것이다(Ÿčas, 1969: 31; Crick, Griffith & Orgel, 1957: 417-418).



위의 두 가정은 구두점 문제를 해결하기 위한 것이었지만, 그것에 마법의 숫자 20을 도출하는 수학적 메커니즘도 함축되어 있음이 곧 밝혀졌다. 크릭은 오겔과 자신의 아이디어에 대해 토론했고, 이 과정에서 오겔은 64개의 삼중코돈 중 ‘의미 있는 코드’의 개수는 20개를 넘을 수 없음을 발견했다. 우선 동일한 염기가 반복되는 네 가지 삼중코돈(AAA, BBB, CCC, DDD)은 ‘의미 있는 코드’ 목록에서 제외되어야 했는데, AAA가 ‘의미 있는 코드’라면 그것이 연결된 ‘...AAA,AAA...’에서 중첩 부분이 ‘의미 없는 코드’여야 한다는 가정에 위배되기 때문이다. 또한 특정 삼중코돈이 ‘의미 있는 코드’일 경우 그에 따라

필연적으로 ‘의미 없는 코드’가 될 수밖에 없는 삼중코돈도 존재했는데, 가령 ABC가 ‘의미 있는 코드’일 경우, 그것이 연결된 ‘…ABC,ABC…’에서 중첩 부분에 해당하는 BCA와 CAB는 ‘의미 없는 코드’여야 했다. 즉, 하나의 삼중코돈이 ‘의미 있는 코드’에 해당할 경우 그것의 원순열에 해당하는 두 개의 코돈은 ‘의미 없는 코드’에 속해야 하며, 따라서 남은 60개의 삼중코돈 중에서 ‘의미 있는 코드’의 개수는 그것의 1/3인 20개를 넘을 수 없었던 것이다.¹⁴⁾ 사실 이상의 논의는 동일한 삼중코돈이 연결될 때의 제한조건만을 따졌을 뿐이므로, 서로 다른 삼중코돈들이 연결되는 경우까지 고려하면 ‘의미 있는 코드’의 개수가 더욱 줄어들 수도 있었다. 문제는 가짓수가 너무 많아 서로 다른 삼중코돈이 연결되는 경우를 모두 고려하기 어렵다는 점에 있었는데, 캐번디시 연구소의 이론 물리학자 존 그리피스는 이를 모두 조사하는 대신에 체계적인 열거법(systematic enumeration)을 이용해 20개의 ‘의미 있는 코드’를 직접 예시함으로써, ‘의미 있는 코드’의 개수가 20개에서 더 이상 감소하지 않는다는 점을 보여주었다.¹⁵⁾

코드 배열상의 가상의 콤마라는 가정의 그럴듯함은 콤마 없는 코드를 매력적인 번역코드로 돋보이게 했다. 타이 클럽 회원은 아니었지만 콤마 없는 코드에 대해 알게 된 몇몇 생물학자들은 자신의 논문에 그것을 인용하길 원했

14) 이러한 증명과정에서는 아미노산 서열에 특별한 제한이 가해지지 않으며, 따라서 20개의 모든 아미노산이 서로 이웃할 수 있다는 점이 기본 전제로 사용되었다(Grick, Griffith & Orgel, 1957: 418).

15) 체계적 열거법을 이용하여 예시된 20개 코드의 목록은 다음과 같다.

$A \ B \ \begin{matrix} A \\ B \end{matrix}$	$\begin{matrix} A \\ B \end{matrix} \ C \ \begin{matrix} A \\ B \\ C \end{matrix}$	$\begin{matrix} A \\ B \\ C \end{matrix} \ D \ \begin{matrix} A \\ B \\ C \\ D \end{matrix}$
2개	6개	12개

고, 이에 크릭은 내부토론용으로 작성한 원고를 미 국립과학원회보에 정식으로 발표했다. 또한 『생명의 소용돌이』라는 대중서를 집필하고 있었던 작가 루스 무어(R. Moore)는 그의 책에서 콤마 없는 코드를 비중 있게 소개했다(Judson, 1979: 320-321; Crick, 1988: 100-101; Crick, 1966: 5). 콤마 없는 코드에 대한 관심은 생물학자들에게 국한되지 않았다. 당시 몇몇 수학자들은 순열 조합론을 이용해 콤마 없는 코드의 특징을 검토하기도 했는데, 골롬(S. W. Golomb)과 고든(B. Gordon), 웰치(L. R. Welch)의 연구가 대표적이었다. 칼텍에서 수학을 전공했던 이들은 텔브릭을 통해 콤마 없는 코드에 대해 알게 되었고, 그것의 제안자들이 언급했지만 풀 수 없었던 ‘콤마 없는 코드의 일반 해’를 구하는 문제를 대신 해결했다(Golomb, Gordon & Welch, 1958).

한 과학사가는 콤마 없는 코드가 틀린 것으로 판명 난 생물학 이론들 중 가장 우아한 것이었다고 평가했는데(Judson, 1979: 315), 실제로 타이 클럽 외부의 여러 연구자들이 번역코드 연구에 몰두하는 계기가 되는 등 당시에 콤마 없는 코드가 미친 파급효과는 적지 않았다(Woese, 1967). 하지만 콤마 없는 코드에 쏟아진 관심과 개념적 명료함에서 비롯된 설득력에도 불구하고, 핵산-단백질 정보전달 모형으로서 그것이 갖는 유용함에 관해서 크릭이나 타이 클럽 연구자들은 오히려 조심스러운 입장을 취했다. 1957년 9월, 실험생물학회에 참석했던 크릭은 단백질 합성에 대한 유용한 해석틀로 사용될 수 있는 서열 가설, 어댑터 가설, 센트럴 도그마를 제안한 후, 마지막으로 콤마 없는 코드를 소개하면서 “그것이 중요한 개념일 수도 있지만 완전히 어리석은 생각일 수 있다”고 평가했다. 많은 연구자들이 콤마 없는 코드를 높게 평가했지만, 사실 크릭의 관점에서 콤마 없는 코드는 체계적 검증방법이 마련되었던 이전의 중첩코드들과 달리 그럴듯한 가설 이상의 의미를 가지기 힘든, 근본적인 한계가 있는 번역코드였다(Crick, 1958: 160).

5. 결론적 고찰: 생산적 연구의제의 요건에 관하여

콤마 없는 코드가 마지막 번역코드는 아니었다. 이에 매료되었던 몇몇 연구자들은 번역코드 연구에 빠져들었고, 이들의 연구를 바탕으로 새로운 비중첩코드는 1960년대 초반까지 잇달아 등장했다. 1958년, 델브뤽은 칼텍의 수학자 골롬, 웰치와 함께 크릭의 콤마 없는 코드를 개선한 ‘교차 가능한 콤마 없는 코드(transposable comma-free code)’를 제안했다. 1959년, 칼텍의 분자생물학자 신샤이머(R. Sinsheimer)는 두 개의 염기쌍 아데닌-시토신, 구아닌-우라실을 유전부호로 하는 ‘이진 코드(binary code)’를 연구했다. 1962년, 골롬은 이전의 코드를 다시 개량하여 ‘이중직교 코드(biorthogonal code)’를 발표했다(Golomb, Welch & Delbück, 1958; Sinsheimer, 1959; Golomb, 1962). 제안된 번역코드들은 다양했지만 이들의 연구는 수학적이고 추상적인 성격이 두드러졌다는 점과 이상적인 정보전달 방법을 상정하고 그에 걸맞은 기능을 번역코드에 부여하려 했다는 점에서 공통점이 있었다. 정보의 효율적 전달, 정보전달 과정의 오류방지, 자체적 교정기능 등이 번역코드의 중요 요건으로 고려되기 시작했고, 이에 따라 수학자들과 정보이론 전문가들이 번역코드 연구에 보인 관심이 이 시기에 크게 증가했다.¹⁶⁾

이와는 대조적으로, 번역코드 연구의 본산이었던 타이 클럽의 연구 활동은 초기의 활기를 잃고 점점 미미해져갔다. 의견 교환 활동을 위주로 운영되었던 RNA 타이 클럽은, 연구자들 간의 교류가 지지부진해지면서 자연스럽게 와해의 길로 들어섰고, 공동연구의 성과로 간주될만한 번역코드는 1950년대 말 이후에는 더 이상 고안되지 않았다. 특히나 이런 쇠락의 모습은 ‘코드가

16) 페루츠의 회고에 따르면, 유전암호 해독 연구를 함께 하고 싶다는 수학자들과 정보이론 전문가들의 편지가 캐번디시 Medical Research Council에 쇄도했다고 한다(Judson, 1979: 345).

중첩된다'는 근본 가정이 의문시되고 그것이 비중첩코드 가설로 대체되는 1956년 말을 기점으로 나타났는데, 중첩코드 가설의 폐기가 '모임의 결속력'과 '연구의제의 성격' 측면에서 미친 영향은 되짚어볼 가치가 있다. 이러한 측면에 대한 고찰은 타이 클럽 초기의 활기, 즉 다양한 분과의 경계를 넘나는 협동연구를 유지시켜 준 동력이 무엇이었는지 이해하는 데 역으로 도움을 줄 수 있기 때문이다.

첫째, 다양한 접근방식을 활용한 공동연구를 통해 단백질 합성 메커니즘을 규명해내겠다는 모임의 목표는 중첩코드 가설이 폐기됨에 따라 달성하기 어렵게 되었다. 1953년 말, 다이아몬드 코드라는 첫 번째 번역코드가 제기된 이후, 그것에 대한 비판, 논쟁, 보완이 이어지면서 타이 클럽의 유전암호 해독 연구는 체계적 틀을 잡아갔고 관련 연구주체들로 풍부하게 확장되었다. 이 과정에서 중첩코드의 기호간 상관관계 개념은 다양한 접근방식을 활용한 연구들을 매개하는 연결고리 역할을 하면서, 서로 다른 학문적 배경을 지녔던 타이 클럽 성원들을 결속시켜준 핵심적 요소로 기능했다. 하지만 중첩코드 가설이 폐기되어 기호간 상관관계에 대한 분석이 무의미해지면서, 이들이 개발, 활용했던 분석기법들은 상당수 쓸모없는 것이 되고 말았다. 아미노산 배열패턴 분석, 단백질 서열해독, 컴퓨터를 활용한 데이터 처리, 서열 통계분석 등은 그 자체로 여전히 유의미한 연구였지만, 핵산-단백질 정보전달 모형을 밝혀내는 데는 더 이상 도움을 줄 수 없었던 것이다. 이 과정에서 클럽 성원간의 의견교환은 1957년 이후 드물어졌고, 공동의 관심사로서 번역코드 연구가 지녔던 의미도 점차 퇴색되었다.

둘째, 번역코드를 제시함으로써 단백질 합성 현상을 해명하려 한 타이 클럽 고유의 접근방식은 아미노산 서열을 더 이상 활용할 수 없게 되면서 생물학 연구방법으로서의 가치를 상실하게 되었다. 처음에 번역코드 연구는 4비트 숫자열과 20비트 문자열 간의 번역 메커니즘을 고안하는 퍼즐풀이 작업으

로 시작했지만, 이후 서열 데이터를 활용한 체계적인 분석방법들이 마련되면서 하나의 유용한 생물학 연구방법으로 자리 잡을 수 있었다. 타이 클럽 연구자들은 이론적으로 제시된 번역코드를 서열분석 실험결과로 검증했고, 더 나아가 서열해독 방법을 활용해 코드-아미노산의 대응관계까지 규명하려 했다. 하지만 이것을 가능하게 해줬던 중첩코드 가설에 문제가 생기면서 번역코드 연구의 의미가 퇴색되기 시작했다. 이후에 타이 클럽 연구자들이 제시했던 비중첩코드들은 타당성을 검증하기 힘든 순전히 이론적인 코드였다는 점에서, 그리고 당대 생물학의 연구성과들로부터 고립되었다는 측면에서 ‘대단치 않은 연구’에 불과한 것으로 평가되었다.

첫 번째 측면 때문에 타이 클럽의 연구자들을 매개하던 내부의 연결고리가 끊어졌다면, 두 번째 측면은 타이 클럽의 독특한 접근방식을 당대 생화학 및 분자생물학의 성과와 매개한 외부의 연결고리를 끊어 놓았다. 이렇듯 중첩코드 가설의 폐기와 이에 따라 나타난 연구의제의 성격 변화는 이후 타이 클럽의 활동에 심대한 영향을 미쳤는데, 이는 특히 모임의 허브 역할을 했던 가모프에게서 두드러지게 나타났다. 기호간 상관관계 분석에 의거하여 개별 연구들을 총괄, 종합했던 가모프는 중첩코드와 비중첩코드의 유용성의 차이를 가장 잘 인식하고 있었고, 어림짐작에 불과한 비중첩코드 연구에 더 이상 매력을 느낄 수 없었다. 1957년 브레너에게 보낸 편지에서, 가모프는 중첩코드 가설이 틀렸다는 증명은 타당하지만 “이론적 관점에서 볼 때, 코드가 중첩되지 않는다면 번역코드 연구가 훨씬 덜 흥미롭게 된다”고 이야기했고, 번역코드에 대한 참신한 아이디어가 더 이상 떠오르지 않는다면 “기본입자이론과 마찬가지로, 암호화 문제 또한 난관에 봉착한 것으로 보인다”고 말했다(Judson, 1979: 332). 구성원들의 결속력이 유지되고 개별 연구성과들이 교류되는 데 가모프의 역할이 결정적이었음을 감안할 때 이 같은 가모프의 심정 변화는 모임 와해의 직접적 요인이라 할 수 있다.

번역코드의 고안이라는 핵심과제에서 한 걸음 떨어져 있으면서 자기 전공과의 연관성에 착안해 타이 클럽에 동참했던 물리학자, 수학자, 컴퓨터엔지니어들의 경우, 이들에게 연구 동기를 부여한 가모프의 활동이 지지부진해지면서 타이 클럽에서 자연스럽게 멀어져갔다. 언뜻 무관해 보이는 분야들을 연관지어주며 새로운 유형의 연구를 가능하게 했던 기호간 상관관계 개념이 폐기되면서 번역코드 연구에서 이들의 역할은 크게 줄어들 수밖에 없었던 것이다. 하지만 그렇다고 하여 이들의 연구 자체가 중단된 것은 결코 아니었다. 가령 레들리의 서열해독법과 메트로폴리스의 서열 통계분석처럼, 이들 중 몇몇은 번역코드 연구에 자신의 전문분야를 성공적으로 접목시켰을 뿐만 아니라, 번역코드 연구와의 연관성이 끊어진 후에도 독자적인 방향으로 관련 연구를 계속 진척시켜나갔다(Ledley, 1987). 생물학자인 왓슨과 리치의 반응도 이와 비슷했는데, 비록 번역코드 고안의 실마리라는 의미는 사라졌지만, RNA 분자 모형에 대한 연구는 이들에게 여전히 중요한 연구주제였다.

한편, 단백질 합성 연구 자체에 계속해서 열의를 보였던 타이 클럽 연구자들도 있었는데, 캐번디시 연구소의 크릭과 브레너가 대표적이다. 물론 이들은 중첩코드에서 비중첩코드로 전환됨에 따라 그 성격에 크게 변화한 이론적인 번역코드 연구에 그리 큰 의미를 부여하진 않았다. 그럼에도 불구하고 크릭과 브레너는 “핵산의 염기 서열에 의해서 단백질을 구성하는 아미노산의 결합순서가 결정된다”는 초기의 문제의식이 여전히 유효하다고 확신했다. 또한 그들은 단백질 합성 메커니즘 연구가 당대 생물학에서 가장 중요한 주제라고 평가했고, 유전자인 핵산이 단백질 합성 과정에 관여하는 방식을 밝혀내겠다는 목표로 1960년대 초반까지 활발히 연구했다. 그 과정에서 그들은 운반 RNA와 전령 RNA, 서열 가설, 센트럴 도그마 같은 이론적 개념들을 제안할 수 있었고, 그들의 예상처럼 유전정보의 전달이 3개의 염기를 단위로 이루어진다는 점을 실험을 통해 보여줄 수 있었으며, 또 유전변이가 아미노산 서열에 미

치는 영향을 확인하는 실험을 준비할 수 있었다. 이 같은 연구성과들은 1961년에 유전암호가 밝혀지는 과정에서 크게 기여했다(Crick, Barnett, Brenner & Watts-Tobin, 1961).

이상의 분석은, 목표 달성의 실패와 모임의 와해라는 결말에도 불구하고 RNA 타이 클럽의 연구 활동을 ‘방향을 잘못 잡은 실패한 기획’으로 치부하고 넘어가선 안 되는 이유를 알려준다. 가모프의 대담한 문제제기로 결성된 RNA 타이 클럽은 당대 주류 생물학의 연구 프로그램에서는 쉽게 등장할 수 없었던 새로운 아이디어와 방법들이 제기되고 검토되는 시험장 역할을 했다. RNA 타이 클럽에 대한 이상의 분석은, 1950년대 중반 이후 생물학계에 점진적으로 수용되었던 “염기 서열이 아미노산 서열을 결정한다”는 대담한 가설의 출처가 어디였는지, 또 핵산과 단백질에 대한 다양한 연구들을 단백질 합성 메커니즘을 중심으로 종합되도록 해준 이 가설이 애초에 어떤 과정을 거쳐 강화되고 정교해졌는지 이해하는 데 도움을 줄 수 있을 것이다.

참고문헌

- 박상욱 (2012), 「융합은 얼마나: 이론상의 가능성과 실천상의 장벽에 관하여」, 홍성욱 엮음 (2012), 『융합이란 무엇인가: 융합의 과거에서 미래를 성찰한다』, 서울: 사이언스북스, 21-40쪽.
- 이상욱 (2015), 「기후과학의 철학적 쟁점: 성공적 융합 연구를 위한 인식론적 조건과 윤리적 함의」, 『과학철학』 제18권, 제1호, 151-180쪽.
- 홍성욱 (2012), 「창의적 융합의 '벡터 모형」, 홍성욱 엮음 (2012), 41-68쪽.
- Beadle, G. W. and Tatum, E. L. (1941), "Genetic Control of Biochemical Reaction in Neurospora", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 27, pp. 499-506.
- Brenner, S. (1957), "On the Impossibility of All Overlapping Triplet Codes in Information Transfer from nucleic Acid to Proteins", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 43, pp. 687-694.
- _____ (1959), "The Mechanism of Gene Action", in G. E. Wolstenholme and C. M. O'Connor (eds.), *CIBA Foundation Symposium on Biochemistry of Human Genetics*, Boston: Little, Brown, pp. 304-328.
- Crick, F. H. C. (1958), "On Protein Synthesis", *Symposium of the Society for Experimental Biology* 12, New York: Academic Press, pp. 138-163.
- _____ (1963), "On the Genetic Code", *Science*, Vol. 139, pp. 461-464.
- _____ (1966), "The Genetic Code-Yesterday, Today, and Tomorrow", *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 31, pp. 3-9.
- _____ (1988), *What Mad Pursuit: A Personal View of Scientific Discovery*, New York: Basic Book.

- Crick, F. H. C., Griffith, J. S., and Orgel, L. E. (1957), "Codes Without Commas", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 43, pp. 416-421.
- Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961), "General Nature of the Genetic Code for Proteins", *Nature*, Vol. 192, pp. 1227-1232.
- Cummings, J. N. and Kiesler, S. (2005), "Collaborative Research across Disciplinary and Organizational Boundaries", *Social Studies of Science*, Vol. 35, pp. 703-722.
- Gamow, G. (1954a), "Possible Relation between Deoxyribonucleic Acid and Protein Structures", *Nature*, Vol. 173, p. 318.
- _____ (1954b), "Possible Mathematical Relation between Deoxyribonucleic Acid and Proteins", *Det Kongelige Danske Videnskavernes Selskab, Biologiske Meddelelser*, Vol. 22, pp. 1-13.
- Gamow, G. and Metropolis, N. (1954), "Numerology of Polypeptide Chains", *Science*, Vol. 120, pp. 779-780.
- Gamow, G., Rich, A., and Yčas, M. (1956), "The Problem of Information Transfer from the Nucleic Acids to Proteins", *Advances in Biological and Medical Physics*, Vol. 4, pp. 23-68.
- Gamow, G. and Yčas, M. (1955), "Statistical Correlation of Protein and Ribonucleic Acid Composition", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 41, pp. 1011-1019.
- _____ (1956), "The Cryptographic Approach to the Problem of Protein Synthesis", in H. P. Yockey (eds.), *Symposium on Information Theory in Biology*, New York: Pergamon Press, pp. 63-69.
- Golomb, S. W. (1962), "Efficient Coding for the Desoxyribonucleic Channel", *Proceedings of Symposia in Applied Mathematics*, Vol. 14, pp. 87-199.
- Golomb, S. W., Gordon, B., and Welch, L. R. (1958), "Comma-Free Codes",

- Canadian Journal of Mathematics*, Vol. 10, pp. 202-209.
- Golomb, S. W., Welch, L. R., and Delbruck, M. (1958), "Construction and Properties of Comma-Free Codes", *Det Kongelige Danske Videnskavernes Selskab, Biologiske Meddelelser*, Vol. 22, pp. 1-34.
 - Hoagland, M. B. (1959), "The Present Status of the Adaptor Hypothesis", *Brookhaven National Laboratory Symposia* June, pp. 40-46.
 - Kay, L. E. (2000), *Who Wrote the Book of Life?: A History of the Genetic Code*, Stanford: Stanford University Press.
 - Klein, J. T. (1996), *Crossing Boundaries: Knowledge, Disciplinarity, and Interdisciplinarity*, Charlottesville: University Press of Virginia.
 - Klein J. T., et al. (2001), *Transdisciplinarity: Joint Problem Solving among Science, Technology, and Society*, Basel: Birknauser.
 - Keller, E. F. (2000), *The Century of the Gene*, Cambridge, Mass.: Harvard University Press.
 - Jeffrey, P. (2003), "Smoothing the Waters: Observations on the Process of Cross-Disciplinary Research Collaboration", *Social Studies of Science*, Vol. 33, pp. 539-562.
 - Judson, H. F. (1979), *The Eighth Day of Creation: The Makers of the Revolution in Biology*, New York: Simon and Schuster.
 - Ledley, R. S. (1955), "Digital Computational Methods in Symbolic Logic, with Examples in Biochemistry", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 41, pp. 498-511.
 - _____ (1987), "Medical Informatics: A Personal View of Sowing the Seeds", *Proceedings of ACM Conference on History of Medical Informatics*, New York: ACM Press, pp. 31-41.
 - Levinthal, C. (1959), "Coding Aspects of Protein Synthesis", *Reviews of Modern Physics*, Vol. 31, pp. 249-255.

- Metropolis, N. (1990), "The Los Alamos Experience, 1943-1954", in S. G. Nash (ed.), *A History of Scientific Computing*, New York: ACM Press, pp. 237-250.
- Nirenberg, M. W. and Matthaei, H. (1961), "The Dependence of Cell-Free Protein Synthesis in E. Coli Upon Naturally Occurring or Synthetic Polyribonucleotides", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 47, pp. 1588-1602.
- Park, Hyung Wook (2010), "Longevity, Aging, and Caloric Restriction: Clive Maine McCay and the Construction of a Multidisciplinary Research Program", *Historical Studies in the National Science*, Vol. 40, pp. 79-124.
- Rich, A. and Watson, J. D. (1954), "Some Relations Between DNA and RNA", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 40, pp. 759-764.
- Rossini, F. A. and Porter, A. L. (1984), "Interdisciplinary Research: Performance and Policy Issues", in R. Jurkovich and J. Paelinck (eds.), *Problems in Interdisciplinary Studies*, Aldershot: Gower.
- Saari, E. and Miettinen, R. (2001), "Dynamics of Change in Research: Constructing a New Research Area: in a Research Group", *Science, Technology, & Human Value*, Vol. 26, pp. 300-321.
- Simonton, D. K. (1984), "Is marginality Effect All That Marginal?", *Social Studies of Science*, Vol. 14, pp. 621-622.
- _____ (2004), *Creativity in Science: Change, Logic, Genius, and Zeitgeist*, Cambridge: Cambridge University Press.
- Sinsheimer, R. L. (1959), "Is the Nucleic Acid Message in a Two-Symbol Code?", *Journal of Molecular Biology*, Vol. 1, pp. 218-220.
- Watson, J. D. (2001), *Genes, Girls, and Gamow: After the Double Helix*, New York: Vintage Books.

- Watson, J. D. and Crick, F. H. C. (1953a), "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid", *Nature*, Vol. 171, pp. 737-738.
- _____ (1953b), "Genetical Implications of the Structure of Deoxyribonucleic Acid", *Nature*, Vol. 171, pp. 964-967.
- Woese, C. R. (1967), *The Genetic Code: The Molecular Basis for Genetic Expression*, New York: Harper and Row.
- Yčas, M. (1956), "The Protein Text", in H. P. Yockey (eds.), *Symposium on Information Theory in Biology*, New York: Pergamon Press, pp. 70-102.
- _____ (1969), *The Biological Code*, New York: Elsevier North-Holland.


논문 투고일	2017년 5월 4일
논문 수정일	2017년 5월 26일
논문 게재 확정일	2017년 6월 11일

Deciphering the Genetic Code in the RNA Tie Club: Observations on Multidisciplinary Research and a Common Research Agenda

Kim, Bong-kook

ABSTRACT

In 1953, theoretical physicist George Gamow attempted to explain the process of protein synthesis by hypothesizing that the base sequence of DNA encodes a protein's amino acid sequence and, in response, proposed the nucleic acid-protein information transfer model, which he dubbed the "diamond code." After expressing interest in discussing the daring hypothesis, contemporary biologists, including James Watson, Francis Crick, Sydney Brenner, and Gunther Stent, were soon invited to join the RNA Tie Club, an informal research group that would also count biologists and various researchers in physics, mathematics, and computer engineering among its members. In examining the club's formation, growth, and decline in multidisciplinary research on deciphering the genetic code in the 1950s, this paper first investigates whether Gamow's idiosyncratic approach could be adopted as a collaborative research forum among contemporary biologists. Second, it explores how the RNA Tie Club's research agenda could have been expanded to other relevant research topics needing multidisciplinary approach? Third, it asks why and how the RNA Tie Club dissolved in the late 1950s. In answering those questions, this paper shows that analyses on the intersymbol correlation of the overlapping code functioned to integrate



diverse approaches, including sequence decoding and statistical analysis, in research on the genetic code. As those analyses reveal, the peculiar approaches of the RNA Tie Club could be regarded as a useful method for biological research. The paper also concludes that the RNA Tie Club dissolved in the late 1950s due to the disappearance of the collaborative research agenda when the overlapping code hypothesis was abandoned.

Key terms | RNA Tie Club, George Gamow, multidisciplinary research, research agenda, genetic code
