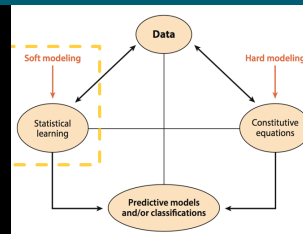
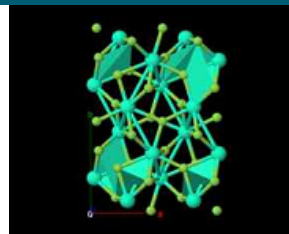
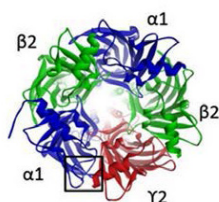
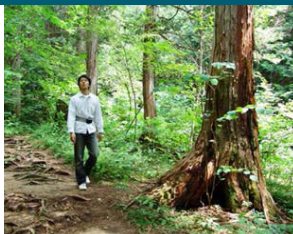
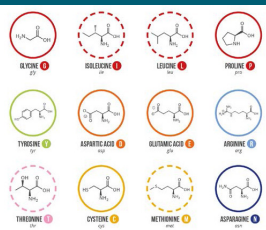


융합연구리뷰

Convergence Research Review

불면증 치료,
피톤치드 연구에서 답을 찾다

R&D 빅데이터,
이제는 연구개발의 경쟁력을 좌우한다



C o n v e r g e n c e R e s e a r c h R e v i e w

Contents

융합연구리뷰 | Convergence Research Review
2017 August vol.3 no.8

- 01 편집자 주
- 04 불면증 치료, 피톤치드 연구에서 답을 찾다
- 28 R&D 빅데이터,
이제는 연구개발의 경쟁력을 좌우한다



발행일 2017년 08월 14일

발행인 하성도

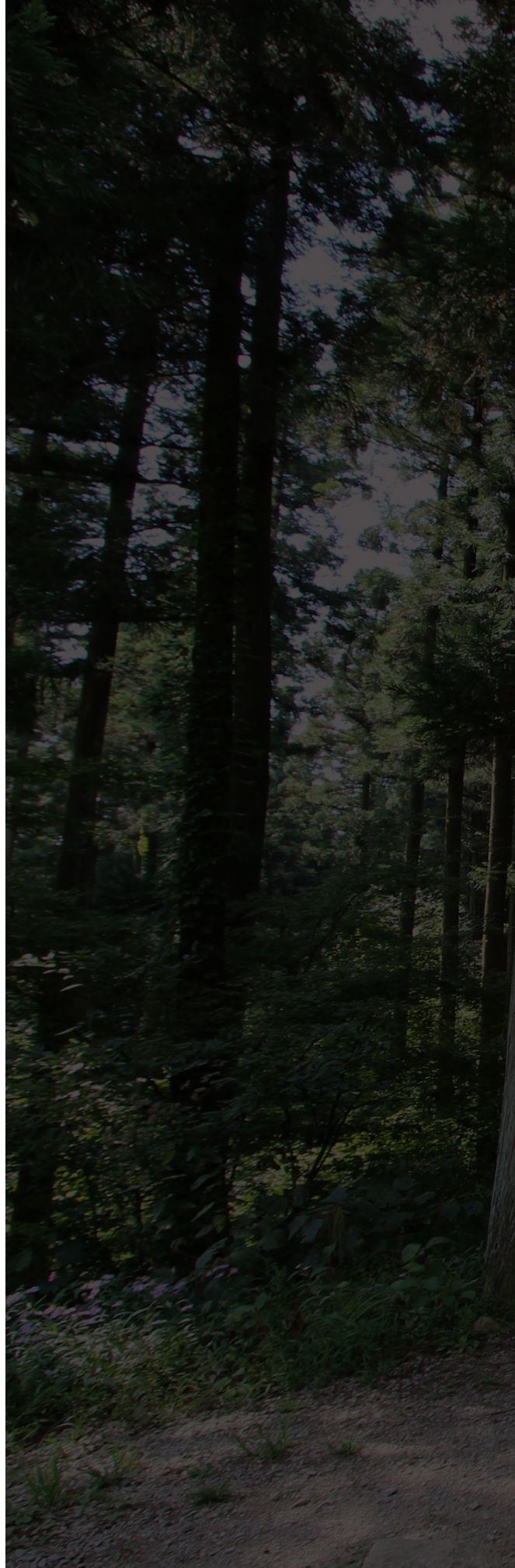
편집인 안주명 김보림

발행처 한국과학기술연구원 융합연구정책센터

02792 서울특별시 성북구 화랑로 14길 5

Tel. 02-958-4984 | <http://crpc.kist.re.kr>

펴낸곳 승일미디어그룹주식회사 Tel. 1800-3673



[편집자주]

| 불면증 치료, 피톤치드 연구에서 답을 찾다

최근 우리사회에서 '수면장애'는 이미 흔한 질환으로써, 현대인의 고질병으로 자리 잡고 있다. 대표적 수면장애 인 불면증은 최근 연구결과 일반인의 1/3이 증상을 경험했다고 응답할 정도로 빈번히 발생하고 있으며, 그 원인은 다양한 정신적 스트레스(입시, 취업, 업무, 퇴직, 노후 등)가 큰 것으로 알려져 있다. 이러한 수면장애는 단순히 잠이 안 오는 문제에서 나아가 신체적·정신적으로 심각한 질병을 야기할 수 있다. 수면장애를 치료하기 위해서는 보통 장기적인 수면장애의 경우에는 심리치료, 행동치료 등 다양한 방법을 시도하나, 단기적인 경우에는 수면제를 이용한 약물치료가 일반적으로 이루어지고 있다. 그러나 대부분의 수면제들은 부작용과 내성이 있어 사용에 주의가 필요하다. 이러한 이유로 최근에는 약물치료 외에 스트레스를 완화하기 위한 시도들이 이루어지고 있다. 그 중 하나로 삼림욕과 등산을 통해 심신의 안정을 찾고자하는 현대인들이 늘어나고 있다. 실제로도 삼림욕 시에 침엽수들이 만들어내는 피톤치드가 심신의 안정을 위한 진정효과가 있다고 알려져 있다. 그동안 이러한 피톤치드의 진정효과는 잘 알려져 있음에도 불구하고 정확한 메커니즘(작용기전)이 밝혀지지 않았으나, 최근 KIST 연구팀이 이를 과학적으로 규명하는데 성공했다.

이에, 본 호 1부에서는 KIST 이창준 박사 연구팀을 통해 피톤치드의 과학적 효과에 대해 살펴보고자 한다. 이러한 연구를 토대로 향후 천연물 기반 수면제가 개발되어 현대인의 수면장애 치료에 새로운 지평을 열길 기대해 본다.

| R&D 빅데이터, 연구개발 경쟁력을 좌우한다

최근 4차 산업혁명의 도래로 우리사회는 초연결, 초지능 사회로 진화하고 있다. 이러한 4차 산업시대에는 빅데이터를 바탕으로 가상공간과 물리공간의 결합에 의한 생산과 서비스가 이루어지고 있다. 빅데이터는 이렇듯 오늘날 21세기 원유라고 불릴 만큼 중요한 개념으로써 사회 전반의 여러 분야로 활용이 확대되고 있다. 빅데이터란 디지털 환경에서 생성되는 기존 데이터보다 방대하고, 생성 주기도 짧고, 다양한 형태를 포함하는 대규모 데이터를 의미한다. 이는 단순히 많은 양의 데이터를 수집하는 것에 의미를 두기보다는 수많은 빅데이터를 어떻게 먼저 수집하여 잘 활용하느냐가 관건이라고 할 수 있다. 특히 R&D 분야에서 이러한 빅데이터를 기반으로 한 연구 플랫폼 개발을 통해 기존의 연구방식을 바꾸는 새로운 움직임이 활발히 일어나고 있다. 그 동안 R&D 분야에서 컴퓨터를 활용하여 과학이나 공학적 문제들을 수치적으로 해석함으로써 보조적인 역할을 해왔던 계산과학이 빅데이터와 만나 새로운 연구 영역을 창출하고 있는 것이다. 연구를 통해 모아진 데이터 속에서 의미 있는 지식을 만들어내고, 직접 실험이 불가능한 현상을 컴퓨터 계산을 통해 모사하고 체계적으로 연구하여 더 많은 경우를 예측하고 검증할 수 있게 되었다.

이에 본 호 2부에서는 빅데이터 기반의 계산과학 플랫폼 연구를 진행 중인 KIST 이광렬 박사 연구팀을 통해 R&D 빅데이터 연구에 대해 살펴보고자 한다. 향후 R&D 빅데이터 연구 플랫폼을 통해 더욱 고도화된 연구성과가 창출될길 기대해 본다.

융합연구리뷰

Convergence Research Review 2017 August vol.3 no.8

<http://crpc.kist.re.kr>

01

불면증 치료, 피톤치드 연구에서 답을 찾다

한국과학기술연구원 신경교세포연구단
이창준 박사, 우준성 박사
(cjl@kist.re.kr, junssam@kist.re.kr)





01 서론

잠이 없는 삶을 생각해 볼 수 있을까? 시험 공부를 위해 며칠 밤을 지새거나 따뜻한 봄날 점심 식사 후의 나른함 속에서 일을 해본 경험이 있다면 수면의 중요성을 구태여 말하지 않아도 공감할 것이다. 수면(垂面)이란 자연스럽게 반복되는, 무의식 상태에서 휴식을 취하는 행위라고 정의된다. 수면은 모든 포유류와 조류, 다수의 파충류, 양서류, 어류에서 발견되므로, 사람을 포함하는 생명체가 정상적인 삶을 살기 위한 필수 요소라 말하고 싶다. 그런데 이러한 수면에 문제가 있으면 어떻게 될까? 수면장애는 인구의 약 20% 이상이 경험한 적이 있거나 앓고 있는 매우 흔한 질환이다. 습관적으로 잠을 이루지 못하며, 짧고 단속적인 얇은 수면이나 꿈을 많이 꾸는 불면증이 대표적인 수면장애 증세이다. 해마다 늘어가는 수면장애 인구와 비례하여 수면, 수면장애 치료 관련 시장은 꾸준히 늘어가고 있다.

가장 흔한 수면 장애인 불면증은 일반인의 약 1/3이 반복되는 증상을 경험하고, 9%가 매일의 일상생활에서 불면증 때문에 괴로움을 느낀다. 장기적인 불면증의 경우 의사의 지시와 처방에 따른 치료가 필요하다. 장기적 불면증 치료는 행동치료, 심리치료, 빛(光) 치료 등 다양한 방법이 있으나 가장 흔한 것은 수면제를 처방 받아 복용하는 약물치료가 있다. 그러나 대부분의 수면제들은 부작용과 내성(tolerance)이 발생하는 등 그 사용에 주의가 필요하다. 대부분의 단기적인 불면증의 경우는 그 원인이 스트레스에 있으므로 심신의 안정을 비롯한 스트레스 완화가 그 치료책이다.

수목이 울창한 산 속에서 상쾌한 공기를 들이마시며 숲 속을 걷거나 머무르는 산림욕(森林浴)은 스트레스 완화의 자연친화적인 방법으로 최근 들어 각광받고 있다. 산림욕이라는 말은 1982년 일본 농림수산부가 만들었으며, 영어로는 forest-bathing 이라고 한다. 산림욕의 효과를 입증하기 위한 많은 연구들 중 대표적인 것은 2010년 일본 연구진에서 발표한 '산림욕의 생리학적 효과'이다. 이 연구에 따르면 성인 남성을 15분 간 숲 속에서 걷거나 머무르게 했을 때, 스트레스 호르몬이 감소하고 맥박과 혈압도 감소하는 스트레스 완화 및 안정효과가 나타났다¹⁾ <그림 1>. 그러나 이러한 효과는 나무들이 배출하는 어떤 물질이 인체에 들어간다는 추측뿐, 산림욕이 어떻게 이러한 효과들을 나타내는지에 관해서는 아직 명확히 밝혀지지 않았다.



그림 1. 숲 속 산책과 감상을 통한 산림욕¹⁾

이런 물질의 근거는 시간을 거슬러 고대 동아시아의 문헌을 통해 엿볼 수 있다. 중국 송나라 또는 명나라 시대에 편찬된 것이라 추정되는 가장 오래된 생약학 서적 신농본초경(神農本草經, 저자 미상)에 따르면, “즉 백나무의 씨앗(백자인, 柏子仁, <그림 2>)은 가슴 두근거림을 치료하고 오장을 안정시키며 기를 보하고 습비(濕痺)를 제거한다”고 하였다. 조선 시대 1596년에 허준에 의해 집필된 동의보감(東醫寶鑑)에서도 백자인의 효능을 비슷하게 언급하고 있으며, 불면증 치료에 백자인을 넣어 처방하였다. 또한 중의학과 한의학에 쓰이는 약초, 동물, 광물의 쓰임새를 집대성해 1977년 중국에서 편찬된 중약대사전(中藥大辭典)에서도, “백자인은 심을 보양하고 정신을 안정시키며 장을 윤희하게 하고 대소변을 통하게 하는 효능이 있어 경계, 불면증, 유정, 도한(盜汗), 변비를 치료한다”라고 나와있다. 산림욕 자체는 아니지만, 숲에서 흔히 볼 수 있는 나무의 성분을 불면증 치료에 사용한 예라 할 수 있겠다.



그림 2. 측백나무 씨인 백자인 (이미지 출처: <http://www.google.com>)

그렇다면 산림욕과 백자인을 통해 볼 수 있는 나무의 어떤 성분이 스트레스를 완화하고 불면증을 치료할 수 있을까? 1937년에 러시아 레닌그라드 대학(현재 상트페테르부르크 대학)의 생화학자인 토크인(Boris P. Tokin)은 식물 유래의 유익한 성분을 피톤치드(Phytoncide)라 이름 지었다. 피톤치드는 희랍어로 ‘식물의’라는 뜻의 ‘phyton’과 ‘죽이다’라는 뜻의 ‘cide’가 합해서 생긴 말로써, 식물이 병원균, 해충, 곰팡이에 저항하기 위해 분비하는 물질이다. 향신료, 양파, 마늘, 찻잎, 참나무, 소나무를 비롯한 여러 식물체들이 피톤치드를 분비한다. 피톤치드는 기본 기능인 항균작용을 비롯하여 스트레스 해소, 심폐기능 강화 등의 효과가 알려지면서 산림욕과 더불어 최근에 각광받고 있다. 그러나 피톤치드의 어떤 성분이 어떻게 작용하여 이러한 다양한 효과들을 나타내는지는 아직 명확하게 밝혀져 있지 않다. 피톤치드는 그 유래에 따라 다양한 물질들이 있으나, 본 리뷰에서는 산림욕과 백자인을 통해 우리나라 산림에서 가장 큰 비중을 차지하고 있는 침엽수인 소나무 유래의 피톤치드의 불면증 치료 또는 수면에서의 역할을 수면 행동학적, 전기 생리학적, 구조화학적 접근을 통해 과학적으로 밝히고자 한다.

02 본론

2.1 소나무 유래의 피톤치드 : 피넨

소나무는 소나무과의 늘 푸른 바늘잎 큰키나무로 전세계적으로 넓게 분포해 있고, 아시아에서는 우리나라와 중국 북동부, 일본 등에 분포해 있다. 우리나라에서는 산림의 약 37%로 가장 높은 비중을 차지하고 있으며(한국 산림자원평가, 산림청, 2013 통계), 특히 산림욕을 하는 침엽수림에서 가장 흔하게 접할 수 있는 품종이다. 소나무는 그 독특하며 향긋한 향과 의학적-약리학적으로 유용한 특성들 때문에 예로부터 지금까지 동서양을 막론하고 실생활에 많이 쓰여져 왔다. 이러한 소나무의 특성들은 소나무의 식물성 기름 또는 정유(精油, essential oil) 때문이다. 식물성 정유는 다양한 약리학적 특징들을 가지는데 그 대표적인 것으로는 항염증, 항균, 통증과 스트레스 완화 등이 있다. 식물성 정유의 주성분은 피넨(pinene), 카렌(careen), 리모넨(limonene), 테르피넨(terpinene) 등을 포함하는 탄소화합물인 테르펜(terpene)이다 <그림 3>.

테르펜은 일반식 $(C_5H_8)_n$ ($n \geq 2$)을 가지는 탄수화물 및 이들 유도체의 총칭으로, 구조상으로 아이소프렌(isoprene) 분자가 기본단위인 아이소프렌 법칙을 따른다. 일반식에서 $n=2$ 인 것, 즉 $C_{10}H_{16}$ 의 조성을 갖는 화합물은 모노테르펜(monoterpene), $n=3$ 인 것을 세스퀴테르펜(sesquiterpene), $n=4$ 인 것을 다이테르펜(diterpene), $n=6$ 인 것을 트라이테르펜(triterpene)이라고 한다 <그림 4>.

소나무 정유에서 가장 대표적인 테르펜은 피넨의 광학이성질체¹⁾ 중 하나인 알파-피넨(α -pinene, 2,6,6-trimethylbicyclo(3.1.1)-2-hept-2-ene)이다. 알파-피넨은 에틸렌계 탄화수소로 4개의 구조 고리를 가지고 있으며, 테르펜이 가지는 항균, 항염증과 같은 특징 외에 유용한 특징들을 가지며, 소나무 피톤치드의 주성분이라 할 수 있겠다. 그렇다면 소나무 유래의 피톤치드인 알파-피넨에 주목하여 수면에서의 그 효과와 작용에 대해 살펴 보자.

i) 구조식이 같으며 분자 중의 원자배열이 좌우 손바닥과 같이 대칭관계가 있는 이성체



① 불면증 치료, 피톤치드 연구에서 답을 찾다










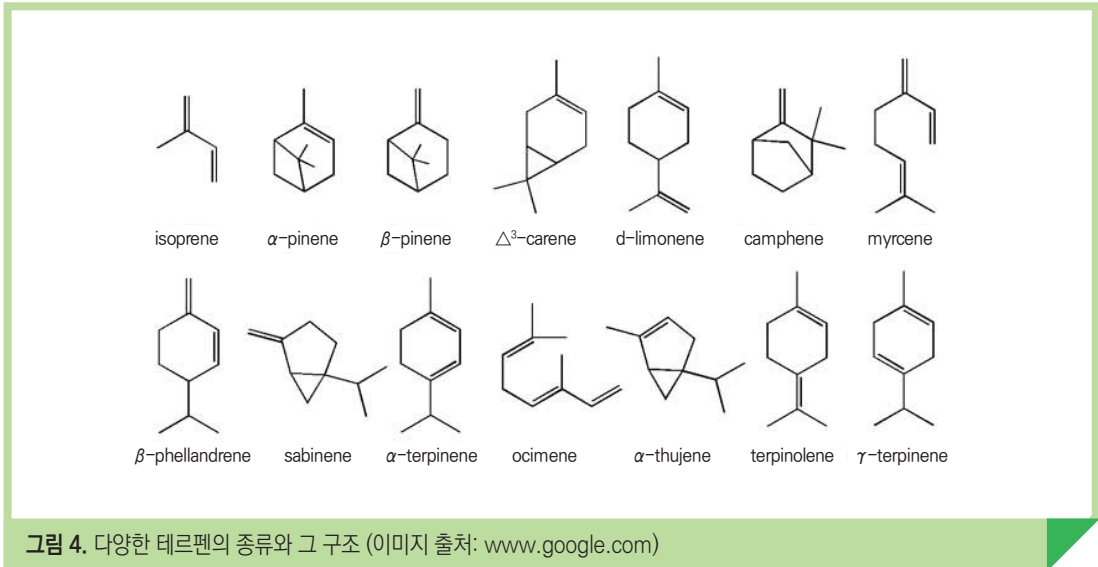
	TERPENE	BENEFIT	AROMA
	Pinene Also found in pine needles	Anti-inflammatory Anti-bacterial Bronchodilator Aids memory	Pine Earth
	Myrcene Also found in hops	Sedative Sleep aid Muscle relaxant	Flowers Pungent Earth
	Limonene Also found in citrus	Treats acid reflux Anti-anxiety Antidepressant	Citrus Fresh spice
	Terpinolene Also found in coriander	Analgesic Pain reduction Digestive aid Stomachic	Pine Herbal Anise Lime
	Linalool Also found in lavender	Anesthetic Anti-convulsive Analgesic Anti-anxiety	Flowers Lavender Citrus Fresh spice
	Terpineol Also found in mugwort	Calming aid Antibacterial Antiviral Immune system	Pleasant lilac Citrus Wood
	Caryophyllene Also found in black pepper	Anti-inflammatory Analgesic Protects cells lining Digestive tract	Citrus Spice
	Humulene Also found in basil	Anti-inflammatory	Robust Herbaceous Earth
	Ocimene Also found in thyme and alfalfa	Decongestant Antiseptic Antiviral Bactericidal	Citrusy green Wood Tropical fruit

그림 3. 여러 식물 유래의 테르펜의 종류 그 효과 (이미지 출처: PHYTATECH™)



2.2 EEG와 EMG를 이용한 알파-파의 수면 증진 효과 분석

동물 또는 사람의 생리적 활동 상태는 크게 잠이 들었을 때(수면상태, 睡眠狀態, sleep)와 깨어있을 때(각성상태, 覺醒狀態, awake)로 구분할 수 있다. 수면상태는 크게 렘수면(REM sleep: rapid eye movement sleep)과 비(非)-렘수면(NREM sleep: non-rapid eye movement sleep)으로 나눌 수 있다. 렘수면과 비-렘수면 단계에서 나타나는 확연한 차이점은 그 이름에서 알 수 있듯, 각각의 수면 상태 동안 일어나는 안구의 움직임이다. 비 렘수면 동안에는 안구의 움직임이 없는데 반해 렘수면 동안에는 안구가 좌우로 움직인다. 렘수면과 비-렘수면으로 구분되는 수면의 단계는 1953년 미국의 디멘트 박사(William C. Dement)에 의해 처음 제시되었다. 수면의 단계를 구분하는 기준으로는 크게 3가지 방법이 쓰인다. 첫 번째는 뇌를 둘러싸고 있는 두피 표면에 전극을 부착하여 신경세포들이 발생시키는 전기적 신호를 측정하는 것으로, 이 방법은 ‘뇌전도(腦電圖, EEG: electroencephalogram)’라고 부른다. 뇌전도에서 측정하여 나타내는 뇌의 활동, 즉 뇌세포가 발생시키는 전기적 신호를 ‘뇌파(腦波, brain wave)’라고 부르며, 두피에서 뇌파를 측정하는 뇌전도 기법은 1937년 미국의 루미스 박사(Alfred Lee Loomis)가 처음 개발했다. 뇌전도 측정은 수면 단계를 구분하는 기준으로 가장 많이 쓰인다. 뇌파는 그 주파수와 진폭에 따라 알파(alpha, α), 베타(beta, β), 델타(delta, δ), 세타(theta, θ)파 네 종류로 나뉘며, 각각의 뇌파는 수면 상태에 따라 다르게 나타난다 <그림 5>.



① 불면증 치료, 피톤치드 연구에서 답을 찾다

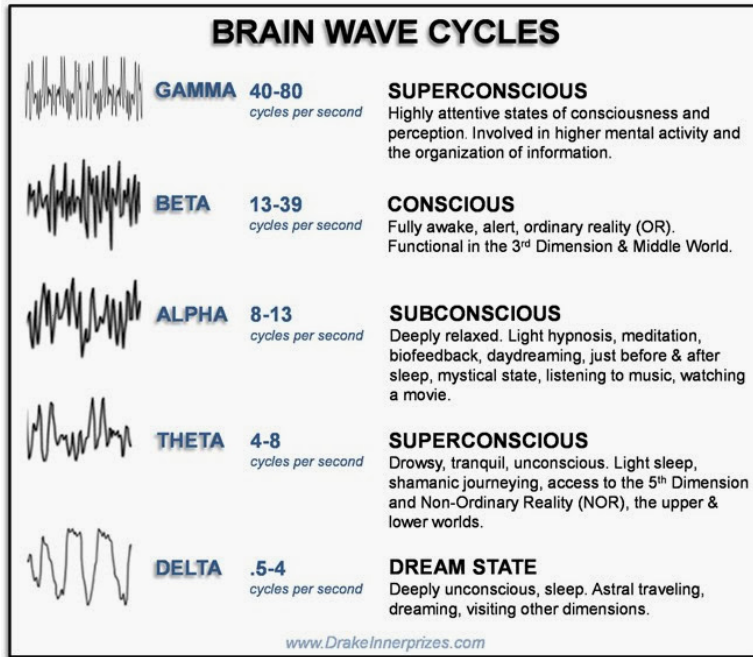


그림 5. 주파수와 수면 단계에 따른 뇌파의 종류 특징 (이미지 출처: www.Drakeinnerprizes.com)

다음으로는 골격근의 활동전위(action potential)를 세포외전극(extracellular electrode)을 이용하여 측정하는 근전도(筋電圖, EMG: electromyogram)이다. 근전도는 수면의 측정뿐 아니라 골격근(skeletal muscle), 신경근접합부(neuromuscular junction) 운동 뉴런(motor neuron) 기능의 생리학적 연구 및 그들 부위가 관련하는 신경, 근질환 진단에도 사용된다. 마지막으로 안구의 움직임을 나타내는 전기신호인 안구전위도(眼球電位圖, EOG: eletrooculogram) 또한 이용된다. 이 세 가지 방법 <그림 6>을 조합하여 분석하면 잠을 자는 동안 일어나는 뇌와 신체 변화에 대한 정보를 거의 대부분 알아낼 수 있다.

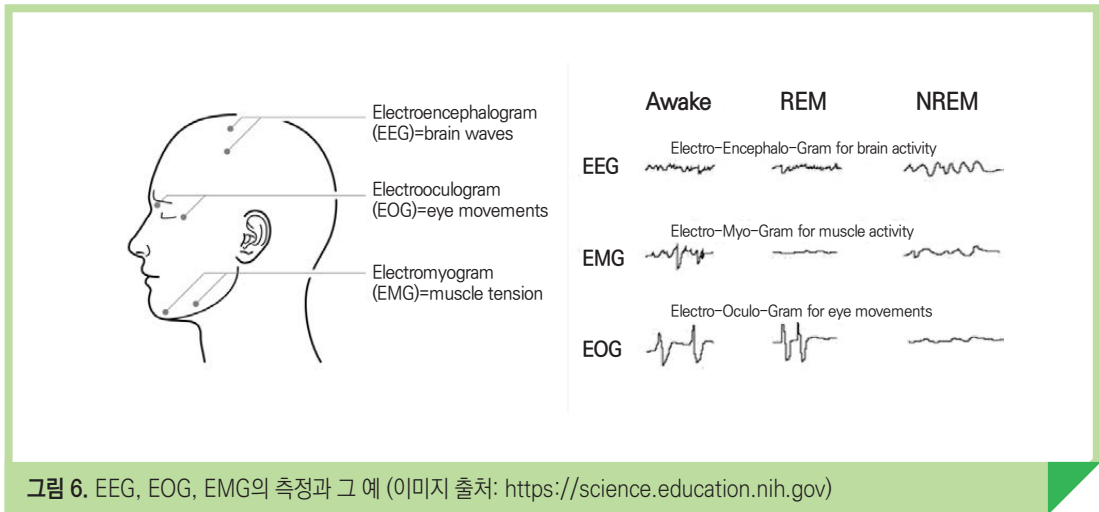


그림 6. EEG, EOG, EMG의 측정과 그 예 (이미지 출처: <https://science.education.nih.gov>)

앞서 언급한대로 수면의 단계는 크게 렘수면과 비-렘수면으로 나눌 수 있고, 사람에서 평균적으로 렘수면은 전체 수면시간의 25%, 비-렘수면은 75%를 차지한다고 알려져 있다. 수면 동안에 각 단계들은 반복되어 나타나는데, 이런 반복의 순서나 각 단계별 지속 시간은 개인별로, 상황별로 매우 다양하게 나타난다. 이러한 수면 단계의 반복성을 나타내는 것을 '수면곡선(Sleep Hypnogram)'이라 한다 <그림 7>.

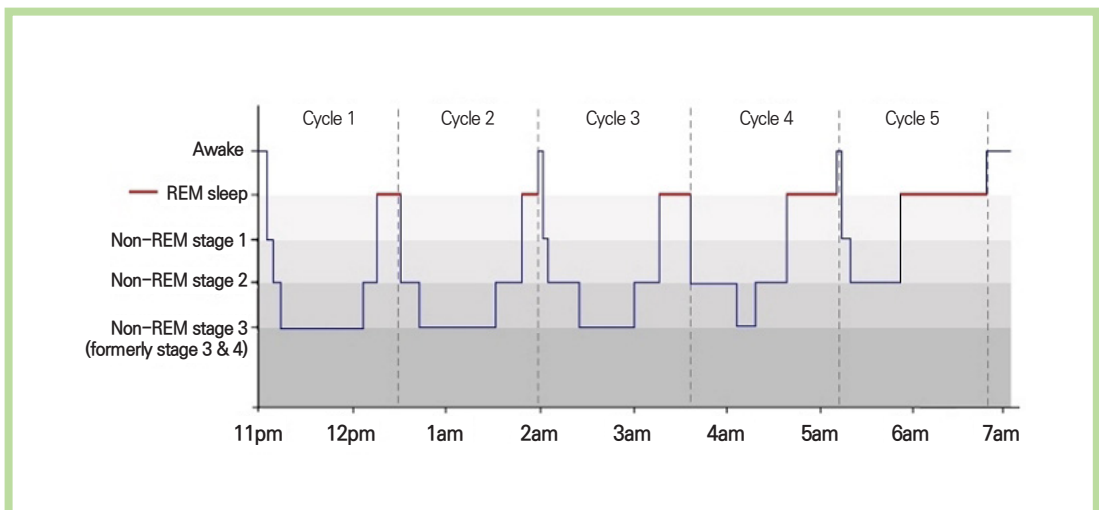
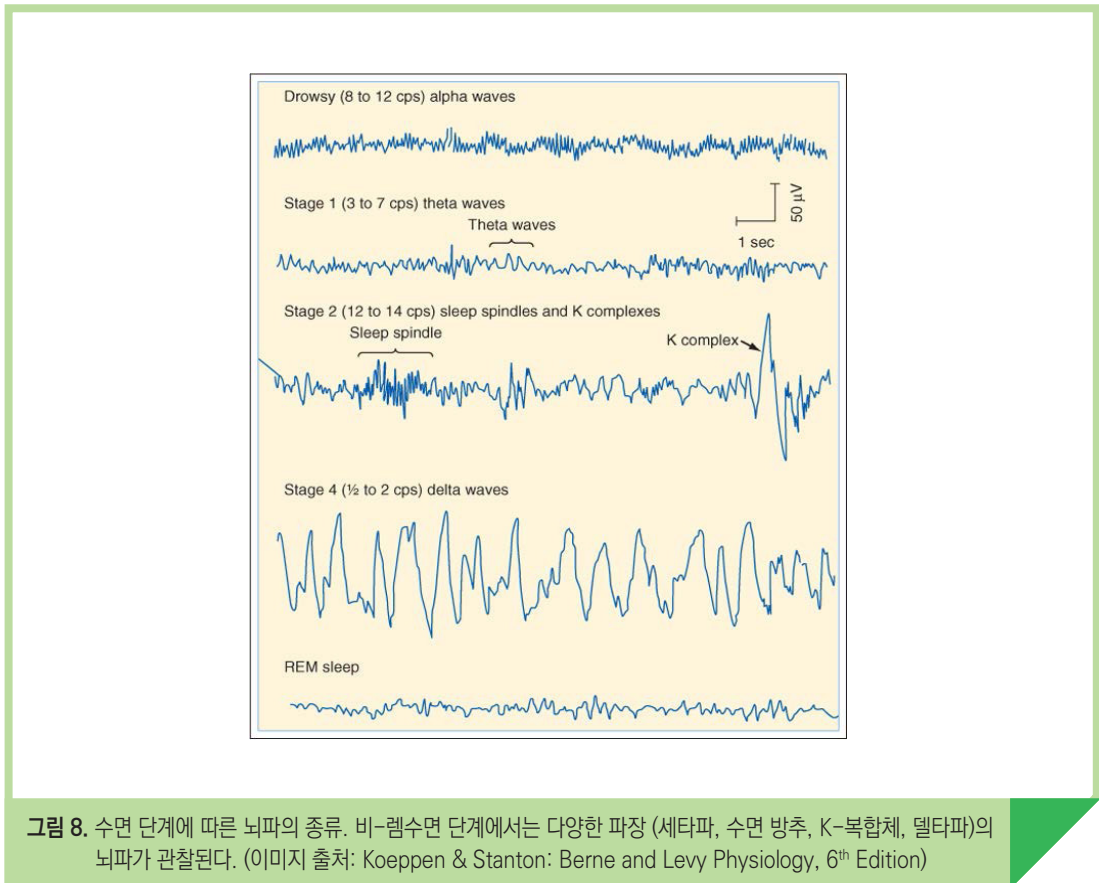


그림 7. 사람에서의 수면 단계와 주기를 나타낸 수면곡선 (이미지 출처: Luke Mastin)



렘수면 또는 급속 안구 운동 수면은 깨어 있는 것에 가까운 얇은 수면이며, 안구의 빠른 운동에 의해 구분된 수면의 한 단계이다. 렘수면은 1950년 애서린스키(Eugene Aserinsky)와 클라이트만(Nathaniel Kleitman)에 의해 정의되었다. 렘수면 단계에 나타나는 뇌파의 양상은 각성상태일 때와 비슷하게 진폭이 작고 주파수가 큰 특성을 보인다. 이 단계는 잠의 깊이가 얇은 상태로 호흡률과 뇌의 활성이 증가하고 잠이 깨기도 쉽다. 반면 근육은 이완되면서 몸의 움직임이 일어나지 않는다. 꿈을 꾸는 단계가 바로 렘수면 단계로, 이 단계 동안 좌우로 움직이는 안구의 운동이 꿈을 꾸면서 느껴지는 시각적인 감각을 발생시킨다고 생각되기도 한다. 성인의 렘수면은 일반적으로 총 수면의 20~25%로 발생하며, 밤시간 수면의 90~120분 정도이다. 보통 밤 수면 동안 사람은 흔히 5번의 렘수면 단계를 경험한다. 렘수면의 양은 나이와 관련되며 갓난 아이는 총 수면의 80%가 렘수면이다.

안구 운동이 관찰되지 않는 비-렘수면은 뇌파의 특성에 따라 4단계로 나뉜다. 1단계는 막 잠이 들었을 때로, 이 때의 뇌파는 알파파보다 주파수는 더 느려지고 진폭은 더 큰 형태로 바뀐다. 이런 형태의 뇌파는 세타파라고 불린다. 지속 시간은 보통 5~10분 정도라고 알려져 있다. 두 번째 단계는 더 깊은 잠을 향해가는 단계로, 빠르고 규칙적인 다발성의 뇌파가 발생한다. 이 같은 형태의 뇌파는 ‘수면 방추(sleep spindle)’라고 한다. 또한 이 단계에서는 K 복합체(K-complex)라고 불리는 뇌파가 나타나기도 한다. 수면 방추는 다발성으로 나타나나, K 복합체는 일회성으로 갑작스럽게 나타나는 진폭이 큰 뇌파를 가리킨다. 2단계 비-렘수면은 한 번에 20분 정도 지속되며 잠을 자는 시간 전체에서 차지하는 비중이 40~45%로 가장 높다. 이 단계 동안 체온과 심박수가 서서히 내려가면서 더 깊은 잠의 단계로 향해간다. 마지막 단계에서는 뇌파의 종류 중에서 가장 주파수가 느리고 진폭이 큰 델타파가 관찰된다. 이 단계를 서파 수면(SWS: slow wave sleep)이라 한다 <그림8>. 이 단계는 가장 깊은 잠이 들었을 때로, 잠을 자는 사람은 소음을 비롯한 주변 환경에서 오는 자극을 감지하기 힘들고 잠을 깨우기도 어렵다. 이 같은 뇌파의 델타파를 수면의 깊이 또는 질을 나타내는 척도로 삼기도 한다²⁾.



알파-피넨의 수면에서의 역할은 2005년 일본 연구진에 의해 처음 소개되었는데, 실험용 쥐에 알파-피넨의 향기를 맡게 했을 때, 렘수면의 지속 시간(sleep latency)이 길어지는 것을 확인했다³⁾. 또한 최근 우리나라의 연구에서는 실험용 생쥐에 직접 알파-피넨을 투여한 결과, 낮은 농도 (25 mg/kg 이상)에서는 진정 작용을 보였으며, 높은 농도 (100 mg/kg)에서는 비-렘수면 시간이 증가하는 수면개선 효과를 나타내었다⁴⁾. 이런 효과는 현재 전세계적으로 가장 많이 쓰이는 수면제인 졸피뎀(zolpidem)을 투여하였을 때에도 나타났다. 여기서 흥미로운 사실은 졸피뎀의 경우 수면의 질과 비-렘수면의 강도를 나타내는 뇌파인 델타파 (frequency range: 0.5-2.5 Hz)의 활성을 저해했으나, 알파-피넨은 델타파의 저해 없이 수면개선 효과를 나타내었다 <그림 9>.

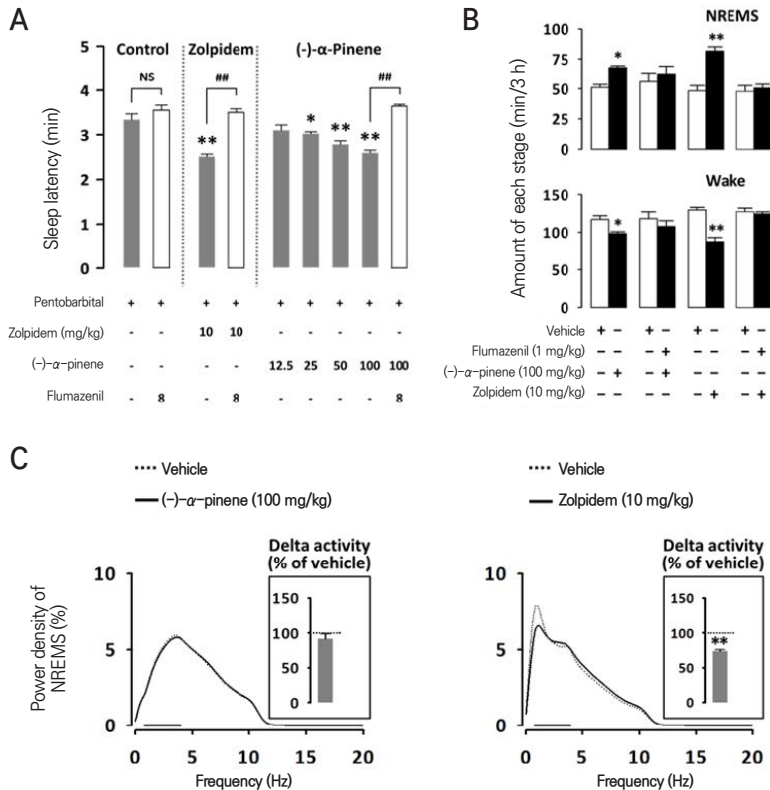


그림 9. 알파-피넨의 수면 개선 효과. 알파-피넨은 수면의 질 저하 없이 비-렘수면의 시간을 늘려 수면 개선 효과를 나타냈다. (A. 졸피뎀과 알파-피넨의 진정작용 B. 알파-피넨의 비-렘수면 증진효과 C. 알파-피넨과 졸피뎀의 델타파에 미치는 영향)⁴⁾

수면제의 수면 질 저하를 동반하는 수면개선 효과는 졸피뎀과 함께 많이 쓰이는 디아제팜(diazepam)의 경우에서도 나타난다. 이들 수면제는 모두 잘 알려진 가바 A형 수용체(GABA_A receptor, GABA: gamma(γ)-aminobutyric acid, <그림 10>)의 벤조디아제핀(benzodiazepine; BZD) 부위에 결합하는 조절제(modulator)들이다. 알파-피넨의 수면개선 효과가 벤조디아제핀 수용체 특이적 저해제(antagonist)인 플루마제닐(flumazenil)에 의해 없어지는 것을 보아, 알파-피넨도 디아제팜과 졸피뎀과 같은 부위에 결합하여 수면 개선을 나타내는 것으로 해석된다.

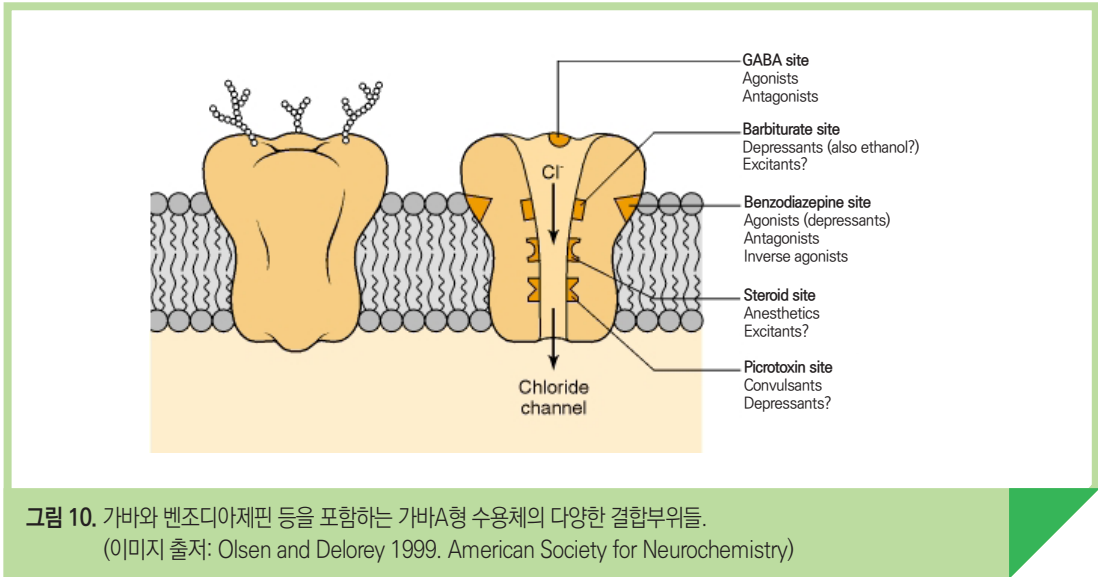


그림 10. 가바와 벤조디아제핀 등을 포함하는 가바A형 수용체의 다양한 결합부위들.
(이미지 출처: Olsen and Delorey 1999. American Society for Neurochemistry)

2.3 패치 클램프 기술을 이용한 알파-피넨 작용기작의 전기생리학적 해석

EEG와 EMG 측정을 통한 수면 분석과 벤조디아제핀 수용체(GABA_A receptor) 저해제(antagonist: flumazenil)와 조절제(modulator: zolpidem)를 사용한 약리학(藥理學, pharmacology)적 방법으로 알파-피넨이 이 수용체에 결합하여 수면 시간을 증가시키는 것이 확인되었다. 그렇다면 알파-피넨이 GABA A형 수용체에 결합하면 어떤 결과가 발생하는 걸까? 이와 같은 질문에 답하기 위해서는 전기생리학(電氣生理學, electrophysiology)적 접근이 필요하다.

전기생리학이란 생체에 대한 전기작용과 생체에서의 전기발생현상을 주요 연구대상으로 하는 생리학의 한 분야이다. 그 연구대상은 단일 세포에서부터 심장이나 뇌와 같은 조직, 생물학적 개체에 이르기까지 다양하며, 주로 측정하는 것은 연구대상들에서의 전위(電位, voltage)와 전류(電流, current)이다. 생체의 전기적 현상은 전기뱀장어와 같은 동물과 관련하여 예로부터 알려졌다며, 1780년 이탈리아의 갈바니(Luigi Galvani)가 개구리 뒷다리의 근육이 전기자극에 의해 수축됨을 발견함으로써 그 존재가 과학적으로 증명되었다. 1952년에는 영국의 과학자 호지킨(Alan Lloyd Hodgkin)과 헉슬리(Andrew Fielding Huxley)가 오징어의 거대 축삭(giant axon)에서 활동 전위(axon potential)와 전위차에 의한 나트륨(Na⁺, sodium)과 칼륨(K⁺, potassium) 전류를 측정하였다 <그림 11>. 이 발견으로 둘은 1963년에 노벨 생리학상을 받았다⁵⁾.

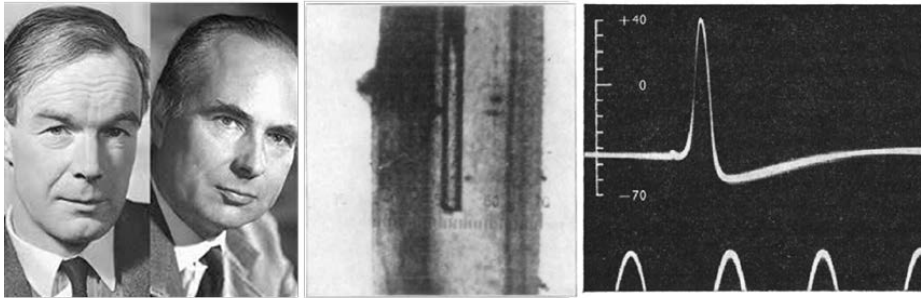


그림 11. 호지킨 & 헉슬리(좌), squid giant axon안의 전극(중간)으로 측정된 action potential(우)

이러한 전기생리학은 1970년대 후반에 독일의 과학자 네어(Erwin Neher)와 자크만(Bert Sakmann)이 세포막의 극히 제한된 부분을 매우 작은 유리 피펫(glass pipette)에 밀착시켜 전위를 고정하고 단일 이온 채널(ion channel)의 전류를 측정하는 패치클램프(patch-clamp) 기술을 발명한 이래로 눈부시게 발전해 오고 있다 <그림 12>. 이들 역시 패치클램프를 통한 신경세포의 전기생리학적 이해를 넓힌 공로로 1991년 노벨 생리학상을 받았다.

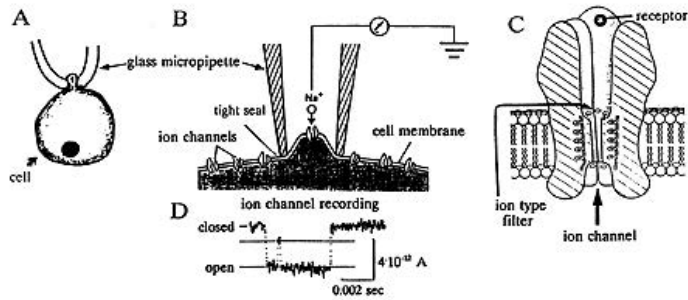


그림 12. Neher와 Sakmann에 의해 발명된 패치 클램프 기술의 모식도와 단일 채널 활성화도 (single channel activity, 이미지 출처: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1991/press.html)

패치 클램프 기술을 이용하면 배양된 세포, 골격근 근세포(myocyte), 뇌 신경세포(neuron) 등 다양한 생리학적 세포들에서 활동 전위를 포함한 전위와 전류의 변화와 같은 전기적인 신호를 측정할 수 있다. 그 중에서도 알파-피넌의 가바 수용체에서의 전기생리학적 의미를 살펴보기 위해 뇌 속 신경세포에서의 역할을 살펴보고자 하자.

신경세포는 신경계의 단위로 기본 기능은 자극을 받았을 경우 전기를 발생시켜 다른 세포로 정보를 전달하는 것이다. 이 때 발생하는 신경세포의 가장 기본적인 전기 신호를 활동전위라고 한다. 뉴런은 구조적으로 크게 핵이 있는 세포체(細胞體, soma), 다른 세포에서 신호를 받는 부분인 수상돌기(樹狀突起, dendrite), 그리고 다른 세포에 신호를 주는 부분인 축삭돌기(逡朔突起, axon) 세 가지 부분으로 나눌 수 있다 <그림 13>. 두 신경세포의 축삭돌기와 수상돌기 사이에 신호를 전달하는 부분을 시냅스(synapse)라고 한다. 신경세포는 그 역할에 따라 감각뉴런(sensory neuron), 연합뉴런(association neuron), 운동뉴런(motor neuron)의 세 종류로 나눌 수 있다. 또한 그 형태에 따라 종류를 나누기도 한다.

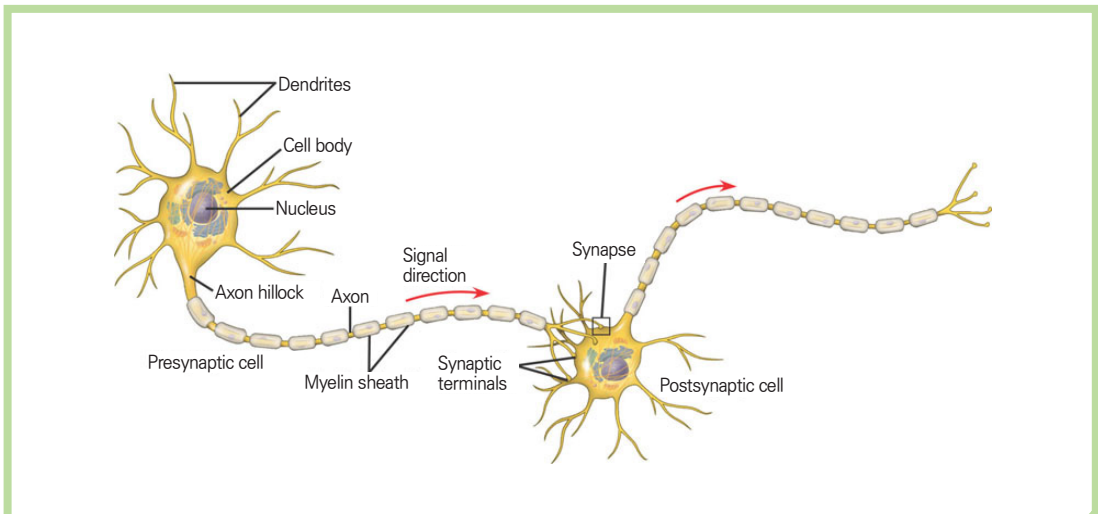


그림 13. Soma, dendrite, axon을 가지는 신경세포(neuron)의 구조 (이미지 출처: www.google.com)

신경세포들 사이의 신경전달과정(synaptic transmission)은 시냅스전세포(pre-synaptic cell)의 활동 전위로 시작되어 칼슘의 유입으로 인한 신경전달 물질(neurotransmitter)의 분비, 신경전달 물질이 시냅

스후세포(post-synaptic cell) 수용체에 결합하여 작용하는 것으로 요약할 수 있다 <그림 14>. 이러한 신경전달과정은 크게 흥분성 신경전달(excitatory synaptic transmission)과 억제성 신경전달(inhibitory synaptic transmission)으로 나눌 수 있다. 이 때 분비되는 대표적인 신경전달 물질은 흥분성인 글루타메이트(glutamate)와 억제성인 가바(GABA)가 있다.

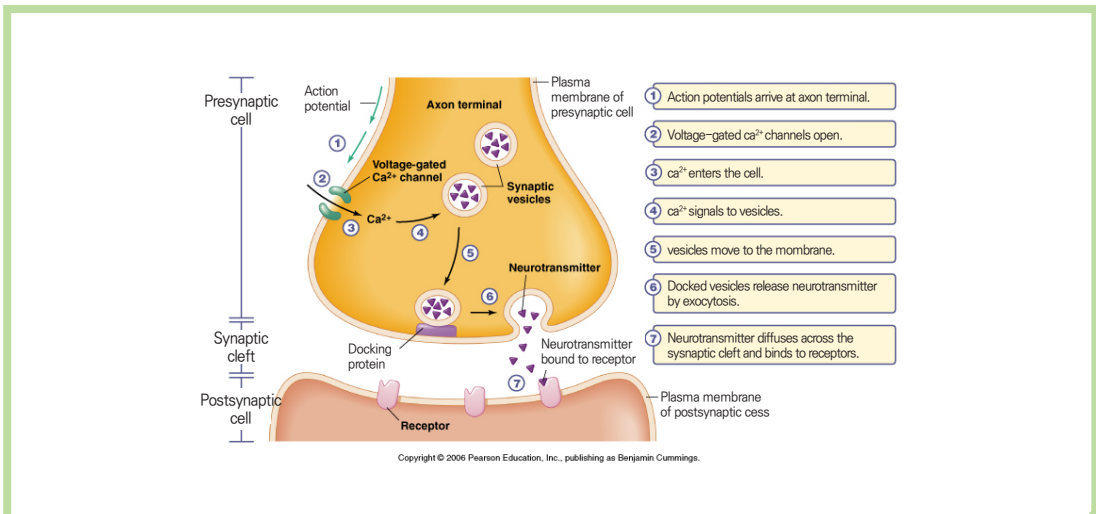
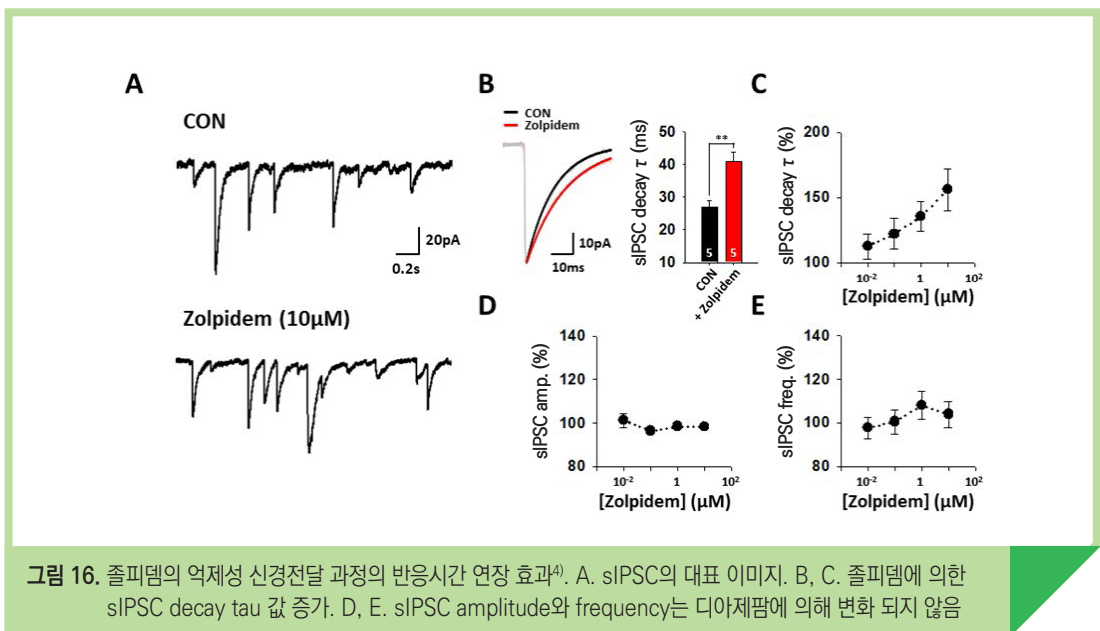
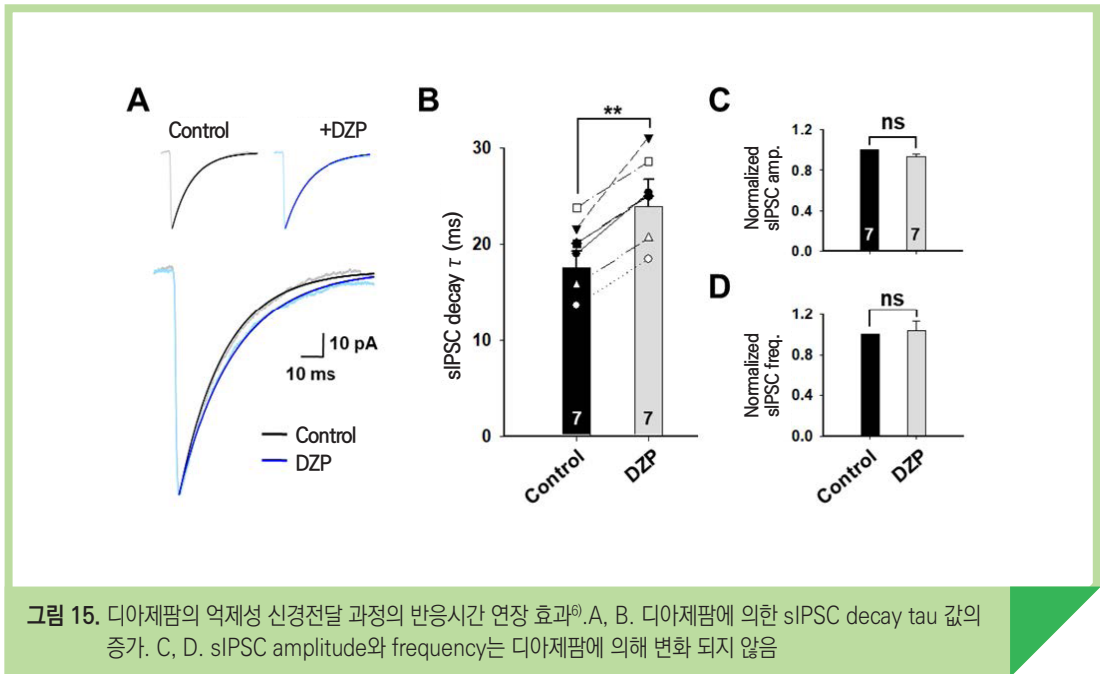


그림 14. 신경전달과정의 모식도 (이미지 출처: Pearson Education. 2006)

최근 연구에 따르면 졸피렘과 디아제팜과 같은 벤조디아제핀 수용체 조절제들은 가바A형 수용체에 결합하여 억제성시냅스후전류(IPSC: inhibitory post synaptic current)의 반응시간을 길게 하는 것으로 보고되었다. 이것은 패치 클램프 데이터의 여러 지표들 중에서 IPSC 반응 시간이 지속되는 것을 나타내는 decay tau(τ) 값ⁱⁱ⁾의 증가가 의미하는 것이다. 그러나 시냅스전세포 신경전달물질 분비와 시냅스후세포 수용체 발현의 변화를 나타내는 IPSC 반응의 발생빈도(frequency)와 크기(amplitude)에는 영향이 없었다^{4),6)} <그림 15.16>.

ii) Decay tau: 어떤 값이 지수적 감쇠 (exponential decay)할 때, 처음 값의 $1/e \approx 0.367879441$ 배가 되었을 때의 시간을 tau 라고 한다



1 ▶ 표본을 치료, 피로치기의 연구에서 표본을 찾다

2 ▶ R&D 표본이 더, 이젠 그 연구는 셀의 전염병에 전이한다

알파-피넨의 경우도 다른 수면제들과 마찬가지로 다른 지표에는 영향을 미치지 않고 IPSC decay τ 값만 증가시켰다 <그림 17>. 이러한 효과는 벤조디아제핀 수용체 억제제인 flumazenil에 의해 사라졌다. 이것은 알파-피넨의 수면 증진효과와 IPSC 반응시간 증가 효과는 졸피뎀이나 디아제팜과 공통된 작용기작을 가짐을 의미한다.

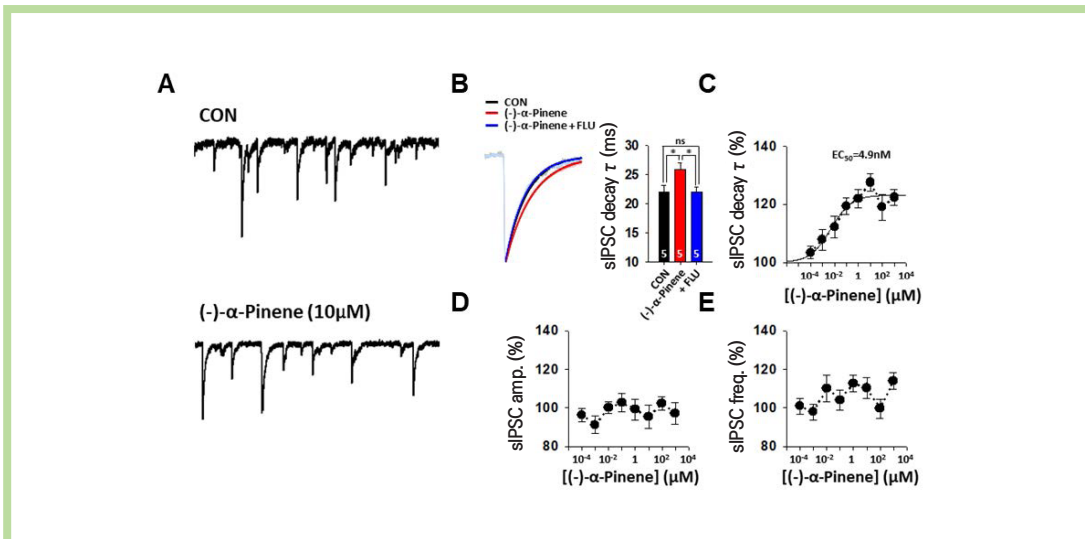
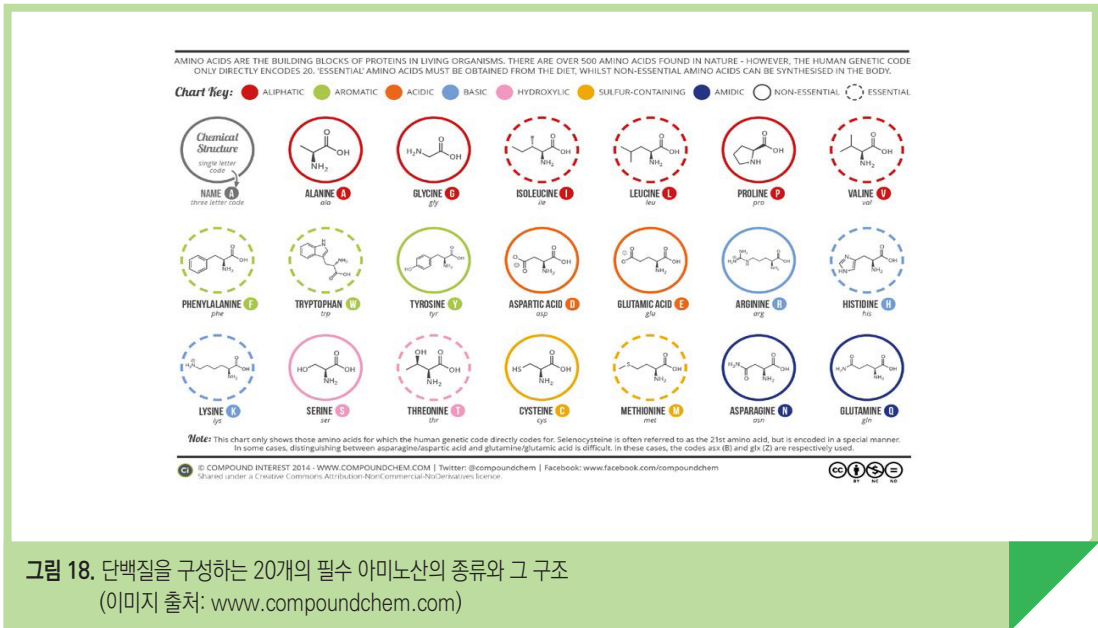


그림 17. 알파-피넨의 억제성 신경전달 과정의 반응시간 연장 효과⁴⁾. A. sIPSC의 대표 이미지 B, C. 알파-피넨에 의한 sIPSC decay tau 값 증가. D, E. sIPSC amplitude와 frequency는 알파-피넨에 의해 변화 되지 않음.

2.4 분자모델링을 통한 알파-피넨의 가바 수용체로의 결합모드와 결합력 예측

알파-피넨의 수면증진효과, 가바에 의한 억제성 신경전달 과정의 연장효과는 모두 알파-피넨이 졸피뎀과 디아제팜과 같은 가바A형 수용체의 벤조디아제핀 부위에 결합함으로써 나타나는 현상임을 앞선 수면 행동학, 약리학, 전기생리학적 방법들로 살펴볼 수 있었다. 하지만 알파-피넨의 수용체로의 결합이 어떠한 형태로 얼마만큼의 결합력을 가지는지에 대해서는 또 다른 관점에서의 접근이 필요하다. 이러한 문제는 알파-피넨을 비롯한 조절체, 수용체와 같은 물질들이 어떠한 단백질 구조를 가지고 있는가를 알면 해결할 수 있다. 이것은 구조생물학(structural biology)의 주요 연구 주제로, 구조생물학은 생물학적 물질이 지니는 분자적 구조가 어떻게 형성되며 그 기능에 영향을 미치는지에 관해 연구하는 분자생물학, 생화학, 생물리학을 포함하는 생물학의 한 분야이다⁵⁾.

생물학적 물질들은 단백질로 구성되어 있는데, 그 원자적 구성은 탄소, 산소, 질소, 수소를 기본으로 하여 다른 미량의 원자들이 작용기(residue, R)에 결합되어 있다. 일반적인 구조식은 R-CH(NH₂)-COOH 이며, 아미노산(amino acid)의 펩티드 결합(peptide bond)에 의하여 연결되어 사슬을 형성한다. 아미노산 서열의 양끝을 N-, C-말단이라고 하며, 이 양 말단에 각각 아미노기(-NH₂)와 카르복실기(-COOH)를 갖는다. 단백질을 구성하는 아미노산은 20종 <그림 18>이 존재하며, 단백질의 종류는 이 아미노산들이 몇 개, 그리고 어떤 순서로 연결되느냐에 따라 달라진다.



단백질에서 이 아미노산의 배열순서를 1차구조(primary structure)라고 한다. 단백질의 1차구조는 1945년 영국의 생화학자(Frederick Sanger)가 51개의 아미노산으로 이루어진 인슐린의 배열을 결정함으로써 확립되었다. 이 1차구조에 의해 형성된 단백질 분자는 구성 아미노산의 분자구조로 인하여 일정한 각도로 구부러지고 꼬여 특정한 형태를 갖추게 된다. 이 형태를 단백질의 2차 구조(secondary structure)라고 한다. 2차구조는 크게 알파나선(α -helix) 구조와 병풍 모양의 베타시트(β -sheet) 구조로 나뉜다. α -helix 구조는 길다란 단백질 분자의 사슬이 나선형으로 꼬여 있는 것으로, 3.6개의 아미노산 분자가 나선의 한 바퀴를 돌게 된다. 아미노산은 다른 아미노산과 수소결합(水素結合, H-bond)로 연결되어 있고, 수소결합 자체는 매우 약

1. 불면증 치료, 피로치근 연구에서 단백질을 찾는다

2. RBD 프레임워크, 이체나, 연구개발의 전환점에 찾아온다

한 결합이지만 알파나선을 구성하는 모든 아미노산이 수소결합으로 연결되어 알파나선은 대단히 안정된 구조이다. β -sheet 구조는 단백질 사슬이 일정한 각도로 꺾여 좁은 종이를 접어 늘인 모양을 하고, 이런 사슬이 여러 개 서로 평행으로 늘어서서 각 사슬의 아미노산이 주위의 다른 사슬 아미노산과 수소결합을 형성한 구조이다. 이 구조는 전체적으로 병풍과 같은 모양이라 병풍 구조라고도 불린다. 이러한 단백질의 2차 구조는 1951년 미국의 폴링(Linus Pauling)이 처음으로 제안하였다. 2차구조를 가진 단백질이 다시 공간적으로 꼬이고, 구부러지고, 접혀서 특정 입체구조를 가지게 되는데, 이것이 3차구조(tertiary structure)이다. 이때 폴리펩티드 사슬(polypeptide chain)이 접히는 과정을 단백질 접힘(protein folding)이라 한다. 3차구조를 가진 단백질 분자들이 둘 또는 그 이상 모여서 하나의 집합체를 구성하여 그 집합체가 비로소 어떤 특정 생물학적 기능을 나타내는 경우가 있는데, 이 집합체를 4차구조(quaternary structure)라고 한다⁵⁾ <그림 19>.

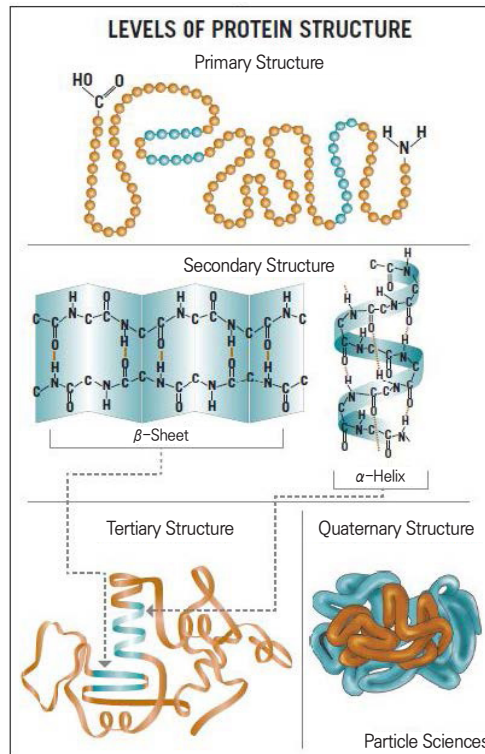


그림 19. 단백질의 단계별 구조들 (이미지 출처: www.particlesciences.com)

알파-피넨이 결합하는 가바A형 수용체는 α , β , γ 3개의 구성요소가 조합을 이루어 5개의 subunit으로 구성된 막관통 단백질(transmembrane protein)이다. 가바A형 수용체의 단백질 3차 구조는 다음과 같다 <그림 20>.

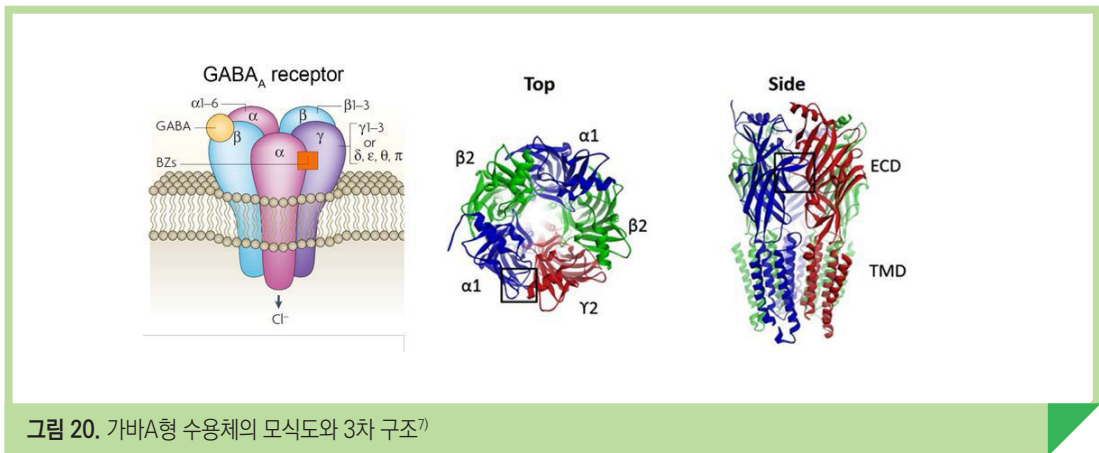


그림 20. 가바A형 수용체의 모식도와 3차 구조⁷⁾

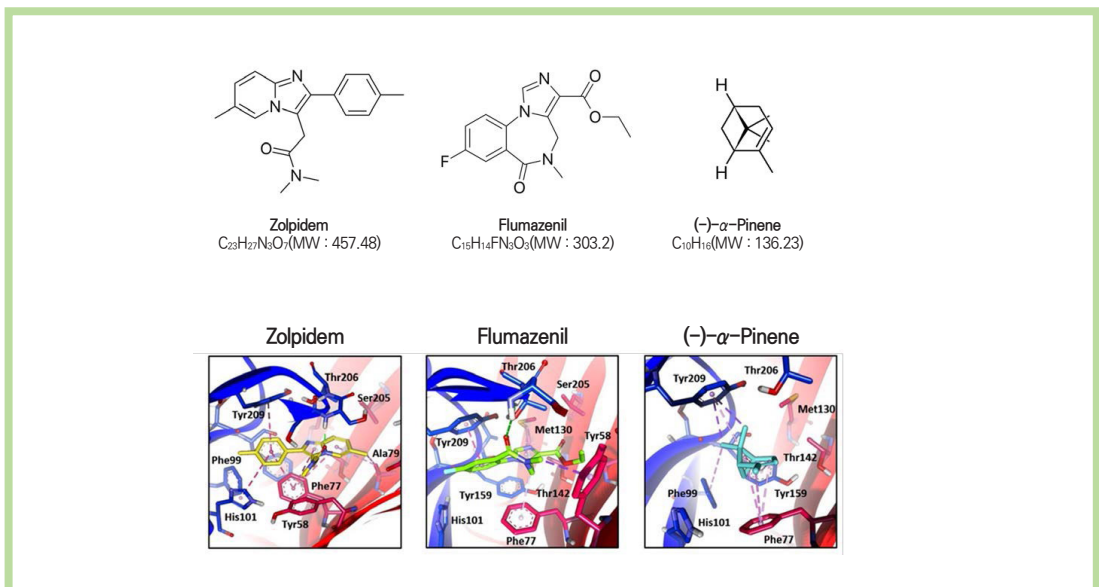


그림 21. 졸피뎀, 플루마제닐, 알파-피넨의 구조식과 분자 모델링을 통한 가바A형 수용체와의 결합력 예측⁴⁾



앞서 논의한 대로 알파-피넨은 졸피뎀, 디아제팜과 함께 벤조디아제핀 부위에 결합하는데, 이 세 가지 화합물이 어떻게 가바A형 수용체에 결합할까? 이것은 단백질 구조학을 바탕으로 하여 분자 모델링(molecular modelling) 기법을 이용하면 예측할 수 있다. 분자 모델링은 분자의 결합이나 움직임을 예측 하거나 모방 하는데 사용되는 모든 이론 및 연산 과정이다. 이 모델링을 통해 알파-피넨, 졸피뎀, 플루마제닐이 모두 가바A형 수용체의 $\alpha 1$, $\gamma 2$ 구성요소로 이루어져 세포외부분(extracellular domain)에 존재하는 벤조디아제핀 부위에 결합하는 것이 보고되었다 <그림 21>. 또한 분자 모델링을 통해 이 세 화합물과 수용체 사이의 결합력 또한 예측 되었는데, 결합력은 졸피뎀, 플루마제닐, 알파-피넨의 순으로 알파-피넨의 결합력이 낮게 나타났다. 이것은 졸피뎀과 플루마제닐의 경우 강력한 결합력을 가지는 파이 공유결합(π - π interaction)으로 수용체와 결합하는데 반해, 알파-피넨이 상대적인 약한 결합력을 가지는 소수성 결합(hydrophobic interaction)으로 수용체와 결합하기 때문인 것으로 예측된다.

03 결론

우리는 EEG와 EMG를 이용한 수면 행동의 분석, 졸피뎀과 플루마제닐과 같은 약물을 사용한 약리학적 방법, 패치 클램프 기술을 이용한 전기생리학적 측정, 구조생물학적 정보를 바탕으로 한 분자적 모델링을 통해 소나무 유래 피톤치드인 알파-피넨의 수면 증진 효과에 대해서 알아보았다. 이와 같은 다양한 과학적 접근 방법들을 통해 알파-피넨이 벤조디아제핀 수용체 조절제로서 가바 유래 억제성 신경전달 과정의 반응시간을 연장함으로써 수면의 질과 강도는 저해하지 않으며 비-렘수면의 양을 증가시키는 것을 확인하였다. 졸피뎀과 디아제팜을 포함하는 수면제의 기억력 저하, 주간 졸림증, 섬망, 환각 등의 부작용 및 약효의 저하와 함께 그 사용량을 증량하게 되는 내성과 같은 다양한 문제는 뉴스와 방송을 통해 빈번히 보도되었다. 이것은 수면 장애를 경험하는 사람들이 증가함에 따라 그 심각성이 점차 크게 대두되고 있다. 이러한 때에 소나무 피톤치드인 알파-피넨의 수면개선 효과에 관한 연구들을 통해 산림욕의 효과를 과학적으로 입증함은 물론 천연물 기반 수면제와 신경 안정제와 같은 신약 개발로의 가능성을 기대해 본다

저자 이창준

Changjoon Justin Lee

학력 Columbia 대학 신경생리학 박사
Columbia 대학 신경생리학 석사
Chicago 대학 화학과 학사

경력 現) 한국과학기술연구원 책임연구원,
現) 한국과학기술연구원 신경교세포 연구단 단장
前) 한국교세포연구회 회장

저자 우준성

Junsung Woo

학력 과학기술연합대학원대학교(UST)
생체신경과학 박사
경북대학교 생명공학부 학사

경력 現) 한국과학기술연구원 박사 후 연구원



참고문헌

1. Park BJ, Tsunetsugu Y, Kasetani T, Kagawa T, Miyazaki Y. The physiological effects of Shinrin-yoku (taking in the forest atmosphere or forest bathing): evidence from field experiments in 24 forests across Japan. *Environ Health Prev Med*. 2010 Jan;15(1):18-26. doi: 10.1007/s12199-009-0086-9.
2. 박술. 생물산책 '잠의 단계'
3. Satou T, Kasuya H, Maeda K, and Koike K (2014) Daily inhalation of α -pinene in mice: effects on behavior and organ accumulation. *Phytother Res* 28:1284-1287.
4. Yang H, Woo J, Pae AN, Um MY, Cho NC, Park KD, Yoon M, Kim J, Lee CJ, Cho S. α -Pinene, a Major Constituent of Pine Tree Oils, Enhances Non-Rapid Eye Movement Sleep in Mice through GABAA-benzodiazepine Receptors. *Mol Pharmacol*. 2016 Aug 29. pii: mol.116.105080.
5. 한국생물과학협회. 생명과학대사전. 2008.
6. Woo J, Cho S, Lee CJ. Isoliquiritigenin, a chalcone compound, enhances spontaneous inhibitory postsynaptic response. *Exp Neurobiol*. 2014 Jun;23(2):163-8. doi: 10.5607/en.2014.23.2.163. Epub 2014 Jun 13.
7. Farrant M & Nusser Z. Variations on an inhibitory theme: phasic and tonic activation of GABAA receptors. *Nat Rev Neurosci*. 2005 6, 215-229

02

R&D 빅데이터, 이제는 연구개발의 경쟁력을 좌우한다

한국과학기술연구원
김찬수, 오정수, 이광렬 박사
(tree@kist.re.kr, soo5@kist.re.kr, krlee@kist.re.kr)



01 서론

1.1 빅데이터 시대

야말로 홍수의 시대다. 매년, 매초 쏟아져 나오는 데이터가 정말 홍수와 같아 그 물살 위에 중심을 잡고 서야 한다. 제대로 중심을 잡으면 우리는 발 아래 있는 데이터의 형태를 이해하고, 그 의미를 알고, 나아가 데이터를 활용·가공하여 미래를 볼 수 있다. ‘빅데이터’로 대표되는 순수한 자료 묶음과 만나 과학이 만개한 사례가 있다. 바로 생물정보학이다. 생물정보학(bioinformatics)은 2000년대 초 인간유전체인 게놈(genome)에 담긴 30억 개 염기의 서열을 해독했다. 엄청난 데이터로부터 생물학적 이해를 도모하도록 그를 가공하고 이해하여 새로운 형태의 의미를 구성해낸 것이다.

인간게놈프로젝트(Human Genome Project)는 공식적으로 그 서열을 완전히 해독했다고 발표했으며, 지금은 이를 가지고 다양한 생물공학적 응용이 활발히 시도되고 있다. 예를 들어, 유전체 정보로부터 질병을 예측하거나, 이를 바탕으로 한 적정 치료 방법을 개발하고 있다. 근래에는 인체 수준의 데이터인 의료 정보까지 더하여, 최적의 약물처방 효과를 얻을 수 있는 표적신약(privatized medicine)의 개발이 무척 활발하다.

현재 우리사회는 정보통신(IT) 기술이 발달하면서 보이지 않는 곳까지 촘촘하게 연결되었다. 사물인터넷(IoT) 등으로 대표되는 최근의 초연결(hyper-networked) IT 기술은 비약적으로 발전했다. 이제 사회의 거의 모든 분야에서 데이터를 거의 실시간으로 수집하거나, 이런 엄청난 빅데이터를 바탕으로 새로운 해석을 시도해 볼 수 있게 됐다.

미국에서 빅데이터는 기존의 amazon과 같은 온라인 사업, ebay와 같은 경매, paypal과 같은 지불시스템 등의 무척 다양한 분야에서 사용되고 있다. 특히 축적된 빅데이터를 이용하여 알지 못했던 가치를 발견하거나, 새로운 차원의 가치를 만들어내고 있다.¹⁾ 지난 9월 Forbes 지에서 소개한 Palantir Technologies는 기존의 테러 발생 정보와 사람들의 실시간 이동 행태, 그리고 뉴스의 분석을 통해서 테러가 어느 지역에서 발생할지를 예측하는 기술을 내놓았다.



그림 1. 빅데이터를 이용한 서비스 개발 사례 1_테러 예측 기술. Palantir CEO Alex Karp

IBM의 왓슨(Watson)은 환자와 병원의 빅데이터를 직접 활용하여 질병을 진단한다. 실제 국내의 한 의료 현장(대학병원)에도 투입되었고, 질병 및 예후를 예측하는데 전문의보다 더욱 높은 정확성을 보여 엄청난 주목을 받고 있다.



그림 2. 빅데이터를 이용한 서비스 개발 사례 2_암진단 의료 서비스

미국의 주 경찰은 경찰차 안에서 달리거나 정지한 채로도 주변의 차량을 실시간으로 모니터링하는 ALPR (Automated License Plate Recognition) 시스템을 도입하였다. 우범차량 혹은 교통 범죄 가능성이 있는 차량들을 빅데이터를 통해 판별할 수 있는 시스템이다. 이는 범죄가 벌어진 후에 해당 차량의 도주로를 예상하는 데까지 사용되고 있다.



그림 3. 빅데이터를 이용한 서비스 개발 사례 3_범외차량 추적 시스템

농업이 어떤 산업보다 비중이 큰 미국에서는 농업에 대한 자료도 많이 축적되어 있을 뿐 아니라, 사용 가능한 분야도 다양하다. 주 정부에서는 파종시기, 파종방법 등을 포함한 그 동안의 농사 관련 빅데이터를 수집 및 정리했다. 이 자료를 민간에 이관하여, 효율성 높은 농업 방식을 제시하고 비용을 효과적으로 낮추는데 일조하고 있다.



그림 4. 빅데이터를 이용한 서비스 개발 사례 4_빅데이터 농업(출처:<http://www.monsanto.com/>)

02 본론

2.1 빅데이터와 연구 패러다임 전환

빅데이터 연구는 거대규모의 자료를 기반으로 한 과학적 방식의 연구이다. 빅데이터와 머신러닝·인공지능(AI, Artificial Intelligence) 연구는 인간의 사고를 도와주는 역할을 넘어 이제는 새로운 규칙 혹은 법칙을 찾는 데까지 활용되고 있다. 특히 연구의 패러다임을 바꾸고 있으므로 전 영역에 걸친 학자들은 관심을 기울여 할 수 없다. 논리성에 기반한 기존의 연역적인 과학에서 ‘합리성에 기반한 귀납적’인 과학으로 연구 방법과 철학, 그리고 사고의 중심이 이동하는데 가장 중추적인 역할을 하고 있다.²⁾ 이처럼 ‘초연결사회’로 불리는 실시간 빅데이터 축적 시스템은 연구개발(R&D)의 새로운 장을 열고 있다. 빅데이터를 활용한 연구개발은 과학적 연구 및 탐색의 과정이 실험(experiment) → 이론(theory) → 모사(simulation)에서 데이터에 기반한 모형 생성(data-driven modeling) 연구의 형태를 갖추고 있다.

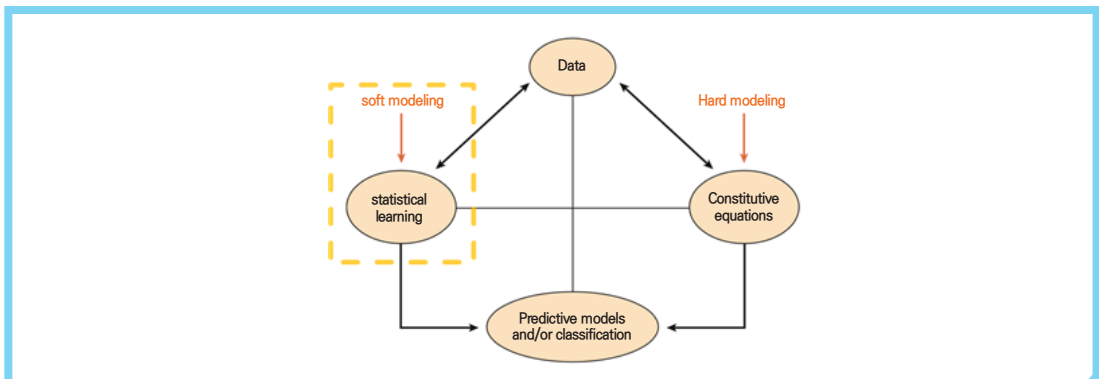


그림 5. 합리성에 기반한 귀납적 지식창출

머신러닝은 이미 모형이 수립되어 있는 단순한 선형회귀 방법을 통해서서는 수행할 수 없다. 대신 비선형 동역학 등의 접근법을 활용하여 다차원의 입력과 출력을 연결지어주는 방법이 연구돼 왔다. ① 신경 그물(neural network), ② Support Vector Machine (SVM), ③ 마구잡이 나무 모형(random forest model),

④ k-방법 갈래치기(k-means Clustering), ⑤ 깊이 배우기(Deep learning) 등의 방법들이 최근 30년간 개발되었다.³⁾ AI는 ① 빅데이터를 기반으로 하여 만들어진 귀납적 수리모형을 활용하여 ② 우리가 원하는 최적의 목적함수를 찾아가기 위하여 ③ 최적화 방법을 적용한다. 이때에 ④ 최적화의 방향(convergence direction)을 컴퓨터 스스로가 정해 판단할 수 있게 설계하는 것이 핵심이다. 미국은 빅데이터 기반 연구개발의 선두에 있다. 미국의 행정부는 백악관의 보고서를 통해 해당 기술을 장려하고 있다. 뿐만 아니라 빅데이터 기술의 표준 구현 형태를 논의하고, 경제에 미칠 영향까지 고려하여 기술 육성 전략 등을 수립하고 있다.⁴⁾ 그 전략으로서 다음의 7가지를 제시하고 있다.

- ① 빅데이터 기술 이해 제고 : 빅데이터 기반 기술을 개발함으로써 차세대 연구능력을 제고할 것
- ② 자료의 질 향상 : 기존 연구들이 신뢰성 높은 데이터를 양산해낼 수 있도록 지원할 것
- ③ 플랫폼의 개발 : 각 응용분야에서 빅데이터를 활용할 수 있는 인프라를 개발할 것
- ④ 정책 지원 : 데이터를 공유하고 활용할 수 있는 정책을 수립하여 데이터 활용의 가치를 제고할 것
- ⑤ 플랫폼 개선 : 빅데이터의 수집과 관련하여 개인정보를 보호하는 연구를 진행할 것
- ⑥ 기술 교육 : 빅데이터 기술의 보급을 위한 교육을 개선할 것
- ⑦ 연구 생태계 혁신 : 빅데이터를 활용한 연구를 위해 생태계를 조성할 것

무엇보다 빅데이터 활용을 위한 연구개발 플랫폼을 구축하도록 장려하고 이를 최우선 과제로 명시하고 있다.

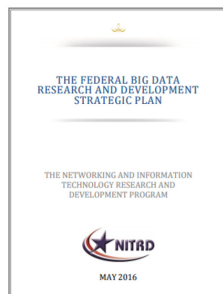


그림 6. 미국 NITRD(The Networking and Information Technology Research and Development)의 '연방 정부의 빅데이터 연구 및 개발 전략 계획' 보고서

2.2 빅데이터 플랫폼 : 새 연구의 실장(implementation)

빅데이터뿐 아니라, 일반적으로 데이터는 그 형태가 띠는 정형성을 기준으로 구분한다. 특히 정형데이터(Structured data)는 고정된 필드에 저장되는 데이터를 의미하는데, 이는 데이터가 필드값을 가진 테이블에 저장되어 있어 처리와 분석이 용이한 형태를 띤다. 반면 비정형 데이터(Unstructured data)는 이미지, pdf 파일, 동영상, 오디오, 통화내용, 위치정보 등으로, 여기에서 의미 있는 정보를 추출하는 것은 어려운 일이다. 즉, 전통적으로는 사람이 직접 정보를 추출해 올 수밖에 없다.

대개의 비정형데이터는 그 특성에 따라 각기 다른 관리법, 활용방안, 그리고 분석을 위한 별도의 처리방법이 요구된다(혹은 XML이나 HTML과 같은 메타데이터나 스키마 등의 반정형 데이터의 3가지로 분류하기도 한다 : <http://blrunner.com/12>).

그리고 과학기술의 빅데이터를 산출방법 측면에서 분류해보면 실험데이터와 계산데이터로 양분이 가능하다. 실험데이터는 실험을 통해 얻어진 데이터로, 실험인자의 다양성 때문에 구조화하기가 쉽지 않다. 즉, 실험환경, 실험장비, 방법론 등이 매우 다양하고 또한 매 실험마다 구체적인 실험인자도 변화되는 까닭이다. 이 경우 평균값 또는 대표값을 저장하여 활용하는 것이 일반적이다. 반면에 계산데이터는 계산과 시뮬레이션 등의 컴퓨터 기법을 통한 데이터를 뜻한다. 실험기반의 데이터 보다 그 관리가 용이하지만, 사용하는 계산방법론이 산출해주는 물성만을 확보할 수 있다는 한계가 분명하다. 예를 들어 제일원리계산은 원자스케일에 대한 데이터만을 산출한다. 다른 수준의 자료를 확보하기 위해서는 분자동역학(Molecular Dynamics), 몬테카를로(MonteCarlo), 유한요소법(Finite Element Method) 등의 다양한 계산을 수행해야 한다. 즉 없는 자료를 생산해내어야 하는 부담이 상존한다.

빅데이터를 위시한 대개의 정보학이 데이터에 기반하는 만큼 데이터의 수집을 비롯한 체계적인 관리 운영 체계 역시 매우 중요하다. 데이터 수집의 대상으로는 '중립성'이 요구된다. 왜냐하면 대개의 연구자들은 실험과 계산을 통해서 나온 긍정적인 결과들에만 집중하고, 부정적인 결과나 예상을 못했던 결과들은 무시하거나 폐기하거나 삭제하는 것이 일반적이기 때문이다. 그러나 빅데이터를 수행하기 위해서는 긍정/부정의 결과를 따지 않고, 모든 데이터들이 축적되어야 하며, 부정 데이터 역시 기계학습을 위해서는 반드시 필요하다.



많은 방법과 기술이 존재하는, 그래서 가장 일반적인 기술로 꼽는 것은 바로 저장 기술이다. 데이터가 포함하는 정보는 표준양식(schema)으로 저장되어야 탐색과 분석이 용이하다. 그 저장 방식에 따라서 자료의 형태에 따른 저장기술과, 처리/접근 패턴(접속빈도 및 보안 등)이 고려되어야 한다. 대개는 복수개의 저장시스템을 구현하기도 한다.

빅데이터 구현에서 빠질 수 없는 요소는 데이터베이스(DB)화이다. 이미 많은 DB가 존재하지만, 접근성(accessibility)에 한계가 있다. 특히 DB가 비공개(주로 상업용)인 경우, 접근성이 낮을 뿐 아니라, 데이터의 정확도(accuracy)도 높지 않다. 분야별 특성에 따라 단일화된 데이터 인프라 혹은 작은 DB 시스템들의 연합 형태로 발전되어 왔는데, 각 특정 분야별 특화된 데이터 리소스들은 상당수 존재하고 있다. 이와 관련하여 빅데이터의 '표준화' 역시 중요한 이슈이다. 표준화가 돼 있지 않으면 데이터의 저장과 DB 구축에 통일성이 없어, 데이터를 재활용/공유/분석하는 과정에서 충분히 활용할 수 없다. 빅데이터가 통일성 없이 저장된 경우라면, 어떤 통찰도 얻을 수 없다는 치명적인 결과를 만들게 된다. 안타깝게도 표준의 필요성을 인식하고는 있으나 이해 당사자들 간의 합의 도출이 쉽지 않은 상황이다.

빅데이터가 공유될수록 더 좋은 결과가 나온다고들 믿는다. 하지만 분야별로 많은 차이가 있다. 입자물리, 바이오, 천문학 분야는 거대규모 실험장치에서 생산되는 자료의 양이 워낙 많으므로 그 공유가 필수불가결하지만, 재료분야는 전통적으로 작은 데이터 세트를 통한 분석과 활용에 익숙하기에 아직 그 인식이 좋지 못하다. 하지만 지금은 빅데이터 및 데이터 세트를 공유하려는 분위기는 더욱 확산될 것이며, 이에 따른 부가적 인센티브도 증가할 것으로 생각된다. 왜냐하면 빅데이터를 공유하여 기계학습과 분석의 정밀도를 증가하고 예측력을 더욱 강력하게 하여 우수한 연구결과로 이어질 것이 자명하기 때문이다. 전술한 이유로 인해 빅데이터를 ① 수집하고 ② 관리하기 위한 플랫폼이 절실히 요구되는 실정이다. 특히 KIST의 경우 ① 계산 분야에 있어 연간 약 5000TB의 미가공 빅데이터(raw data)가, ② 실험 분야에서는 연간 약 20GB의 미가공 데이터가, 그리고 ③ 분석 분야에서는 이미지 자료를 포함하여 연간 약 100TB의 미가공 데이터가 생산되고 있다(2016년 KIST 조사 기준).

인포메틱스 플랫폼은 빅데이터를 실질적으로 구현하는 기술로 이해할 수 있다. 플랫폼을 통한 데이터 기반의 지식 창출은 데이터의 잠재 가치로부터 실질 가치를 실현하는 의미를 지닌다. 이러한 플랫폼 내에서는

통합적 framework가 중요하다. 통합형 플랫폼 설계를 위해서는 단일화된 프로그래밍 언어, 인터넷 연결 등의 요소가 고려되어야 한다.

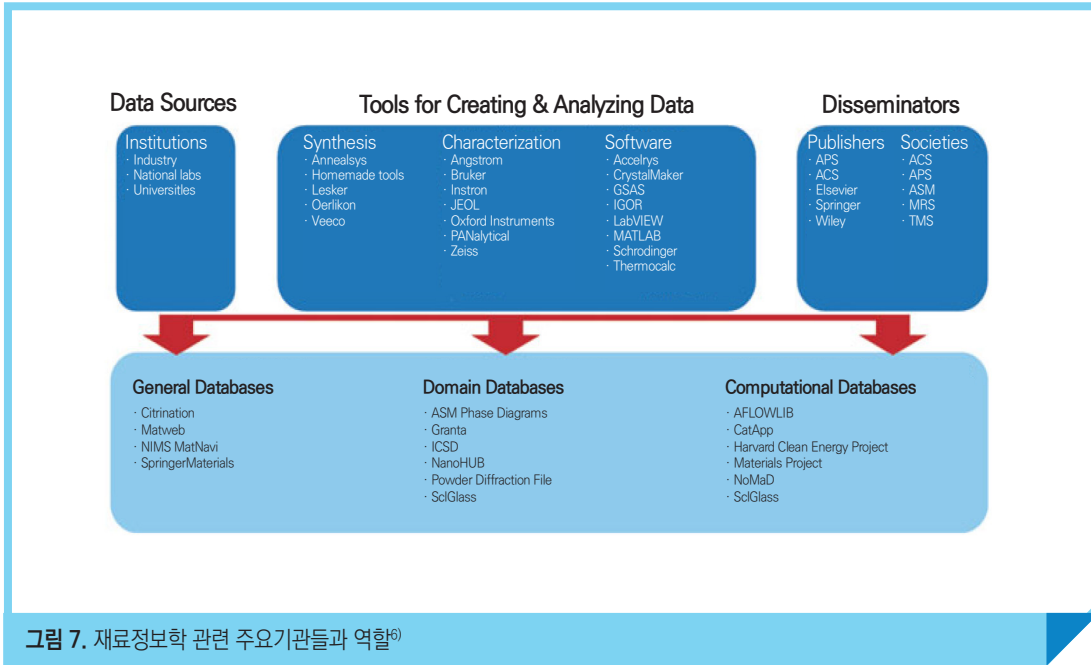
인포메틱스 플랫폼은 방대한 데이터 정보로부터 숨어있는 지식에 접근하고, 연구개발 과정에서 빅데이터를 활용할 수 있도록 하는 연구 기반으로서 모델링, 계산, DB구축, 기계학습, 예측/실제 등의 기능이 구현될 것이다. 즉 사용자가 정보를 처리하고 최종 통찰을 얻기 위한 플랫폼과의 접촉면으로, 데이터의 입력, 접근, 쿼리, 분석, 다운로드 등 다양한 기능의 지원이 필요하다. 가시화 도구들은 데이터를 패턴화하여 유용한 트렌드를 전달해주며 사용자들이 이해 가능한 포맷으로 변경을 지원하는 강력한 수단이다.

데이터와 알고리즘을 이용해서 과학적 지식을 창출하는 R&D 정보학에서는 분야 전문가와 정보처리 툴을 위한 프로그래밍과 데이터 분석/처리에 익숙한 전산전공자 또는 데이터 사이언티스트들이 협력하는 융합연구가 필수적이다.

2.3 빅데이터 : 미래 경쟁력의 핵심

빅데이터를 활용한 연구는 데이터에 기초하므로, 데이터 수집뿐 아니라 체계적인 관리와 운영의 체계가 중요하다. 재료과학 분야에서 데이터 기반의 연구개발에 대한 노력이 본격화된 것은 비교적 최근의 일이다.⁵⁾ 그것이 가능해진 이유는 IT 기술의 급격한 발전에 힘입어 재료 분야의 데이터들이 본격적으로 축적되기 시작했기 때문이다. 특히, 기존의 연구방식에 따라 새로운 소재를 발견하고 상업화하기까지는 대개 10~20년의 시간이 소요되는 소재개발의 시간을 크게 단축시킬 수 있는 가능성 때문에 주목을 받고 있다. 그러나 이러한 성과가 가시화되기 위해서는 우선 재료분야 데이터의 체계적 축적과 활용수단의 확보가 선결과제이다.

데이터가 재료 연구의 주요한 조력자로 등장하면서, 미국 Materials Genome Initiative(MGI) (2011~), 미국 Air Force Research Laboratory, NIST, NSF 공동의 Materials Science and Engineering Data Challenge 등이 시작되었다(2015). 이들은 주로 공개된 데이터 기반으로 insight를 확보하고, 이를 통해 신소재 발굴이 가능함을 증명하는데 주력하고 있다.



이외에도, Lawrence Berkeley National University는 Elastic properties에 대한 세계 최대 데이터베이스를 구축하였고, NASA는 우주정거장 재료에 관한 자료 등이 포함된 Physical Science Informatics 데이터베이스를 구축하고 있으며 유럽과 일본에서도, 재료 데이터 인프라 구축이 활성화되고 있다. European Commission Joint Research Centre's MatDB와 제일원리계산 데이터를 중심으로 하는 European Union's NoMaD (<http://nomad-repository.eu>) 일본의 National Institute of Materials Science (NIMS)에서도 재료데이터 확보 및 재료정보학 연구를 수행 중이다.

앞으로 사회적으로 빅데이터를 공유하려는 분위기는 더욱 확산될 것이다. 그리고 빅데이터를 표준화하고 이의 수용 및 공유를 통해서만 지금과 다른 차원의 여러 부가적 보상도 증가할 것으로 예상된다. 빅데이터 공유는 머신러닝과 AI분석의 속도를 증가시킬 뿐 아니라 신뢰성도 제고함으로써 귀납적 모형의 합리성을 더욱 탄탄히 만들 수 있기 때문이다. 이에 기반한 우수한 연구성과는 자명한 결과가 아닐까?

2.4 빅데이터 기반의 계산나노과학 - 소재 분야

2-4-1 개관

재료분야에서 빅데이터를 적용하는 시도는 계산나노과학(Computational Nano-Science)에서 최초로 시도되고 있다. 계산나노과학은 컴퓨터를 이용한 시뮬레이션과 빅데이터 및 정보기술을 이용하여 나노소재와 그 공정, 나노시스템을 설계하고 최적화하는 기술을 통틀어 일컫는다. 오래 전부터 물리학자들은 물질을 구성하는 원자와 전자들의 움직임을 이해한다면 물질의 특성 대부분을 이해할 수 있을 것으로 기대해왔다. 그러나 물질을 구성하는 원자나 전자의 수가 계산이 불가능할 정도로 많은 것이 한계였다. 이는 나노산업의 근간인 나노물질에서도 여전히 어려운 과제였다. 그러나 최근 혁명적으로 발전해 온 IT 기술과 빅데이터 기술을 바탕으로 컴퓨터의 계산 능력과 판단 능력이 크게 증가하였다. 이에 따라서, 합리적인 자료에 근거하여 더욱 실질적인 나노물질을 시뮬레이션 할 수 있는 시대를 맞게 되었다.

논리성에 기반한 또다른 모형적 접근인 밀도범함수론(Density Functional Theory) 등 첨단 계산 방법의 발전과 원자간 상호반응 모델의 개발을 발판으로 시뮬레이션 기술은 점차 현상의 해석 차원을 넘어 현상을 예측하고 실험연구의 방향을 제시할 수 있는 수준으로 발전하고 있다.

대부분의 나노물질은 저차원의 형상을 갖거나 비표면적(부피에 대한 표면의 비율)이 매우 크므로, 표면에서 일어나는 현상이 물성에 커다란 영향을 미친다. 표면의 구조가 미세하게 변하거나 의도치 않은 분자가 흡착되는 것만으로도 나노물성이 크게 변하여 물성 제어에 어려움을 겪는 경우가 많다. 이러한 실험상의 노이즈를 없애기 위해 고순도 물질, 단결정, 고진공, 극저온 등의 조건에서 실험해야 하므로 연구개발에 많은 비용과 노력이 필요하다. 또한 나노 수준의 재료 거동을 명확하게 이해하기 위해 도입된 기존의 연역적 모형은 수치해석의 한계뿐 아니라 모형 수립의 한계와 같은 논리성의 한계도 존재한다.

빅데이터에 기반한 계산나노과학은, 자료를 기반으로 하여 귀납적인 접근을 수행하므로 이러한 한계를 돌파할 수 있는 중요한 연구방법으로써 나노 연구개발의 효율성 증진 뿐만 아니라 그 성공 여부에도 직접적으로 영향을 미친다. 따라서 빅데이터 계산나노과학은 실험적으로 구현하기 어려운 이상적인 조건에서의 연구를 구현해줄 뿐 아니라, 새로운 현상의 모형화를 수립하는데 무척 유용하다.



이처럼 나노소재의 이론적 모델과 첨단 빅데이터 기술을 융합한 계산나노과학 기술은 소재개발의 새로운 패러다임을 여는 커다란 흐름으로 부각되고 있다. 2013년도 노벨화학상이 실험과 계산의 협력연구 기반을 구축한 과학자들에게 주어진 것처럼 앞으로 나노산업의 연구개발에서는 실험과 계산이 중요한 두 축을 이룰 것이다.

2-4-2 현황

수많은 시행착오를 수반하는 실험적 방법에만 의존하는 나노제품의 개발은 연구개발에 소요되는 시간과 비용 면에서 커다란 위험요인들을 안고 있다. 그러나 본격적인 개발에 들어가기 전 컴퓨터 시뮬레이션과 빅데이터 분석을 활용하여, 탐색 대상의 가능분야를 제한하여 확인할 수 있다면 최적의 공정 설계에 반영할 수 있다. 이는 결국 연구개발이 갖는 위험요인들의 상당 부분을 해소하여 성공 가능성을 높인다.

따라서 빅데이터에 기반한 계산나노과학은 소재 및 나노과학의 성공을 위해 반드시 확보되어야 할 기술로 인식되고 있다. 미국에서는 나노기술의 핵심 경쟁력이 빅데이터 기반의 설계 및 정보화에 달려 있다는 인식이 강하다. 즉 2011년 Materials Genome Initiative (MGI)를 발표하고 기존의 막강한 IT 인프라(세계 computing power의 약 60%를 점유)를 활용한 효율적인 나노소재의 연구개발 체계를 구축하였다. 일본 역시 World Premier Institute 프로그램의 MANA(International Center for Materials Nanoarchitectonics, NIMS) 및 AIMR(Advanced Institute for Materials Research, Tohoku University)를 통해 자료 기반의 나노기술의 발전을 위해 노력하고 있다.⁸⁾ 유럽에서는 FP7 프로그램에 이어 Horizon 2020의 NMP 프로그램에서도 빅데이터와 컴퓨터를 통한 모델링은 중요한 연구개발의 축을 구성하고 있으며, 2014년 설립된 European Materials Modeling Council을 중심으로 유럽 산업에서 계산과학 기술의 활용을 더욱 증진시키기 위해 노력 중이다.⁹⁾

현재 빅데이터에 기반한 계산나노과학과 재료정보학의 발전을 위해 단일통합 시스템을 구성하려고 노력하고 있다. 다양한 재료연구 데이터를 함께 저장하고 구조화를 추구하는데, 미국의 National Institute of Standards and Technology (NIST), Citrine Informatics 社 (Citrination Platform) 등을 대표적으로 꼽을 수 있다.

2011년 시작된 Materials Project는 Lawrence Berkeley National Laboratory가 주축이 되어 수행되고 있다. 빅데이터에 기반한 나노연구의 선두이자 대표격으로, 수학적 알고리즘, 정보과학, 계산재료과학을 융합하여 새로운 기능성 소재를 발굴하는 것을 그 목표로 하고 있다. Ab initio 계산으로 65,000 개 이상 inorganic compounds 의 구조, 열역학 성질, 전기/ 기계적 성질과 같은 빅데이터를 보유하고 있다. 특히 open-source python-based analysis library(pymatgen) 와 고성능컴퓨팅(HPC)에 기반을 두고 있어, 확장성이 높다고 평가된다.

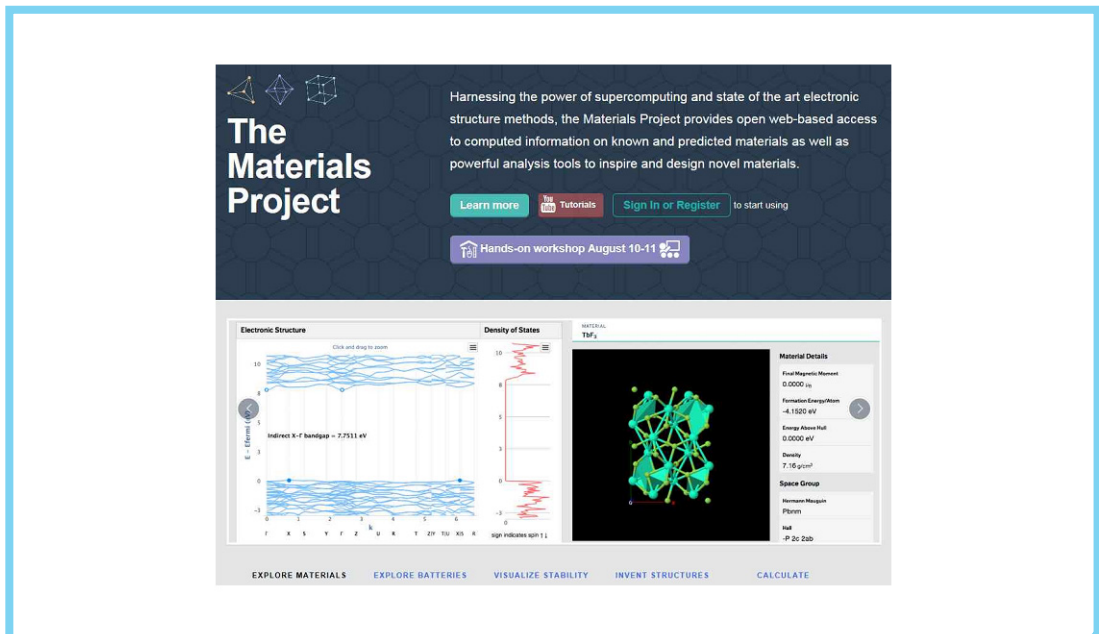


그림 8. The materials project

Open Quantum Materials Database(OQMD)는 ICSD의 합성물을 대상으로, DFT 총 에너지 계산데이터 400,000여 개를 확보한 빅데이터 계산나노 기술이다. 합성 재료를 탐색하거나, 상전이 그림(phase diagram)을 작성한다든지 혹은 안정 상태의 화합물(groud-state composition)의 예를 계산해낸다. 이는, 나아가 결정구조를 가시화하거나 필요한 데이터의 다운로드도 가능하게 구성되어 빅데이터의 활용에 초점을 두고 있다. 활용처로는 Structural 금속 합금, Li 배터리용 재료, 고효율 나노구조 thermoelectrics 스크리닝 등이 있다.

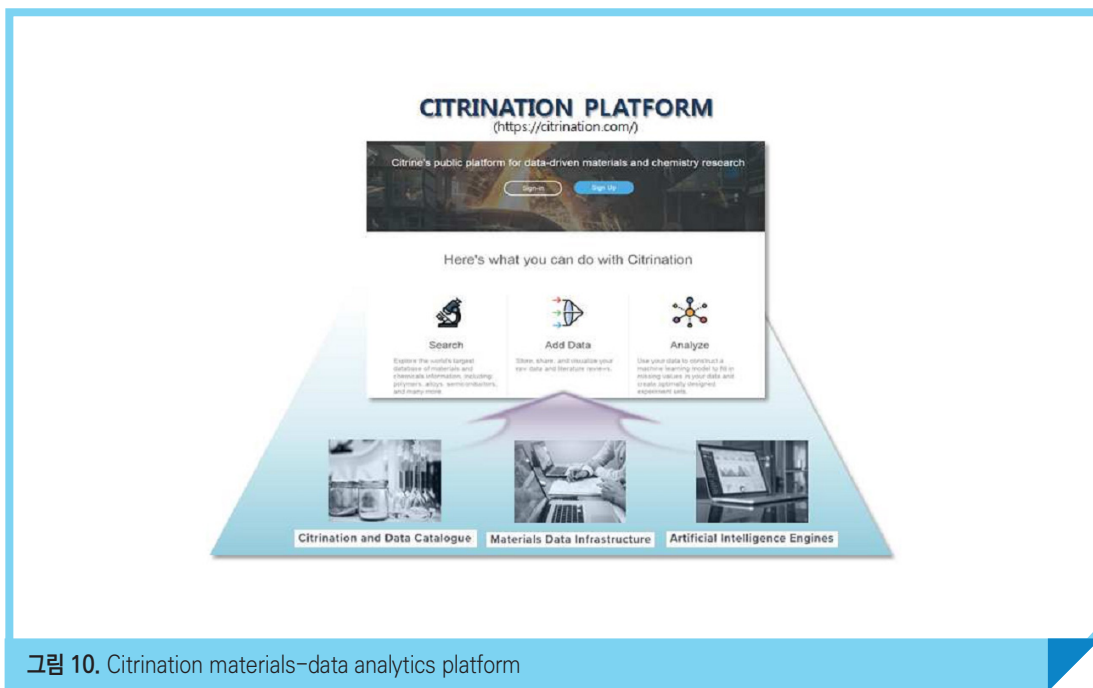


그림 10. Citrination materials-data analytics platform

Expert-led database building from literature은 빅데이터 기반의 재료과학에서는 ICSD(Inorganic Crystal Structure, 무기결정구조 데이터)가 비로소 DB로서의 입지를 닦고, 기초를 세웠다는 데 의의가 있다. 그러나 아직 기능성 재료에 대한 유용한 repository는 없다는 단점이 있어, 현재로는 플랫폼으로서의 역할에만 주목할 만하다. 대개 논문으로부터 자료를 수집하기 위해서는 논문 내의 key data를 결정하고, 연구자가 직접 논문 안의 데이터를 수작업으로 추출하는 등 사용상의 불편함이 있다. 또한 SW를 활용해 논문 내 그림과 표를 디지털화하거나 가시화하는 작업도 거쳐야 한다.

Open-access movement는 지난 20여 년간 재료분야의 빅데이터가 누적되자, 바이오분야와 유사한 open access 운동이 접목되어 만들어진 연구 운동이다. 거대 출판사인 Elsevier에서는 'Open Data'라는 인프라를 구축하기 시작하여, 13개의 주요 출판사와 재료과학 관련 데이터를 공유 중에 있다. Nature Publish Group에서는 Scientific Data Journal(2014)을 운영 중인 것도 이런 영향을 반영한 것이다.



이러한 세계적 추세에 맞춰 우리나라에서도 빅데이터를 활용하고 계산 나노과학을 생산의 한 수단으로 활용하는 움직임이 뚜렷하다. 상당수의 선진 연구그룹은 나노기술의 개발을 위한 멀티스케일 시뮬레이션 연구를 진행해 오고 있으며 산업적으로 의미 있는 결과를 추출하기 위해 노력하고 있다. 미국의 경우 무기 개발과 우주 개발 분야로부터 나노기술 분야에 이르기까지 빅데이터를 활용한 계산과학의 필요성에 대한 인식이 매우 높으며 국방성 및 에너지성 산하의 국립연구소에 거대 규모의 계산과학 연구팀을 유지하고 있다. 대표적인 예로서 Oak Ridge National Lab의 계산과학 연구팀을 들 수 있다. 최근까지 세계 최고 속도의 슈퍼컴을 운영하고 있으며 나노기술 및 소재 설계 분야의 높은 경쟁력을 확보하고 있다.

국내의 경우 빅데이터를 활용한 계산나노과학은 아직 꽃피어 있지 않다. 빅데이터 플랫폼을 구축하고 빅데이터에 기반한 귀납적 모형을 만드는 연구는 KIST를 필두로 하여 올해 큰 규모로 시작되고 있다.¹⁰⁾

현재 개별 길이 규모에서의 전산모사 경험은 비교적 많이 축적되어 있다. 제일원리계산 분야에서는 물리학 및 화학 분야의 연구자들을 중심으로 꾸준한 연구가 진행되고 있으며, 나노물질 구조, 표면과학, 수소저장, 나노튜브, 그래핀, 나노촉매, 반도체 등 매우 다양한 분야에서 활발히 응용되고 있다. 전술한 바, 우리나라는 아쉽게도 선진국에서 특정 과학 분야별로 개발된 S/W를 도입하여 응용하는 단계에 머물러 있으며, 기존의 계산 방법을 넘어서는 첨단화된 방법론의 개발 등의 연구는 취약하다. 분자동역학(MD) 시뮬레이션은 재료공학과 화학공학, 전자공학 분야의 연구자들이 주로 연구를 진행하고 있다. MD 시뮬레이션은 국내 소재 관련 거의 모든 연구기관에서 활발하게 활용되고 있으며, 상당한 수준의 연구성과를 거두고 있다.

국내 연구의 대부분은 소재의 물성 평가 혹은 단편적 현상의 이해를 위한 연구에 머물러 있으며 해석기술의 체계적인 융합을 통한 소재설계 시스템으로는 아직 발전하지 못하고 있다. 이는 선진국들에 비해 우리나라에서는 아직까지 계산과학 분야에 대한 체계적인 지원이 없었기 때문인데, 이로 인해 소규모 연구그룹이 개별 길이 규모의 전산모사 연구를 수행하고 있을 뿐 종합적이고 전략적인 관점에서 나노계산과학 기술의 개발이 진행되지 못하고 있다. 그러나 빅데이터에 기반한 귀납적 모형에 대한 인식이 제고되고 있는 지금의 실정은, 그 중요성에 눈을 돌리게 하고 있다.

2-4-3 산업화

빅데이터 모델링 및 시뮬레이션에 기초한 계산과학기술은 본격적인 실험에 착수하기 전에 성능에 영향을 미치는 주요 인자를 추출함으로써 고비용의 실험연구를 최소화하여 연구비용과 시간을 절약하는 효과가 있다. 2005년도 IDC Market research white paper를 보면 빅데이터와 시뮬레이션 기술은 연구효율을 증진시켜 연구비용과 시간을 감소시키고 새로운 물질과 제품에 대한 이해의 폭을 확장시키고 있다. 특히 1 달러 투자 당 3~9 달러의 이익을 얻는 효과를 갖는 것으로 평가받고 있다. 연구개발 툴로서 가치를 가진 빅데이터 계산과학의 시장은 나노산업 및 첨단 소재 산업의 연구 개발 투자 및 시장규모에 비례하여 성장할 것으로 예측된다. 대표적인 첨단 소재인 나노소재를 바탕으로 한 나노산업의 경우 시장 규모는 2000년 이후 매년 25%씩 증가하여 3년마다 2배씩 성장하여 왔다. 2011년 이후에는 그 성장속도가 더욱 증가하여 연평균 44%에 이르고 있다.

빅데이터 기술의 혁신적인 발전은 일상생활 뿐 아니라 과학기술 연구에서도 큰 변화를 가져오고 있다. 빅데이터에 기반하여 새로운 소재를 발견하는 등의 연구가 이제는 나노소자 및 공정을 설계하고 나노소재의 특성을 미리 예측하는 '합리적 나노기술의 개발'로 이어질 것이다. 이러한 빅데이터 기반의 연구방식은 나노기술의 패러다임 변화로 계속될 것이며, 가속화될 것이다. 재료과학을 통한 공정 및 구조 설계는 매우 중요한 분야로써 계산나노과학이 큰 역할을 할 수 있는 분야이다. 자동차, 전자, 조선 등 국내 주요 산업이 세계 일류 수준으로 성장함에 따라 소재 자체의 개발 필요성이 크게 증대되어 국내 유수의 대기업들은 소재 분야의 연구개발에 본격적으로 뛰어 들고 있다. 이러한 연구개발 추세에 따라 원자 혹은 아원자 수준의 소재설계 필요성이 크게 증가하고 있고 이러한 경향은 앞으로도 계속될 것이다.

빅데이터에 기반한 나노물질 시뮬레이션 기술의 산업적 가치 창출을 위해서는 (1) 빅데이터 플랫폼의 구축, (2) 빅데이터 구현 및 수집, (3) 이를 바탕으로 한 멀티스케일 시뮬레이션, 그리고 (4) 합리성에 기반을 둔 베이지안(Bayesian) 방식의 귀납적 모형화 연구가 반드시 필요하며 이러한 기술은 유기적 협업을 통해서만 가능하다. 이러한 기술들의 개발과 융합을 통해 다양한 나노기술 개발에 광범위하게 활용될 수 있는 시뮬레이션 기술을 구현하고, 시뮬레이션 계산결과들이 축적되는 데이터베이스와 이들로부터 의미 있는 자료를 도출하면, 계산나노과학 기술의 산업적 가치가 실현될 수 있을 것이다.

2.5 R&D 빅데이터 플랫폼의 모델 사례

데이터 기반의 R&D 인프라 구축의 사례로서 KIST에서 추진 예정인 데이터 기반 R&D 인프라 구축의 경우를 살펴보고자 한다. 이 계획은 R&D 빅데이터 및 정보학 기술을 이용할 수 있는 연구개발 인프라를 구축함으로써 데이터 기반의 연구개발 환경을 제공하는 것을 목표로 하고 있다. <그림 11>은 본 인프라의 핵심 플랫폼 구성과 데이터의 흐름을 보여주고 있다. 본 인프라 구축에는 크게 3개의 플랫폼 및 다수의 서비스 개발이 포함되어 있다.

첫째는 KIST 내부 R&D 데이터의 수집을 위해 KIST R&D 플랫폼 “KiRI (KIST R&D Informatics) Note”를 구축하여 제공하는 것이다. 연구자들은 이 플랫폼 상에서 연구개발을 수행함으로써 연구의 각 과정이 기록되고, 시료 혹은 시스템의 특성분석 결과들이 축적된다. 이 시스템은 KIST의 내부정보망과 프로젝트의 정보를 공유하고, 특성분석센터 및 계산재료과학 플랫폼의 데이터와 연계되며, 연구수행 기록과 분석결과들을 바탕으로 연구노트를 자동으로 생성하여 연구노트 인증시스템과 연동된다. 이를 통해 지적재산의 보호가 이루어지도록 설계되었으며 연구개발 활동 자체를 편리하게 수행할 수 있는 많은 서비스를 제공한다. <그림 12>

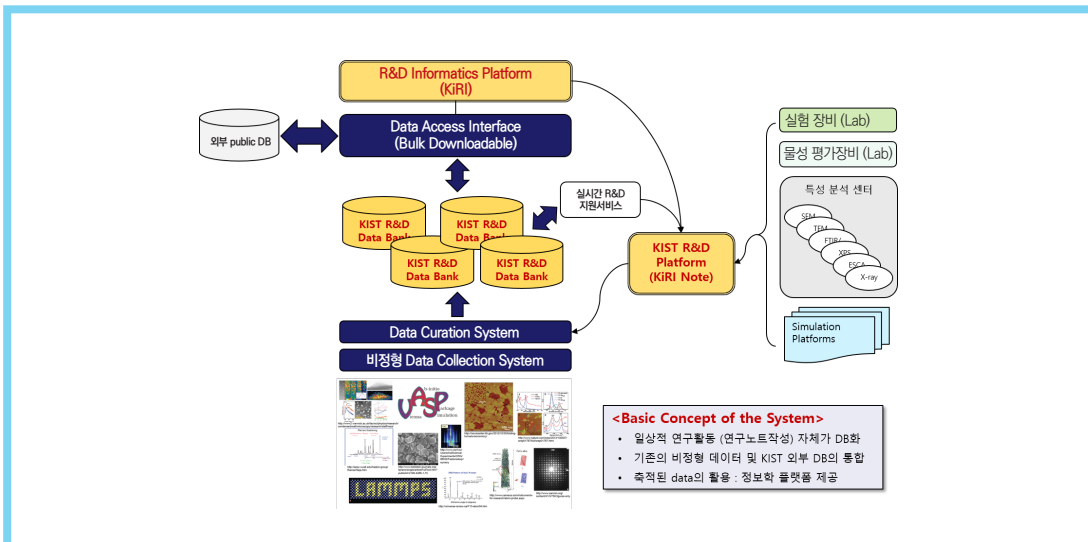


그림 11. KIST 데이터 기반 R&D 플랫폼의 구성도. 본 플랫폼은 ① 일상적 연구활동의 기록 자체가 DB화 될 수 있는 시스템, ② 기존의 비정형 데이터 수집 가공을 통한 다양한 R&D 자원 서비스 ③ 원내의 DB를 연동한 빅데이터 기반의 인포메틱스 서비스를 구축하여 제공함.

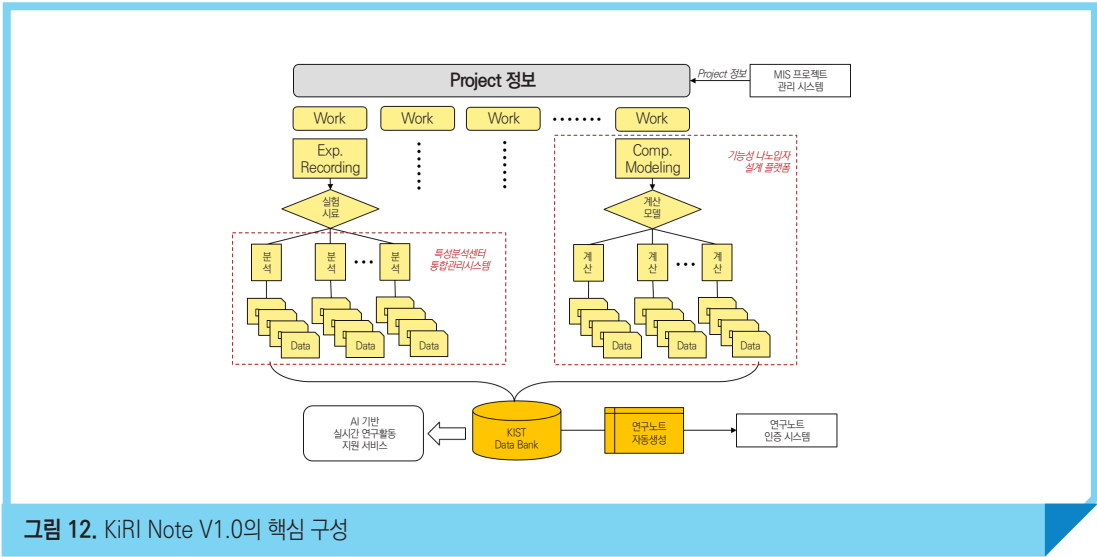


그림 12. KiRI Note V1.0의 핵심 구성

또한, KiRI Note에서는 KIST Data Bank에 축적되어 있는 데이터를 기반으로 실시간 R&D 지원 시스템을 구축하여 서비스 할 것이다. 이 지원 시스템은 데이터가 입력되면 유사한 연구를 수행하고 있는 원내 연구진의 정보와 관련이 높은 논문 및 특허 정보를 제공한다. 이를 통해 협력연구를 활성화하고 연구의 방향을 설정할 수 있는 환경이 제공될 것이다.

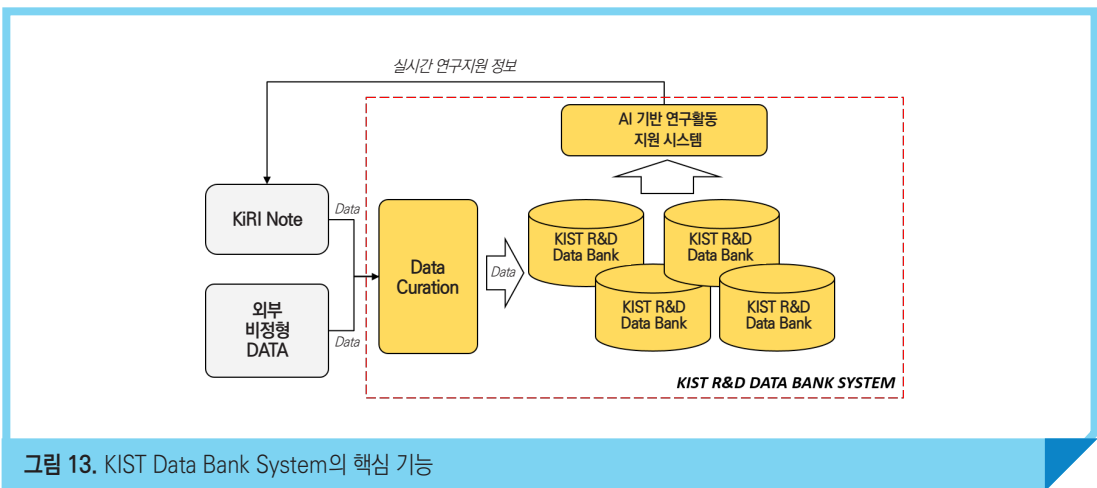


그림 13. KIST Data Bank System의 핵심 기능

1. 프레임워크, 플랫폼, 연구역사, 데이터 관리

2. R&D 프로젝트, 연구진, 연구역사의 정보에 기반한 연구역사

KIST 내에서 축적되는 데이터만으로는 지식을 창출하기에 충분한 규모의 데이터가 될 수 없기 때문에 글로벌 스케일의 데이터 수집 시스템을 구축하여 KIST Data Bank의 데이터 보유량을 확대할 것이다. <그림 13>은 이러한 시스템의 데이터 흐름을 보여주고 있다. 미래에는 데이터의 국제적 공유가 일반화될 것이므로 국제 표준과 연계된 DB 시스템을 구축하고, KiRI Note의 데이터 뿐 아니라 기 출판된 논문과 특허 그리고 공개된 모든 데이터 등 외부 비정형 데이터를 수집할 수 있는 시스템을 구축한다. 이 데이터들은 활용가능한 형태로의 curation단계를 거쳐 Data Bank에 축적관리 되어야 한다. Data Bank의 데이터는 곧 KIST의 자산으로서 이를 바탕으로 다양한 서비스가 개발되게 된다. 그 한 예가 전술한 KiRI Note의 실시간 연구활동 지원시스템이다.

마지막으로 데이터 기반의 지식창출을 위한 정보학 플랫폼 “KiRI” 플랫폼이다. <그림 14> KiRI 플랫폼은 분석대상 데이터를 KIST Data Bank와 외부 공개 DB에서 가져와 관리하는 데이터 통합관리 시스템을 근간으로 KIST Data Bank Service Portal을 제공한다. 또한 정보학 기반 연구의 개념을 검증하기 위해 머신러닝 기반의 촉매 설계 모델을 개발하여 플랫폼으로 제공하게 될 것이다. 개념 검증에 성공한다면 향후 응용분야는 계속 확대할 필요가 있을 것이다. 빅데이터를 제대로 활용하기 위해서는 각 분야의 전문지식(domain knowledge)을 가진 연구자들과 정보처리 전문가, 머신러닝 전문가 등 수학적 분석에 능통한 데이터 과학자들이 협력하는 융합연구가 반드시 필요하다. 각 분야의 전문가들에 의해 명확히 정의된 문제에 대해 데이터 기반의 해법을 모색해야 하기 때문이다.

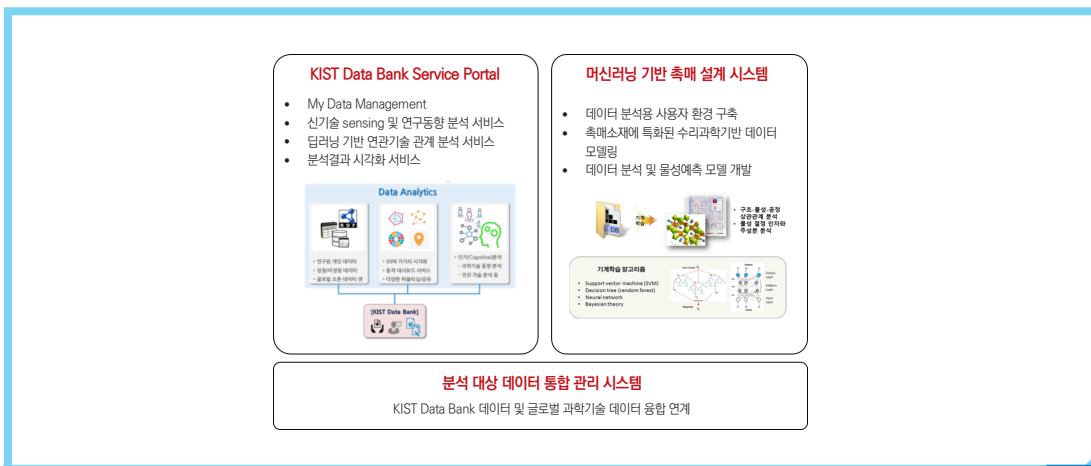


그림 14. KIST R&D Informatics Platform (KiRI v1.0)의 핵심 기능

03 결론

빅데이터를 제대로 활용하기 위해서는 각 분야의 전문지식(domain knowledge)을 가진 연구자들과 정보 처리 전문가, 머신러닝 전문가 등 수학적 분석에 능통한 데이터 과학자들이 협력하는 융합연구가 반드시 필요하다. 이는 각 분야의 전문가들에 의해 명확히 정의된 문제에 대해 데이터 기반의 해법을 모색해야 하기 때문이다. 최근 미래창조과학부는 ‘미래소재 준비계획 수립을 위한 공청회(2016년 11월 18)’를 갖고, 소재 관련 미래 기술을 개발하기 위해서 빅데이터와 계산과학을 필수적으로 사용해야 한다고 천명한 바 있다.⁷⁾ 이에 따라 나노소재연구개발사업에서 최근 나노정보학 플랫폼 사업이 시작되어 우리나라도 이제 데이터 기반의 소재연구에 첫걸음을 디뎠다고 할 수 있다.

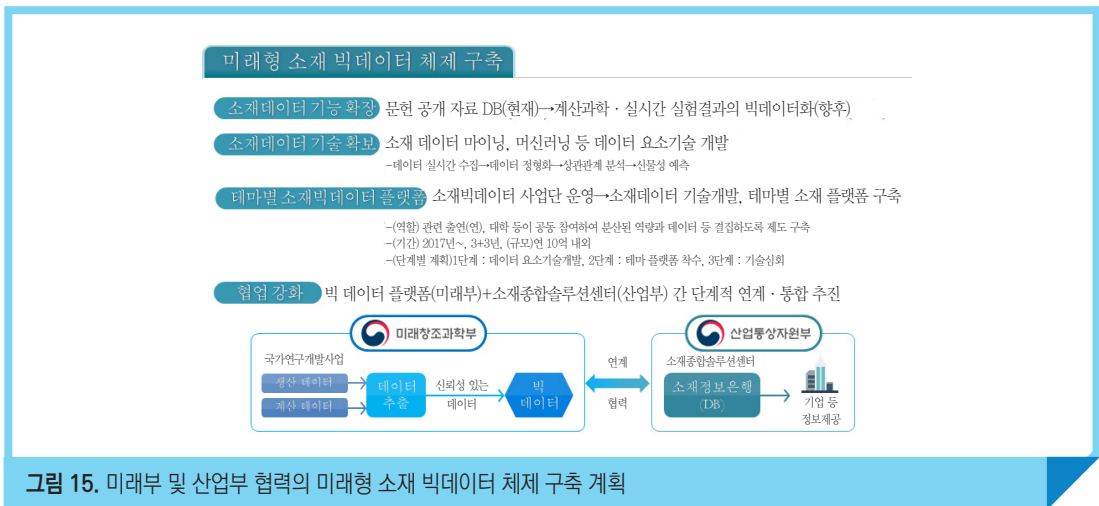


그림 15. 미래부 및 산업부 협력의 미래형 소재 빅데이터 체제 구축 계획

그러나 우리나라의 데이터 기반 연구개발 경쟁력을 향상시키기 위해서는 지속적이고 전략적인 노력이 필요하다. 빅데이터에 기반한 정보플랫폼은 합리성에 기반한 귀납적 모형을 만들어 내는데 필수적이다. 연구 관련 자료들을 포괄하여 수집·관리·통합하는 빅데이터 정보플랫폼은 아직 실현되지 못한 자료의 '잠재가치(potential value)'로부터 '실질가치(real value)'를 창출하는 의미를 지닌다. 즉 정보플랫폼은 방대한 정

① 페르메이팅, 디텍팅, 인가, 분석, 모델링, 학습

② R&D 플랫폼, 인프라, 데이터, 정보, 플랫폼, 연구개발

보로부터 숨어있는 지식에 접근하고, 연구개발 과정에서 빅데이터를 활용할 수 있도록 하는 연구 기반인 것이다. 여기에는 필수적으로 모델링, 계산, DB, 머신러닝, AI를 축으로 한 예측과 설계의 기능이 구현된다. 또한 정보플랫폼은 사용자가 빅데이터를 입력하여 연구의 의미와 과학적 보편지식을 취득하기 위한 접촉면으로도 기능한다. 빅데이터를 입력하고 접근하며, 질의어를 던져 분석하는 등의 기초적 기능도 플랫폼의 중요한 성격이다. 가시화 도구는 빅데이터가 가지는 유용한 추세(trends)를 시각적으로 전달해주며, 사용자가 이해할 수 있는 형태로의 변경을 지원하는 수단이 된다. 직관적이고 사용이 편리한 플랫폼의 형태를 설계한다거나, 내부의 효과적인 계산 알고리즘이 개발되어야 하는 것은 자명하다.

우리나라의 미래 R&D 경쟁력을 위해서는 분야별 빅데이터의 구축, 빅데이터를 활용한 머신러닝과 인공지능 연구개발, 그리고 빅데이터 플랫폼 개발에 역량을 집중해야 할 때이다. 따라서, 데이터 과학과 공학 및 자연과학 분야 전문가뿐만 아니라, 국가 R&D 생태계의 모든 관련자들이 힘을 모아 R&D 데이터를 축적할 수 있는 체계와 표준화, 활용기법의 개발 및 확산, 그리고 과학기술계의 데이터 공유 문화의 정착에 노력해야 할 것이다.

저자 김 찬 수

Chansoo Kim

학력 MIT Nuclear Science 석사
서울대학교 컴퓨터공학 학사

경력 現) 한국과학기술연구원 연구원

저자 오 정 수

Jung Soo Oh

학력 서강대학교 정보통신 석사
한국과학기술원 수학 석사
서울여자대학교 수학 학사

경력 現) 한국과학기술연구원 책임전문원
前) 삼성종합기술원 전문연구원

저자 이 광 렬

Kwang-Ryeol Lee

학력 한국과학기술원 재료공학 박사
서울대학교 금속공학 학사

경력 現) 한국과학기술연구원 기술정책연구소 소장(책임연구원)
前) 한국과학기술연구원 계산과학센터장

참고문헌

1. Bernard Marr, “6 Incredible Ways Big Data Is Used by the US Government”
2. 이영의, “베이지주의”, 한국연구재단 저술총서, 2015, 서울.
3. J. Grus, “Data Science from Scratch : First Principles with Python”, O’reilly.
4. The White House, “The Federal Big Data Research and Development Strategic Plan”, 2016, Washington DC.
5. Materials Genome Initiative for Global Competitiveness, National Science and Technology Council, USA (June 24, 2011). https://www.whitehouse.gov/sites/default/files/microsites/ostp/materials_genome_initiative-final.pdf
6. Materials science with large-scale data and informatics : Unlocking new opportunities, Joanne Hill, Gregory Mulholland, Kristin Persson, Ram Seshadri, Chris Wolverton, and Bryce Meredig, MRS BULLETIN, VOLUME 41, MAY 2016
7. Computational Materials Science Initiative, Japan. <http://www.cms-initiative.jp/en>
8. WPI program in Japan Society for the Promotion of Science. https://www.jsps.go.jp/english/e-toplevel/04_centers.html
9. Roadmap for Materials Modelling (2nd Ed., European Materials Modeling Council, 2015).
10. 이광렬 등, 나노소자 설계용 다차원 가상 실험장치 및 그 방법, 대한민국 특허 10-1474331 (2014.12.12.)
11. 미래부 및 산업부 협력의 미래형 소재 빅데이터 체제 구축 계획, 미래소재 준비계획을 위한 공청회 발표자료, 2016