令和7年度特色入試問題

《薬学部》

論文試験

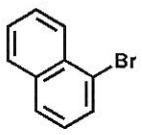
200点満点

(注　意)

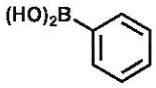
1. 問題冊子および解答冊子は係員の指示があるまで開かないこと。
2. 問題冊子は表紙のほかに18ページある。
3. 解答冊子は表紙のほかに，下書き用紙を含め8ページある。
4. 試験開始後，解答冊子の表紙所定欄に受験番号・氏名をはっきり記入すること。
   表紙には，これら以外のことを書いてはならない。
5. 解答はすべて解答冊子の指定された箇所に記入すること。
6. 解答は指定された場合を除き日本語で記述すること。
7. 解答に関係のないことを書いた答案は無効にすることがある。
8. 解答冊子は，どのページも切り離してはならない。
9. 問題冊子は持ち帰ること。解答冊子は持ち帰ってはならない。

問題 I

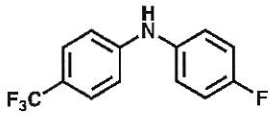
分子を連結する化学反応は多様な医薬品を合成するために必要不可欠である。資料 1 は 2024 年に開発されたある化学反応に関する論文の紹介記事であり、資料 2 はその原著論文の一部である。これらの英文を読み、以下の問いに日本語で答えよ。構造式は以下の化合物 **I**–**III** にならって示せ。



I



II



III

問 1　Aminative Suzuki–Miyaura coupling とはどのような化学反応か。資料 1 および資料 2 に記載された内容に沿って 100 字以内でまとめよ。

問 2　Aminative Suzuki–Miyaura coupling を開発するために、Onnuch らはどのような課題を克服する必要があったか。また、それらの課題をどのように工夫して解決したのか。資料 1 に記載された内容に沿ってまとめよ。

問 3　Aminative Suzuki–Miyaura coupling の開発は Onnuch らの研究構想に基づき設定した仮説の検証に相当する。Onnuch らの研究構想はどのような点で伝統的なクロスカップリング研究と異なった価値観を有するか。資料 2 の 2 段落目に記載された内容に沿って説明せよ。

問 4　上に示す化合物 **I** と化合物 **II** を用いて (a) aminative Suzuki–Miyaura coupling および (b) カルボニルの挿入を含む aminative Suzuki–Miyaura coupling の四成分変法を実施すると、どのような化合物が主生成物として得られるか。解答欄にそれぞれ期待される主生成物の構造式を示せ。

問 5　アリールトリフラートとアリールボロン酸を用いて上に示す化合物 **III** を aminative Suzuki–Miyaura coupling で合成したい。必要となるアリールトリフラートとアリールボロン酸の組み合わせとして適当な 1 例を解答欄に構造式で示せ。

問 6　創薬研究を進展させるためにはどのような新しい化学反応の開発が必要か。資料 2 の著者らの考えを踏まえて、具体的な案を挙げながら、あなたの考えを述べよ。

（注釈）アルファベット順

Ar: アリール基。芳香環 (ベンゼン環など) から水素原子一つが失われてできる残基の名前。

Aryl amine: アリールアミン。芳香環の水素原子をアミノ基で置換した化合物のこと。

Arylboronic acid: アリールボロン酸。芳香環の水素原子を $B(OH)_2$ 基で置換した化合物のこと。

Aryl halide: ハロゲン化アリール。芳香環の水素原子をハロゲン原子で置換した化合物のこと。

Aryl triflate: アリールトリフラート。芳香環の水素原子を CF_3SO_3 基で置換した化合物のこと。クロスカップリング反応においてハロゲン化アリールと同様に用いられる。

Biaryl compound: ビアリール化合物。芳香環同士が炭素原子同士で直接連結された化合物のこと。

Bu: C_4H_9 (ブチル基)。

Cross-coupling: クロスカップリング反応。異なる二種類の化合物が異なる官能基の間で縮合反応を起こし、新しい共有結合を生成する反応のこと。

Cy: C_6H_{11} (シクロヘキシル基)。

DBU: ジアザビシクロウンデセン。水酸化カリウムの代わりに利用している強塩基の一種。

Diarylamine: ジアリールアミン。アミンにアリール基が二つ置換した化合物のこと。

DPPF: 1,1′-ビス(ジフェニルホスフィノ)フェロセン。ホスフィン配位子の一種。

Electrophile: 求電子剤。求電子的な性質があり、求核剤と反応する。

Electrophilic: 求電子的な。

Equiv: equivalent の略。当量のこと。

Et: C_2H_5 (エチル基)。

Halide: ハロゲン化物イオン。

Leaving group: 脱離基。

Ligand: 配位子。金属に配位結合で結合する分子のこと。

Me: CH_3 (メチル基)。

Ms: CH_3SO_2 (メシル基)。

Nitrene: ニトレン。窒素原子と水素原子一つずつからなる分子。

Nucleophile: 求核剤。

Nucleophilic substitution: 求核置換反応。

Organoboron: 有機ホウ素。

Organometallic reagents: 有機金属試薬。狭義では金属元素と炭素が直接結合した化合物のこと。

Oxidative addition: 酸化的付加。

Ph: C_6H_5 (フェニル基)。ベンゼンから水素原子一つが失われてできる残基の名前。

Phosphine: ホスフィン。リン化合物の一種。

Pr: C_3H_7 (プロピル基)。

Precursor: 前駆体。

Reductive elimination: 還元的脱離。

Tf: CF_3SO_2 (トリフリル基)。

資料 1

Nitrogen cuts in during C–C cross-coupling

Transition metal-catalyzed cross-coupling, which creates a bond between two target molecules, is a workhorse method for synthesizing pharmaceuticals, agricultural chemicals, electronic materials, and other fine chemicals. The generation of biaryl compounds by the Suzuki–Miyaura coupling and of aryl amines by the Buchwald–Hartwig coupling reactions are two of the most used transformations in the pharmaceutical industry. Their prevalence has resulted in biaryl and arylamine structures as common motifs in drug candidates. New methods to prepare these structures will expand the chemical space that can be accessed in cross-coupling reactions. Onnuch *et al.* recently reported a palladium-catalyzed aminative Suzuki–Miyaura coupling reaction in which the traditional Suzuki–Miyaura coupling of an aryl (pseudo)halide and an arylboron compound is interrupted by the insertion of nitrogen. This results in the formation of two new C–N bonds in one reaction.

Transition metal-catalyzed coupling reactions are the most widely used methods for C–C and C–heteroatom bond formation. In these reactions, the metal promotes nucleophilic substitution at an electrophilic carbon bearing a leaving group. The nucleophiles are typically organometallic reagents, such as organomagnesium (Grignard compounds), organozinc, or organoboron reagents, and the leaving group is typically a halide or sulfonate. These reactions are highly useful synthetic methods because they can typically be carried out under mild conditions and are tolerant of a wide range of functional groups.

The mechanism for these bond-forming reactions involves three basic catalytic steps: oxidative addition of the carbon electrophile, typically a haloaromatic compound, to the metal center; substitution of the nucleophilic coupling partner for the leaving group on the metal; and reductive elimination to form the product. In Suzuki–Miyaura coupling, the nucleophile is an organoboron reagent, which results in the formation of a new C–C bond. In the case of the Buchwald–Hartwig coupling, a C–N bond is formed from a nitrogen nucleophile.

One method to expand the scope of these reactions is to introduce additional catalytic steps before the bond-forming reductive elimination step, creating multicomponent reactions. For example, addition of carbon monoxide (CO) to cross-coupling reactions leads to CO incorporation between the electrophilic and nucleophilic coupling partners. In Suzuki–Miyaura coupling, the result is a ketone product, whereas amides are prepared through the palladium-catalyzed coupling of aromatic halides, CO, and amines. These reactions are highly selective for the carbonylative product because CO insertion into the metal-carbon bond occurs much faster than the subsequent steps of the catalytic cycle.

Onnuch *et al.* extended this atom insertion concept by achieving an aminative Suzuki–Miyaura coupling through the three-component coupling of a nitrene precursor, an aryl halide or triflate (CF_3SO_3–), and an arylboronic acid. In this reaction, the nitrene unit is introduced during the Suzuki–Miyaura catalytic cycle, resulting in the formation of two new C–N bonds. Successful development of this reaction required overcoming several potential challenges. Modern Suzuki–Miyaura catalysts afford high rates for C–C bond formation. Selective formation of the diarylamine product requires efficient nitrogen insertion before the C–C

reductive elimination step. In addition, the nitrene reagent must efficiently react with the arylpalladium(II) intermediate but not with the palladium(0) species responsible for insertion into the carbon-leaving group bond.

These challenges were overcome by Onnuch *et al.* through the appropriate choice of the nitrene precursor and palladium catalyst. *O*-Diphenylphosphinyl hydroxylamine (DPPH) was the optimal nitrene precursor, whereas less electrophilic nitrogen sources were less selective for nitrogen insertion. Sterically demanding, electron-rich phosphine ligands were also critical to achieving selective formation of the desired unsymmetric diarylamine product.

This methodology opens new avenues for the synthesis of pharmaceuticals and other fine chemicals through late-stage functionalization of halogen-containing drug molecules to incorporate arylamine moieties. The nitrogen insertion approach can be used to create new potential drugs from existing drug compounds containing a biaryl structure. No other changes are required to the other steps in the synthesis to introduce nitrogen in this way.

Onnuch *et al.* further demonstrated the potential utility of this approach with a four-component coupling of an aromatic bromide, CO, DPPH, and an arylboronic acid to give an *N*-aryl benzamide derivative. The nitrogen insertion approach was also applied to the coupling of allyl acetates and arylboronic acids to give *N*-allylaniline derivatives in modest yield.

The nitrogen insertion approach developed by Onnuch *et al.* represents a groundbreaking new avenue in metal-catalyzed cross-coupling in which heteroatoms can be introduced in traditional C–C bond–forming reactions. There is the potential to apply this approach to other classes of coupling reactions beyond the Suzuki–Miyaura coupling. Although an exciting development, the method must overcome some challenges to become widely applicable. Yields range from modest to high, which likely stems from undesired side reactions. In some cases, the arylboron electrophile and the nitrogen reagent react, leading to undesirable aniline side products. In addition, further optimization is needed to afford consistently high selectivity for the diarylamine product over the biaryl product of Suzuki–Miyaura coupling. Additional development of the catalyst system will open the door for wider application of this method for late-stage introduction of nitrogen and other heteroatoms into target molecules.

(出典)

Author: Kevin H. Shaughnessy

Source: Science, Vol. 383, 954 (2024)（一部改変）

資料 2

Aminative Suzuki–Miyaura coupling

Transition metal–catalyzed cross-coupling reactions have become indispensable tools for the synthesis of important organic compounds, such as therapeutics, agrichemicals, energy-storage materials, and functional polymers. Over the past half century in medicinal chemistry, three of the 20 most frequently practiced transformations are palladium (Pd)–catalyzed cross-couplings (Suzuki–Miyaura, Sonogashira, and Buchwald–Hartwig). Over time, the popularization of cross-coupling has considerably influenced which sectors of chemical space are heavily emphasized during drug discovery and, therefore, the structures of recently approved small-molecule pharmaceuticals. For example, there has been a proliferation of (hetero)biaryl and aryl amine motifs because of the reliability and generality of Suzuki–Miyaura and Buchwald–Hartwig catalysis (Fig. 1A). These examples suggest that new, general strategies to expand the product space of essential cross-coupling schemes can enhance structural diversity during candidate generation and improve the speed and success rate of pharmaceutical development.

Traditionally, research aimed at broadening the synthetic utility of cross-coupling methodology has focused on the development of catalysts and reaction conditions that engage distinct reactive partners (electrophiles or nucleophiles), a campaign punctuated by major recent achievements such as the fluorination and trifluoromethylation of aryl electrophiles, carbon-carbon (C–C) coupling from alkyl electrophiles, reductive cross-electrophile couplings, and activation of carbon-hydrogen (C–H) bonds for cross-coupling. An attractive but rarely explored research strategy involves the repurposing of classical, widely available coupling partners by means of diverted pathways that generate alternative, high-value products. In the past decade, late-stage insertion and deletion reactions have attracted tremendous attention as a strategy to generate structural diversity. By analogy, we asked whether the insertion of a bridging atom between the nucleophilic and electrophilic components could be a universal modification to cross-coupling reactions that generates new products from existing partners. The well-established carbonylative Stille cross-coupling represents an example of this approach, but the generalization of the concept to insertions of other ambiphilic components, especially heteroatomic ones, appears to have escaped systematic consideration.

To demonstrate the proposed concept, we pursued the introduction of a formal nitrene insertion into the Pd-catalyzed Suzuki–Miyaura cross-coupling pathway, rerouting its endpoint from biaryl products (C–C linkage) toward diaryl amines (C–N–C linkage), a privileged substructure class among bioactive compounds (Fig. 1B). Many industrial research operations maintain extensive libraries of custom aryl halides (or pseudohalides) and boronic acids (or esters), and we envisioned that through the addition of a simple reagent, these building blocks could be conveniently repurposed to furnish amines. This type of scaffold change from biaryl to diaryl amine, previously inaccessible in a single operation, could be useful for fine tuning the geometry, polarity, and H-bonding ability of many candidates. Achieving this goal would effectively unite the two most prominent metal-catalyzed coupling manifolds (Suzuki–Miyaura and Buchwald–Hartwig) by connecting their products to common precursor pools. Without requiring the separate synthesis and purification of new reagents, the chemical space accessible from existing functionalized intermediates could

be multiplicatively increased.

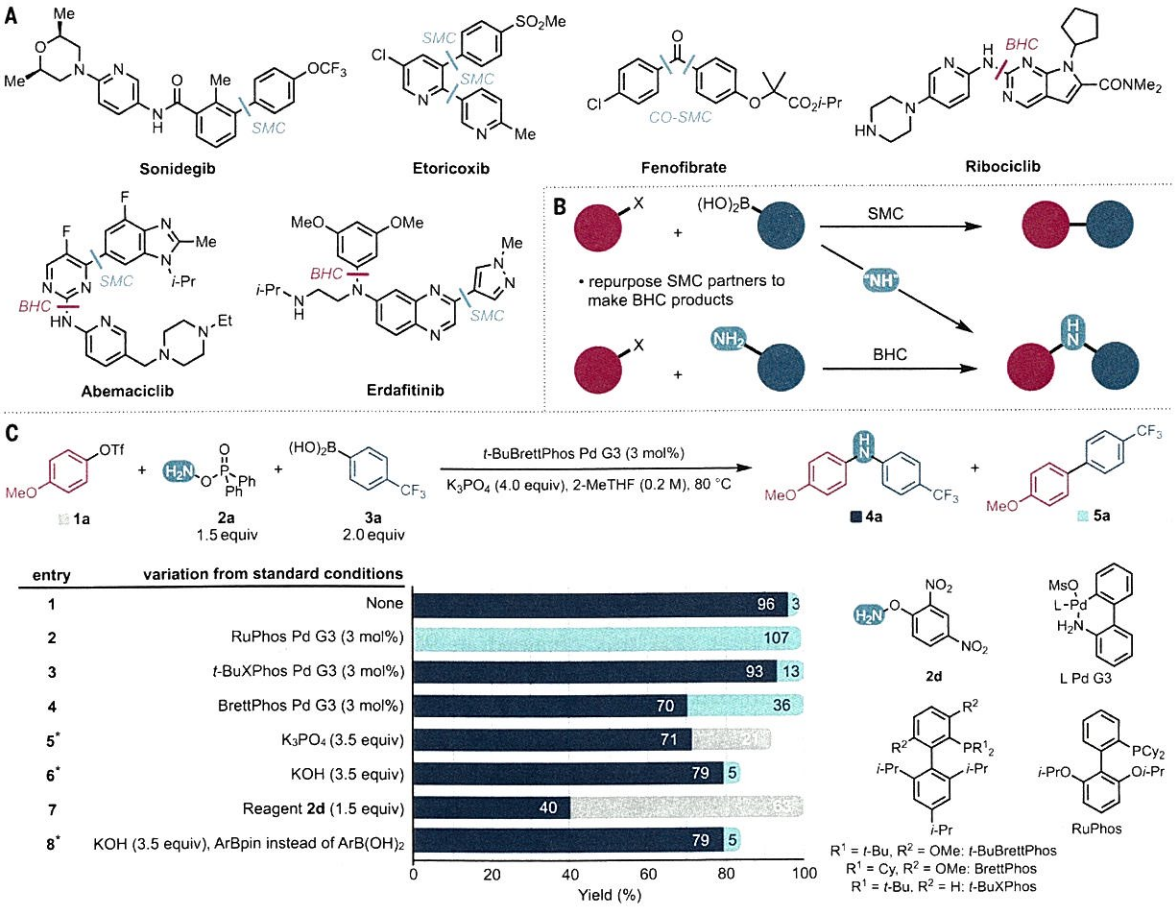


Fig. 1. Background and concept.
(**A**) Suzuki–Miyaura coupling (SMC) and Buchwald–Hartwig coupling (BHC) in drug development. (**B**) This work: SMC with NH insertion to access BHC products. (**C**) Selected results from reaction optimization. Asterisk (*) indicates **2a** (1.1 equiv), **3a** (1.5 equiv).

We anticipated that realization of the intended three-component coupling might be met with several distinct challenges. First, modern Pd-based catalysts perform Suzuki–Miyaura coupling so efficiently that for an amine insertion to intercede, the original process would likely need to be decelerated, either through deactivation of the catalyst toward reductive elimination or inhibition of transmetallation with the aryl boron nucleophile. However, any attenuation of reactivity must be carefully balanced; after *N*-insertion, the metal center must still be capable of achieving the second C–N bond formation. Likewise, the reactivity of the nitrene reagent must be precisely adjusted: It must be electrophilic enough to efficiently insert, yet it should avoid reacting prematurely with Pd(0) before oxidative addition of the aryl halide. Last, there remains the task of avoiding homocoupling processes that install two of the same aryl group on the product rather than one derived from each coupling partner.

We report that the combination of a bulky catalyst and commercially available amination reagent affords a convenient and highly general solution. Our protocol is effective for all common classes of electrophiles (aryl chlorides, bromides, triflates, and tosylates) and compatible with an exceptionally broad scope of polar

functional groups, substitution patterns, and heterocyclic partners relevant to medicinal chemistry. The strategy is easily used on late-stage intermediates to prepare amine-inserted variants of drug candidates, and preliminary results suggest that the insertion concept can be extended to other reaction classes (allylic substitution) and even four-component variants (ArX + CO + NH + Ar'M). The cross-selectivity of this reaction is notable because mechanistic experiments indicate that multiple competing mechanisms likely operate simultaneously.

Reaction development

At the outset of our investigations, we examined the reaction between 4-methoxyphenyl triflate (**1a**) and 4-(trifluoromethyl)phenylboronic acid (**3a**) in the presence of a variety of electrophilic amination reagents as formal precursors of parent nitrene ("NH"). Using catalysts supported by typical phosphine ligands (such as RuPhos) (Fig. 1C, entry 2), complete conversion to Suzuki–Miyaura coupling products was observed after 12 hours, with no apparent participation of the amine reagent *O*-diphenylphosphinyl hydroxylamine (DPPH, **2a**). By contrast, when a *t*-BuBrettPhos-modified Pd catalyst was used under optimized conditions (Fig. 1C, entry 1), the desired aminative coupling product (**4a**) was obtained in 96% after 12 hours with only trace Suzuki–Miyaura product (**5a**). The use of *t*BuXPhos, a ligand with similar steric properties and scaffold to *t*-BuBrettPhos, was found to be nearly equally effective (Fig. 1C, entry 3). However, BrettPhos, a ligand typically used in Buchwald–Hartwig cross-coupling between (het)aryl (pseudo)halides and anilines, could only catalyze the reaction with diminished yield (70%) (Fig. 1C, entry 4) and selectivity (36% **5a** was formed). Relative to the optimized conditions, decreasing equivalents of **2a** and **3a** were associated with incomplete conversion and lower yield of **4a** (Fig. 1C, entry 5). A stronger base, such as potassium hydroxide (KOH), could be used to restore high yield and full conversion, but in polar solvents—such as *N*,*N*′-dimethylformamide (DMF), acetonitrile (MeCN), and 2-methyltetrahydrofuran (2-MeTHF) (Fig. 1C, entry 6)—triflate decomposition to phenol was a competing side reaction. Other ambiphilic aminating agents such as **2d** (Fig. 1C, entry 7), which have been previously reported to effect amination of aryl boronic acids, were ineffective in this context. Instead of boronic acid **3a**, its pinacol ester displayed equally efficient reactivity provided that KOH was used as the base (Fig. 1C, entry 8).

〈中略〉

Applications and extensions

〈中略〉

Late-stage modification of Loratidine and Fenofibrate provided **4al** and **4am**, respectively, in good yields (Fig. 2A). These examples illustrate the power of this reaction in providing direct access to new drug candidates without introducing additional operations or intermediates on all stages of drug synthesis.

Formal nitrogen insertion into cross-coupling reactions is a concept readily generalizable beyond the

Suzuki–Miyaura couplings shown above. For example, by tandem insertion of NH and a carbonyl (C=O) group, Suzuki–Miyaura coupling partners can be used to make amides as an alternative to traditional amide-bond formation, which is one of the most frequently used reactions in medicinal chemistry. As an example, **1w** and **3b** were coupled in the presence of **2a** and iron pentacarbonyl [$Fe(CO)_5$] as the carbonyl source to produce **6ao** in good yield (55%) (Fig. 2B).

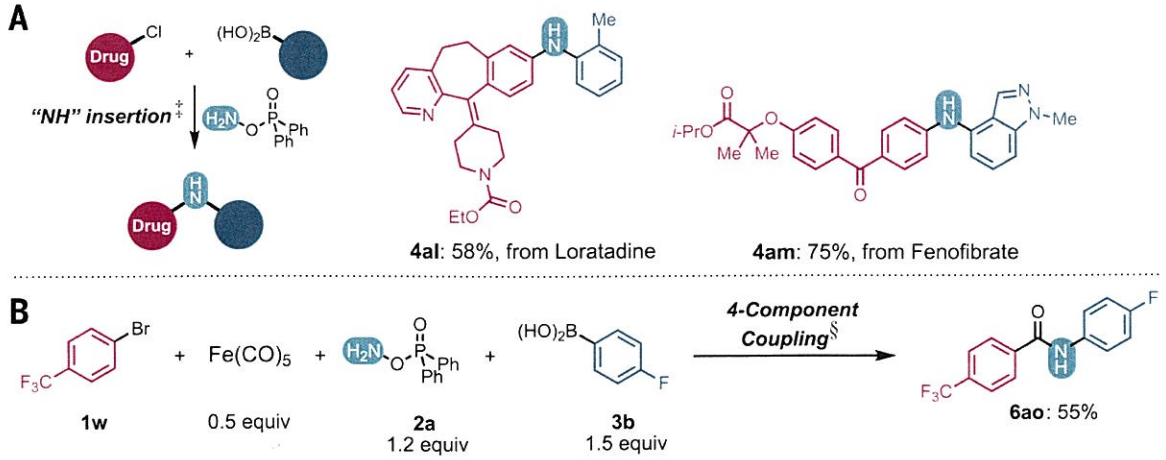


Fig. 2. Applications of aminative Suzuki–Miyaura coupling.
(**A**) Late-stage modification of ArCl-containing drugs. (**B**) Four-component coupling involving sequential NH and CO insertion. "‡" indicates [Pd] (2 mol %), *t*-BuBrettPhos (1 mol %), **2a** (1.1 equiv), $ArB(OH)_2$ (1.2 equiv), KOH (3.0 equiv), MeCN (0.2 M), 80°C. "§" indicates [Pd] (2 mol %), DPPF (0.6 equiv), DBU (3.0 equiv), MeCN (0.2 M), 80°C.

〈中略〉

Mechanistic insights

A plausible pathway for initial C–N bond formation to take place from the aryl electrophile (Fig. 3, mechanism I, "electrophile-first") could involve a 1,2-shift of the aryl ligand from Pd(II) to a coordinated and deprotonated **2a**, which generates the amido complex $LPd(NHAr)OPOPh_2$. This process resembles that proposed by Knochel and coworkers for the electrophilic amination of organozinc reagents with organic azides. The resulting Pd(II) phosphinate can undergo transmetallation with a boronic acid and reductive elimination to afford the desired product. The idea that amination of the aryl electrophile might precede C–N bond formation with the aryl nucleophile is consistent with observation of aniline derived from the former in some cases.

〈中略〉

During the course of our studies, it became clear that initial C–N bond formation from the aryl nucleophile side could also be viable (Fig. 3, mechanism II, "nucleophile-first"). Electrophilic amination of boronic acids by reagents such as **2d** has been reported, although such transformations are typically limited to electron-rich substrates, and many reagents competent for this process are not effective in our three-component coupling.

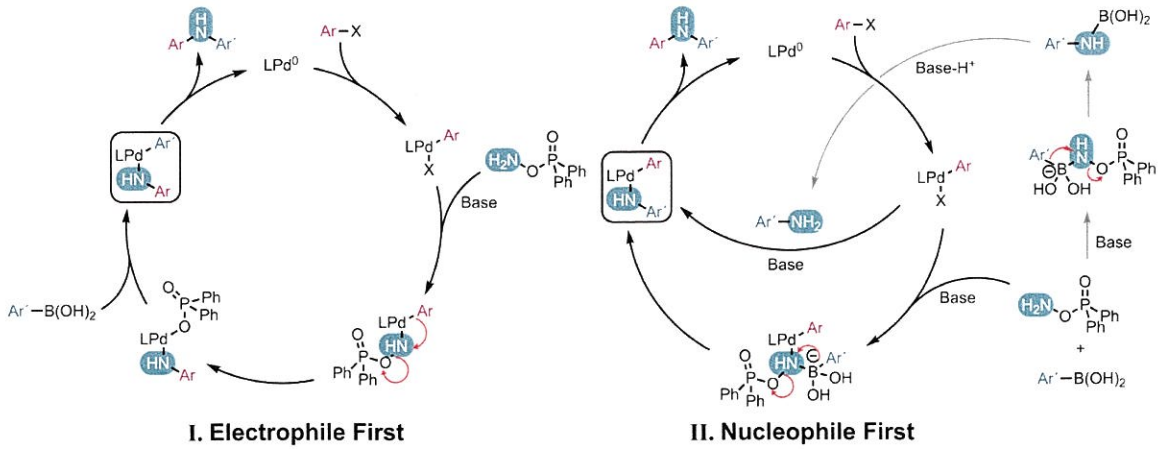


Fig. 3. Proposed mechanisms.

Under our optimized conditions, for some substrate combinations, Pd-independent formation of aniline from arylboron can occur.

〈中略〉

Work continues in our laboratory to achieve a more detailed characterization of available pathways and to elucidate the effect of substrate structure on which of the nucleophile-first or electrophile-first mechanisms is preferred. However, at this point, our studies have definitively established the possibility of forming insertive cross-coupling products through either order of events. Looking forward, we argue that this mechanistic flexibility portends favorably for the extension of the aminative cross-coupling concept to diverse classes of both nucleophiles and electrophiles.

(出典)

Authors: Polpum Onnuch *et al.*

Source: Science, Vol. 383, 1019-1024 (2024)　(一部改変)

<白 紙>

問題Ⅱ

2023 年のアルバート・ラスカー医学研究賞は、タンパク質構造予測プログラム AlphaFold2 を開発した 2 名の研究者に授与された。AlphaFold（AlphaFold2 およびその第一世代である AlphaFold を含む）に関する資料 1 および資料 2 を読み、以下の問いに答えよ。

問 1　下線部(1)は、AlphaFold2 の開発チームが立てた開発方針を述べたものである。これを達成するために開発チームが採用したシステムの訓練方法は、進化の過程におけるどのような生物学的現象を模倣したものか、具体的に説明せよ。

問 2　下線部(2)における“quest”とは何か。その具体的な内容を資料 1 から抜き出し日本語で答えよ。

問 3　Fig. 2a は、ヒト、*Trypanosoma cruzi* (*T. cruzi*)、*Mycobacterium tuberculosis* (*M. tuberculosis*)、*Escherichia coli* (*E. coli*)の 4 つの生物種がそれぞれ発現するタンパク質に対して、AlphaFold を使って立体構造を予測したときの信頼度スコアをまとめたものである。ヒト、*T. cruzi* での信頼度スコアは、*M. tuberculosis*、*E. coli* とは明らかに異なった分布を示す。このような分布の違いが生じる理由を推定せよ。

問 4　3-oxo-5-alpha-steroid 4-dehydrogenase 2 は男性ホルモンのテストステロンを、より強力なジヒドロテストステロンへ変換する酵素である。Fig. 3a は AlphaFold により構築されたこの酵素の立体構造である。Fig. 3b は Fig. 3a を矢印の方向から見たもので、その二次構造をリボンモデルで示したものである。Fig. 3b 中で赤色スティックで示される部位はどのようなアミノ酸の側鎖か答えよ。またその位置や配向方向からどのようなことが推定できるかを述べよ。

問 5　AlphaFold は将来、医薬品開発にどのように利用されると期待されるか。資料 1、資料 2 に記載の内容にもとづいて解説せよ。

（注釈）アルファベット順

Å: オングストローム。0.1 nm（1×10^{-10} m）に相当。タンパク質での原子間の距離を議論する際に適している。

CASP: Critical Assessment of Structure Prediction の略。参加者は2年ごとに「立体構造が解明されたが、まだ論文等としては未発表のタンパク質」のアミノ酸配列を受け取り、自身が開発したシステムを適用し、その立体構造のモデルを計算する。そのモデルを構造の解かれたモデルと比較することで、採点が行われる。

ClinVar: ヒトゲノムの多様性とそれらに関連する疾患についての情報を収集した公開アーカイブ。

DrugBank: 薬物に関する情報を包括的に収集したオンライン公開データベース。

EMBL-EBI: European Molecular Biology Laboratory-European Bioinformatics Institute の略。欧州分子生物学研究所（EMBL）の一部門の欧州バイオインフォマティックス研究所（EBI）。

Escherichia coli (*E. coli*): 大腸菌。腸内細菌の一種。

MSA: Multiple Sequence Alignment の略。同程度の長さの複数のアミノ酸配列を並べて比較したもの。

Mycobacterium tuberculosis (*M. tuberculosis*): 結核菌。ヒトの結核の原因となる細菌。

Proteome: 遺伝子配列をもとに発現するタンパク質の総体。

PDB: Protein Data Bank の略。タンパク質、核酸、糖鎖など生体高分子の3次元構造の原子座標を蓄積している国際的な公共のデータベース。

Trypanosoma cruzi (*T. cruzi*): クルーズトリパノソーマ。シャーガス病を引き起こす原虫。

Uniprot: アミノ酸配列とその機能情報を掲載している代表的なデータベース。

X-ray crystallography: X線結晶解析。タンパク質結晶にX線を照射することで生じる回折像を元にタンパク質の立体構造を計算する手法。

資料 1

A daunting problem

The body's proteins execute a plethora of vital roles inside cells. Their diverse capabilities are intimately linked to the forms that they take after they fold from linear amino-acid chains into three dimensions. Insights into structure can illuminate function and unlock biological mysteries.

More than 60 years ago, the late Christian Anfinsen (National Institutes of Health) showed that an unfurled protein could regain its shape unaided and concluded that its amino-acid sequence encodes its final organization. As a nascent chain configures itself, it cannot try every possibility: Sampling all arrangements would take longer than the age of the universe, even for a modest-sized protein. Yet inside cells, folding can occur within milliseconds, so nature somehow deciphers the problem. In theory, at least, scientists could discern the guidelines that steer amino-acid chains into correct conformations.

Using multiple approaches, hordes of investigators forged tactics that they hoped would capture this information well enough to mold a protein's architecture from its sequence. They attempted to express physical interactions in energy equations and looked toward X-ray crystallography and, eventually, other methods to produce templates that could serve as blueprints for related proteins. They also combined knowledge about the chemical proclivities of specific amino acids—whether they carry a charge, for instance—with their location along a chain to gain hints about a protein's structural features.

〈中略〉

Bringing AI into the fold

By 2018 and the 13th CASP competition, scientists had introduced machine learning into their prediction schemes. In contrast to traditional AI approaches that rely on pre-conceived logic, machine-learning systems discover patterns for themselves from data. By making machine learning the central component of their protein-structure prediction network, Hassabis and Jumper's team won CASP13 with a hefty lead in accuracy over the runner-up and almost a 50% improvement since the best of CASP12. Despite this success, the DeepMind researchers were unsatisfied: They wanted a tool that experimentalists would find useful, with errors of less than one angstrom, the size of an atom.

Hassabis, Jumper, and the AlphaFold team started over and brainstormed intensively. They added geometric and genetic concepts, and they integrated established wisdom about proteins. Atoms have characteristic radii, for instance, and bonds have characteristic angles. The group aimed to include these factors in such a way that they didn't interfere with the system's power to learn for itself.

The researchers devised ways to extract maximal information from limited experimental data, and they deployed strategies that force AlphaFold2 to learn efficiently. They allowed the network to adjust calculations anywhere in its process—all the way back to the beginning—as it toils. This innovation avoided a previous pitfall of locking in early errors. Throughout, the system iteratively hones its developing structural model by re-feeding itself tentative solutions.

Hassabis, Jumper, and colleagues also discarded principles that had guided traditional algorithms. For

example, they ignored linear proximity in favor of three-dimensional relationships, as amino acids that lie hundreds of subunits away can reside together in a folded protein. Moreover, the team boosted the importance of physical closeness by inventing a mechanism that pays special attention to amino acids that are in contact.

No single element was decisive on its own; rather, many ingenious new ideas combined to achieve a breakthrough performance.

〈中略〉

Rigorous training

To train the system, Hassabis and Jumper's team used the PDB's experimentally established structures. AlphaFold2 repeatedly compared its proposals to the real answer and gradually nudged its solutions closer to reality. By repeating this process on every member of the training set, the algorithm absorbed principles of protein structure.

The researchers exploited tricks that pushed the network to learn better. For instance, they hid amino acids in the MSA and asked it to fill in the gaps. (1)In this way, they demanded that the system master rules of evolutionary relationships. They also recursively supplied outputs from any given step, which provided many opportunities for AlphaFold2 to reconsider and refine.

AlphaFold2 also computed how much to trust its predictions, and these confidence ratings allowed the researchers to squeeze more information from the available data and thus, to enhance its performance. After they fed it the roughly 140,000 PDB sequences, they ran another set whose structures had not been solved. From the predictions, they plucked the most reliable 350,000 sequence/structure pairs and trained the system on those data as if they had been experimentally verified.

Retooling protein science

In 2020, AlphaFold2 soared past the competition in CASP14. Its predictions were accurate to atomic precision and it generated excellent results in minutes even for proteins that lacked a template. This was the first approach that could construct high-resolution predictions in cases where no similar structure is known.

In July 2021, Hassabis and Jumper published their method as well as structure predictions of almost every human protein. In only two years, their manuscript's impact has vaulted over almost all of the 100,000 research articles that have been published in *Nature* since 1900. It ranks 50th, having been cited in more than 7000 papers from top journals.

In collaboration with the European Molecular Biology Laboratory's European Bioinformatics Institute, Hassabis and Jumper have shared the program and the database with the scientific community, and more than a million investigators have used these resources. The DeepMind team has since expanded its catalog to almost every known protein in organisms whose genomes have been sequenced. Listings include proteomes of, for instance, viruses that pose epidemic threats and the World Health Organization's high-priority pathogens.

The technology has already made a dramatic impact in myriad biomedical spheres and beyond. It helped researchers fill holes in their visualization of the nuclear pore complex, an enormous and complicated

molecular machine that controls transport into and out of the nucleus. Scientists used the tool to analyze a bacterial syringe that shoots molecules into insect cells. By applying understandings that AlphaFold2 revealed, the investigators reengineered the protein to target human cells, opening a new avenue toward medication delivery and gene therapy. Academic laboratories and companies are harnessing AlphaFold2 to develop vaccines, design drugs, craft enzymes that chew up pollutants, and much more. The prospects are endless.

(2)By letting their imaginations and talents fly, Hassabis, Jumper, and their team completed a quest that had flummoxed scientists for half a century. This triumph has launched a new era in studying and manipulating proteins. It has already catalyzed substantial advances, and its impact and reach promise to explode as workers in a vast range of fields dream up new ways to mine its potential.

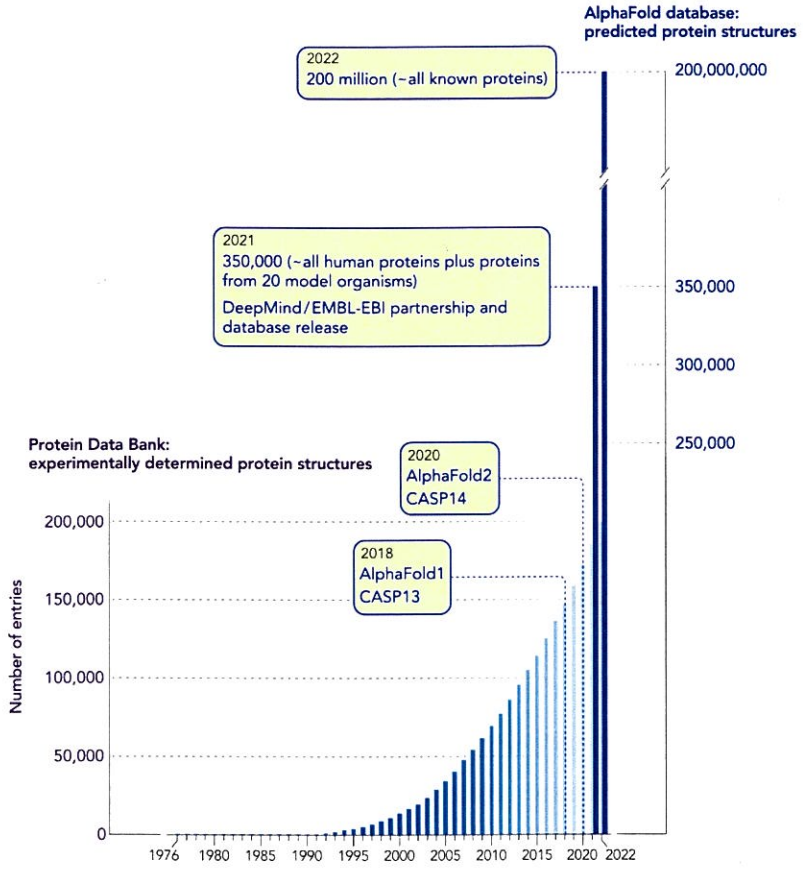


Fig. 1. Predicting success.

Beginning in 1957, when John Kendrew first determined the three-dimensional shape of a protein, scientists have made slow and steady progress on solving protein structures by experimental means. In the last several years, machine learning has catapulted the protein-structure field into a new realm. By 2022, a year after the DeepMind team introduced AlphaFold2, the group had generated predicted structures for almost all known proteins—approximately 200 million of them—an increase of three orders of magnitude over the total number of experimentally solved structures.

（出典）

Author: Evelyn Strauss

Source: AlphaFold—for predicting protein structures, 2023 Albert Lasker Basic Medical Research Award, https://laskerfoundation.org/winners/alphafold-a-technology-for-predicting-protein-structures/（一部改変）

資料 2

資料2は出典のみ公開する。

（出典）

Authors: Janet M. Thornton, Roman A. Laskowski and Neera Borkakoti

Source: Nature Medicine Vol. 27, 1666-1671 (2021)（一部改変）