



創造全新蛋白質助藥物研發

AI模型快速預測兩億種結構

Google AI創辦人等英美三傑奪化學獎



哈薩比斯(Demis Hassabis)

出生日期及地點：1976年7月(48歲)
英國倫敦
任職機構：Google DeepMind
獲獎原因：
利用AI技術預測蛋白質結構
獎金份額：25%

詹珀(John M. Jumper)

出生日期及地點：1985年(39歲)
美國阿肯色州小石城
任職機構：Google DeepMind
獲獎原因：
利用AI技術預測蛋白質結構
獎金份額：25%

蛋白質分子通常由20種氨基酸組成，組合方式千變萬化。1960年代以來，研究人員主要使用一種名為「X射線晶體學」的方式，確認蛋白質結構。科學家們發現，蛋白質的三維結構完全由其氨基酸序列決定，但一個蛋白質可能的結構種類多達10的47次方，如果氨基酸鏈是隨機摺疊，單是分析一種蛋白質的結構，就要耗費較宇宙年齡還要長的時間，窮舉尋找答案。

準確預測蛋白質三維結構

伴隨研究進展，人們發現知道蛋白質的氨基酸序列後，可分析其不同部分作用力，預測蛋白質的結構。1994年，一項名為「蛋白質結構預測批判性評估」(CASP)的競賽誕生，全球各地的科學家們定期相聚「猜謎」，謎面是一種蛋白質的氨基酸序列，謎底則要依照該序列，預測蛋白質的三維結構。

2018年，哈薩比斯團隊利用AI模型AlphaFold參加CASP競賽，其預測準確率接近60%，以絕對優勢奪冠。不過要將準確率提升到90%以上，需要詹珀的協助，推出全新版本AlphaFold2，將所有已知蛋白質結構和氨基酸序列資料庫，全部用於AI訓練。2020年競賽中，AlphaFold2的準確率幾乎與X射線晶體學一致，相較後者需耗時數月甚至多年，AI只需幾分鐘便能給出結果。

Top7人造蛋白質改變世界

人們知道蛋白質的結構，又能否創造新的蛋白質？1990年代，貝克開始研發預測蛋白質結構的電腦軟件Rosetta，他思考如果反向利用軟件，輸入需要的蛋白質結構，由軟件輸出構建氨基酸序列的建議，或許能創造出新的蛋白質。

在貝克團隊的努力下，一款名為「Top7」的人造蛋白質問世，它由93個氨基酸組成，結構極為穩定，與任何天然蛋白質都不同，自此打開「從頭設計蛋白質」世界的大門。這意味着科學家們想要設計有特定功能的蛋白質，除調整天然蛋白質結構外，還可以從零開始將它製造出來。

諾貝爾獎委員會稱，截至本月，全球有來自190個國家超過200萬人使用過AlphaFold2，人類探索蛋白質結構的效率已大幅提升，更可以自行創造蛋白質，其成果在納米材料作標靶藥和疫苗等領域都能發揮作用。貝克將獲得1,100萬瑞典克朗(約827萬港元)獎金的一半，另一半由哈薩比斯和詹珀平分。

母親新加坡華人 哈薩比斯曾研AlphaGo橫掃棋壇

香港文匯報訊 身為軟件工程師和人工智能(AI)專家的哈薩比斯，似乎並非傳統的諾貝爾化學獎得主，不過他在AI領域雄心勃勃。《華爾街日報》早前報道，哈薩比斯曾告訴員工，應該利用AI技術作出「值得獲得諾獎的發現」。哈薩比斯也將助其今次獲獎的「AlphaFold2」模型，稱作他最喜歡的AI現代應用。

哈薩比斯出生於倫敦，母親是新加坡華人，父親有希臘和塞浦路斯血統。年少的哈薩比斯熱衷國際象棋與程式設計，從劍橋大學獲得電腦科學學士學位後，曾創辦遊戲工作室，其後轉向探索AI，在倫敦大學學院獲得認知神經科學博士學位，畢業後創立DeepMind公司。

去年奪「諾獎風向標」

2014年，DeepMind被科企Google收購；2016年，其AI產品「AlphaGo」戰勝多名圍棋世界冠軍，自此走紅。不過哈薩比斯認為，AlphaGo是為研發更好的AI模型積累經驗，解決對人類更重要的問題，他因此參加2018年「蛋白質結構預測批判性評估」(CASP)競賽、向

獲得諾獎邁出重要一步。

《華爾街日報》訪問多名哈薩比斯的同事，他們形容哈薩比斯要求嚴格、富有好勝心，常身穿低調圓領冷衫。哈薩比斯常年留在家鄉倫敦，他崇拜英國數學家圖靈，認為自己足夠成為被譽為「電腦界諾獎」的圖靈獎得主。

哈薩比斯與今屆共同獲獎的同事詹珀，去年已憑藉AlphaFold2獲得被視作「諾獎風向標」的拉斯克基礎醫學研究獎。哈薩比斯相信現有技術對於日後建立更強大的AI系統至關重要，「我一直相信人們會普遍使用AI，只是無法預測時機。」



● AlphaGo擊敗包括中國棋手柯潔等冠軍人馬。
網上圖片



本報專訪清華助理教授 師從貝克讚優秀管理者

香港文匯報訊(記者 張瑞豪)北京生命科學研究所研究員、清華大學生物醫學交叉研究院助理教授徐純福，曾在今屆化學獎得主之一貝克的實驗室擔任博士後。徐純福形容貝克待人親和，即使實驗室人數眾多，他仍親力親為，「只有貝克找不到學生的時候，沒有學生找不到貝克的時候。」

設計人造蛋白質如搭積木

就今次化學獎獲獎成果，徐純福用搭積木來解釋，「使用人工智能(AI)模型預測蛋白質結構，就似預測手中的積木，最終會建成什麼建築。貝克的研究則是計算出哪些積木，可以搭建成最終的建築。」

徐純福回憶，貝克不但是偉大的科學家，也是優秀的管理者，能夠遊刃有餘地管理規模龐大的實驗室，在小組會議上親自報告。徐純福在回國建立自己的實驗室前，曾向貝克尋求如

何保持良好心態的建議，貝克亦主動分享經驗稱，「若覺得心情不好，那就不要來實驗室了。」



●徐純福(右)與貝克在今年7月合照。
受訪者提供

專訪北大教授「貝克不斷挑戰更高山峰」

香港文匯報訊(記者 張瑞豪)北京大學化學與分子工程學院教授王初(圖)曾師從今屆化學獎得主貝克，在西雅圖華盛頓大學生物化學系取得理學博士學位。王初形容，師從貝克為他奠定了科研基礎，令他受益終生。王初回憶貝克在工作之餘喜愛爬山，不斷挑戰更高的山峰，永遠在「邁向最高點」的路上。

王初形容，貝克是一名天才且純粹的科學家，始終積極思考科研，他已獲得諸多重量級獎項，一直勇於接受新挑戰。對於今次化學獎，王初認為這是對貝克研究成果的肯定，同時也印證了人工智能(AI)技術在科研領域，正發揮愈來愈重要的作用，期待未來AI能夠結合更多傳統科研領域，帶來新成果造福人類社會。

AlphaFold2如何運作？

作為AlphaFold2研發的一部分，AI模型事先接受訓練，利用人類目前所有已知蛋白質的結構和氨基酸序列作為訓練資料。

第一步：輸入資料、檢索資料庫

將結構未知的蛋白質的氨基酸序列輸入AlphaFold2，AI在資料庫中檢索相似的氨基酸序列和蛋白質結構。

第二步：序列分析

AI模型首先會比對不同物種的蛋白質所有相似氨基酸序列，研究哪一部分在物種進化中被保留。

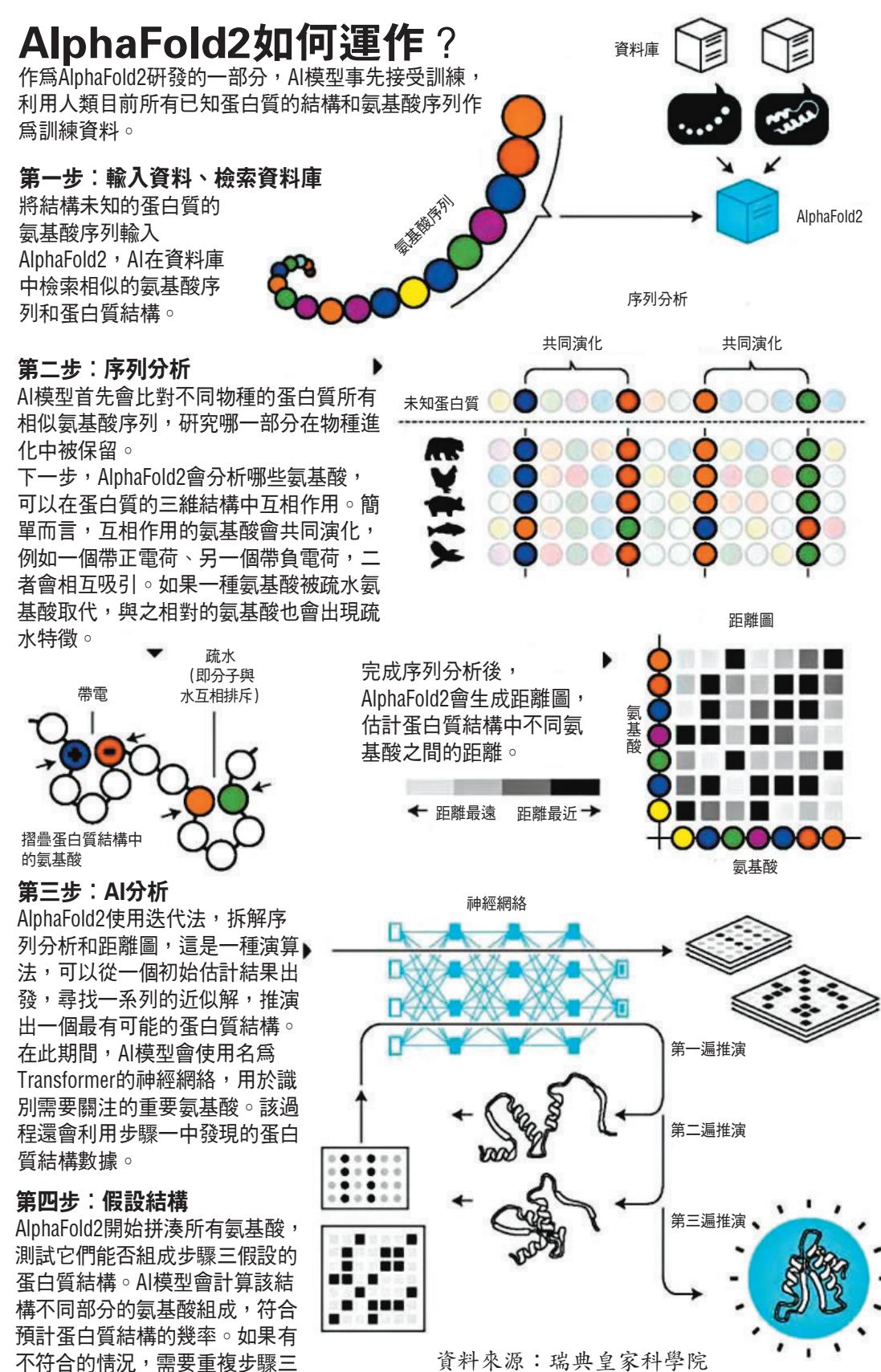
下一步，AlphaFold2會分析哪些氨基酸，可以在蛋白質的三維結構中互相作用。簡單而言，互相作用的氨基酸會共同演化，例如一個帶正電荷、另一個帶負電荷，二者會相互吸引。如果一種氨基酸被疏水氨基酸取代，與之相對的氨基酸也會出現疏水特徵。

第三步：AI分析

AlphaFold2使用迭代法，拆解序列分析和距離圖，這是一種演算法，可以從一個初始估計結果出發，尋找一系列的近似解，推演出一個最有可能的蛋白質結構。在此期間，AI模型會使用名為Transformer的神經網絡，用於識別需要關注的重要氨基酸。該過程還會利用步驟一中發現的蛋白質結構數據。

第四步：假設結構

AlphaFold2開始拼湊所有氨基酸，測試它們能否組成步驟三假設的蛋白質結構。AI模型會計算該結構不同部分的氨基酸組成，符合預計蛋白質結構的幾率。如果有不符合的情況，需要重複步驟三。



資料來源：瑞典皇家科學院