**ฐานข้อมูล SCOP ในปี 2004: การปรับแต่งรวมโครงสร้างและข้อมูลลำดับครอบครัว**

ได้รับวันที่ 17 กันยายน 2546 : ยอมรับวันที่ 18 กันยายน 2003

**บทคัดย่อ**

ฐานข้อมูลการจำแนกโครงสร้างของโปรตีน (SCOP) เป็นคำสั่งที่ครอบคลุมของโปรตีนทั้งหมดของโครงสร้างที่รู้จักตามความสัมพันธ์วิวัฒนาการและโครงสร้าง โดเมนโปรตีนใน SCOP นั้นแบ่งออกเป็นลำดับชั้นตามตระกูลแฟมิลี่โฟลด์และคลาส การสะสมลำดับและข้อมูลโครงสร้างอย่างต่อเนื่องช่วยให้การวิเคราะห์อย่างเข้มงวดมากขึ้นและให้ข้อมูลที่สำคัญสำหรับการทำความเข้าใจโลกโปรตีนและเพลงวิวัฒนาการของมัน SCOP มีส่วนร่วมในโครงการที่มีวัตถุประสงค์เพื่อหาเหตุผลเข้าข้างตนเองและบูรณาการข้อมูลเกี่ยวกับโปรตีนที่จัดขึ้นในฐานข้อมูลลำดับและโครงสร้างทั้งหลาย เป็นส่วนหนึ่งของโครงการนี้เริ่มตั้งแต่ 1.63 เราได้ทำการปรับการจัดหมวดหมู่ SCOP ซึ่งแนะนำการเปลี่ยนแปลงจำนวนมากซึ่งส่วนใหญ่อยู่ในระดับต่ำกว่า superfamily การจัดประเภท SCOP ที่รอดำเนินการจะดำเนินไปเรื่อย ๆ ผ่านการเผยแพร่ในอนาคตหลายรายการ นอกจากชุดขยายลิงก์แบบคงที่ไปยังแหล่งข้อมูลภายนอกแล้วที่ระดับรายการโดเมนเราได้เริ่มต้นการปรับปรุงความสามารถของอินเทอร์เฟสของ SCOP ให้มีการเชื่อมโยงแบบไดนามิกมากขึ้นกับฐานข้อมูลอื่น ๆ สามารถเข้าถึง SCOP ได้ที่ http://scop.mrc-lmb.cam.ac.uk/scop

**ประวัติ**

ฐานข้อมูล SCOP (การจำแนกโครงสร้างของโปรตีน) ได้รับการพัฒนาเป็นการจัดหมวดหมู่วิวัฒนาการซึ่งเป้าหมายหลักคือการวางโปรตีนไว้ในกรอบวิวัฒนาการที่สอดคล้องกันโดยยึดตามลักษณะโครงสร้างที่อนุรักษ์ไว้ ฐานข้อมูลมีวัตถุประสงค์เพื่อให้คำอธิบายที่ครอบคลุมและมีรายละเอียดของความสัมพันธ์ระหว่างโปรตีนทั้งหมดที่มีการกำหนดโครงสร้าง 3 มิติ หน่วยพื้นฐานของการจำแนกประเภทในฐานข้อมูล SCOP คือโดเมนโปรตีน โดเมนถูกกำหนดให้เป็นหน่วยวิวัฒนาการที่สังเกตได้ในธรรมชาติไม่ว่าจะแยกหรือในบริบทมากกว่าหนึ่งในโปรตีนหลายโดเมน โดเมนโปรตีนแบ่งออกเป็นลำดับชั้นเป็นครอบครัว , superfamilies ,และคลาสซึ่งมีการพูดถึงความหมายก่อนหน้า

ข้อได้เปรียบของฐานข้อมูล SCOP คือการรวมทฤษฎีวิวัฒนาการของโปรตีนตามที่ผู้เชี่ยวชาญของมนุษย์กำหนดไว้มากกว่าโดยกฎเชิงประจักษ์ที่นำมาใช้ในอัลกอริทึมและเครื่องมือทางชีวสารสนเทศที่หลากหลาย การสนับสนุนการคำนวณใน SCOP ใช้เพื่อขยายความสามารถของมนุษย์ในการวิเคราะห์และตีความข้อมูลและเพื่อให้ความรู้อันล้ำค่าเกี่ยวกับวิวัฒนาการของโปรตีนที่มีอยู่ในวงกว้างสำหรับนักวิจัยทางวิทยาศาสตร์

การเปิดตัว SCOP อย่างเป็นทางการครั้งแรกเมื่อ 9 ปีที่แล้วประกอบด้วยโดเมนโปรตีน 3,179 กลุ่ม แบ่งออกเป็น 498 ตระกูล, 366 superfamilies และ 279 เท่า เจ็ดคลาสหลักในรุ่นล่าสุด (1.65) มี 40 452 โดเมนที่แบ่งเป็น 2,327 ตระกูล, 1,294 superfamilies และ 800 เท่า โดเมนเหล่านี้สอดคล้องกับ 20 619 รายการใน Protein Data Bank (PDB) และการอ้างอิงวรรณกรรมหนึ่งเรื่องถึงโครงสร้างที่มีพิกัดที่ไม่ได้เผยแพร่ สถิติของรุ่นปัจจุบันและก่อนหน้าสรุปและประวัติการเปลี่ยนแปลงและข้อมูลอื่น ๆ สามารถดูได้จากเว็บไซต์ SCOP ( [http: //scop.mrc‐lmb. cam.ac.uk/scop/](http://scop.mrc-lmb.cam.ac.uk/scop/) ) พร้อมด้วยไฟล์แยกวิเคราะห์ SCOP ทั้งหมด ข้อมูล . ลำดับและโครงสร้างของโดเมน SCOP พร้อมใช้งานจาก ASTRAL compendium และโมเดล Markov ที่ซ่อนอยู่ของโดเมน SCOP มีให้บริการจากฐานข้อมูล SUPERFAMILY

ที่นี่เรานำเสนอการปรับปรุงเพิ่มเติมและคุณสมบัติใหม่ที่นำมาใช้ใน SCOP ตั้งแต่การอัปเดตก่อนหน้า  เริ่มต้นด้วยการปล่อย 1.63 ส่วนใหญ่ของการจัดหมวดหมู่ SCOP จะถูกจัดระเบียบใหม่เพื่ออำนวยความสะดวกในการรวมของการจัดประเภทโครงสร้างที่มีลำดับร่วมสมัยและรูปแบบการจัดหมวดหมู่การทำงาน ในระดับบนสุดของลำดับชั้น SCOP การเปลี่ยนแปลงเหล่านี้จะมีผลต่อรายการจำนวนเล็กน้อยเท่านั้น (∼20 เท่าและ superfamilies ใน SCOP ได้ถูกจัดประเภทใหม่แล้ว) การจัดเรียงใหม่ที่มีนัยสำคัญ แต่ไม่ชัดเจนจึงกำลังดำเนินการในระดับที่ต่ำกว่าและมุ่งเป้าไปที่การปรับแต่งความสัมพันธ์ระหว่างโปรตีนและตระกูลโปรตีน การเปลี่ยนแปลงที่สำคัญที่แนะนำใน SCOP 1.63 และ 1.65 มีการอธิบายรายละเอียดเพิ่มเติมด้านล่าง

**การจัดประเภทรายการ**

ลักษณะแบบไดนามิกของ SCOP เป็นหนึ่งในคุณสมบัติหลักและจำเป็นต้องนำมาพิจารณาในแอปพลิเคชันที่ใช้ฐานข้อมูล SCOP การสะสมลำดับและข้อมูลโครงสร้างอย่างต่อเนื่องในปัจจุบันช่วยให้การวิเคราะห์อย่างเข้มงวดมากขึ้นและให้ข้อมูลที่สำคัญสำหรับการทำความเข้าใจโลกโปรตีนและละครวิวัฒนาการ หากมีหลักฐานใหม่เกี่ยวกับความสัมพันธ์ของโปรตีน สิ่งนี้อาจส่งผลให้มีการกำหนดขอบเขตของโดเมนใหม่และ/หรือจัดเรียงโหนดใหม่ในลำดับชั้นของ SCOP ตัวอย่างทั่วไปคือเมื่อส่วนหนึ่งของโปรตีนใหม่ที่มีขนาดใหญ่ก่อนจัดเป็นรายการหลายโดเมนเดียวต่อมาถูกสังเกตว่าเป็นโปรตีนเดี่ยวหรือในการรวมกันของประเภทโดเมนที่แตกต่างกันและดังนั้นจึงถูกจัดประเภทใหม่เป็นโดเมนแยกต่างหาก บ่อยครั้งที่โปรตีนที่จำแนกสองชนิดแยกกันแสดงให้เห็นว่ามีความสัมพันธ์กันผ่านตัวกลาง โครงสร้างที่ได้รับการพิจารณาเมื่อเร็ว ๆ นี้ การปรากฏตัวของโปรตีนดังกล่าวในฐานข้อมูลโครงสร้างสามารถช่วยในการระบุความสัมพันธ์ที่ห่างไกลมากขึ้นระหว่างโดเมนโปรตีนและสามารถนำไปสู่การจัดเรียงใหม่ที่รวม superfamilies โปรตีนที่แตกต่างกัน

อีกปัจจัยที่มีอิทธิพลต่อการจัดประเภทใหม่คือการรวมเข้ากับฐานข้อมูลอื่น โครงการได้เริ่มขึ้นในปีที่ผ่านมาโดยมีวัตถุประสงค์เพื่อหาเหตุผลเข้าข้างตนเองและบูรณาการข้อมูล SCOP กับข้อมูลเกี่ยวกับตระกูลโปรตีนที่อยู่ในลำดับที่โดดเด่นและฐานข้อมูลโครงสร้างรวมถึง InterPro , Pfam , CATH และ MSD เหตุการณ์สำคัญในเป้าหมายที่ทะเยอทะยานนี้คือการให้คำจำกัดความที่เข้มงวดและแม่นยำยิ่งขึ้นภายใต้แผนการจำแนกประเภทต่าง ๆ ที่ใช้ในฐานข้อมูลที่แตกต่างกันเหล่านี้ เพื่อตอบสนองความต้องการเหล่านี้โดยเริ่มจากปล่อย 1.63 เราได้ทำการปรับแต่งการจัดหมวดหมู่ SCOP ที่แนะนำการเปลี่ยนแปลงจำนวนมากซึ่งส่วนใหญ่อยู่ในระดับต่ำกว่า superfamily

**โปรตีนเมมเบรน all‐ α**

หนึ่งในการปรับปรุงครั้งใหญ่ในการเปิดตัว SCOP 1.63 คือการแก้ไขเมมเบรนทั้งหมดที่เรียกว่า all alpha fold สร้างขึ้นใน SCOP เมื่อมีโครงสร้างโปรตีนเมมเบรนจำนวนหนึ่งรู้จักโดเมนโปรตีนชนิดพับนี้จัดอยู่บนพื้นฐานของเนื้อหาโครงสร้างรองโดยไม่พิจารณาโครงสร้างทอพอโลยีอย่างชัดเจน ได้รับการแจ้งให้ทราบถึงความก้าวหน้าอย่างรวดเร็วของผลึกโปรตีนเมมเบรนการวิเคราะห์ที่ครอบคลุมของโดเมนเหล่านี้ได้ดำเนินการและโปรตีนทั้งหมดของเมมเบรนได้ถูกจัดประเภทใหม่จากรอยขีดข่วนเป็นรอยพับใหม่หรือเท่าเดิมในฐานข้อมูล SCOP 24 แห่ง ปัจจุบันการพับโปรตีนเหล่านี้ถูกรวมไว้ภายใต้คำจำกัดความการพับที่แม่นยำยิ่งขึ้นตามจำนวนของเอนริเก้ที่แผ่ขยายเยื่อหุ้ม มีการค้นพบความสัมพันธ์เชิงวิวัฒนาการและโครงสร้างใหม่ที่น่าจะเป็นในระหว่างการจัดประเภทใหม่ การค้นพบรอยพับของเฮม‐ใหม่นั้นน่าสนใจที่สุด การพับโปรตีนนี้ประกอบไปด้วยเอ็นเทอร์เมมเบรนสี่เส้นที่เรียงตัวกันเป็นกลุ่ม up และ down โดยมีกลุ่มฮีมที่อยู่ระหว่างเอนริเก้ สี่เท่าขดลวดถูกสังเกตในโครงสร้างของหน่วยย่อยของไซโตโครมขของไซโตโครม เป็นฉันที่ซับซ้อน(1be3: C) หน่วยย่อยของ Escherichia coli จัดรูปแบบดีไฮโดรจีเนส N (1kqf: C) และในเมมเบรนหน่วยย่อยของคอมเพล็กซ์ระบบทางเดินหายใจ fumarate reductase (1qla: C) ลิแกนด์สามในสี่จะได้รับการอนุรักษ์ไว้ไซโตโครมและครอบครองพื้นที่ที่มีโครงสร้างเทียบเท่ากับโหมด haem ‐ binding ของโปรตีนทั้งสองมีความคล้ายคลึงกันมาก คุณลักษณะเหล่านี้พิจารณาร่วมกับความคล้ายคลึงกันของโครงสร้างโดยรวมที่ดีของโดเมนสี่ขดลวดสามารถตีความได้ว่าเป็นหลักฐานของแหล่งกำเนิดวิวัฒนาการทั่วไป ปัจจุบันโดเมนโปรตีนเหล่านี้ประกอบไปด้วย superfamily ของ ทรานเมมเบรน di ‐ haem ไซโตโครม

**capsid ไวรัสและเคลือบโปรตีน**

การจำแนกประเภท SCOP ในอดีตของ capsid ไวรัสและโปรตีนเคลือบอยู่บนพื้นฐานของการสันนิษฐานว่าไวรัสมีการพัฒนาร่วมกับโฮสต์ของพวกเขา โดเมนโปรตีนของเท่านี้ถูกแบ่งออกเป็นหลายครอบครัวตามโฮสต์ที่ติดเชื้อ อย่างไรก็ตามการเพิ่มจำนวนของข้อมูลที่มีอยู่ในโครงสร้างของไวรัสและลำดับจีโนมทำให้เกิดการประเมินแนวคิดการจำแนกประเภทเก่า สัตว์เลี้ยงลูกด้วยนม picornaviruses (เช่นไวรัส ssRNA ที่ติดเชื้อบวก) มีลักษณะทางสัณฐานวิทยาและพันธุกรรมคล้ายคลึงกันมากกับขนาดเล็กที่เรียกว่า Cricket อัมพาตเช่นไวรัสและไวรัสพืชจำนวนหนึ่ง โปรตีนโค้ทของพวกมันก่อตัวเป็นชุดประกอบและแสดงคุณลักษณะที่อนุรักษ์ไว้หลายประการในการพับของพวกเขา ความคล้ายคลึงกันระหว่างสัตว์เลี้ยงลูกด้วยนมและไวรัสแมลงขยายไปถึงกระบวนการผลิตโพลีโปรตีน ในการเปิดตัว SCOP 1.65 โดเมนโปรตีนเหล่านี้จะถูกจัดกลุ่มเข้าด้วยกันและจำแนกเป็นของซุปเปอร์เอสของไวรัส SSRNA การจัดประเภทใหม่ของ capsid ไวรัสและตราโปรตีนผลในสี่ superfamilies ใหม่และ 11 ครอบครัวใหม่ การจำแนกประเภทใหม่อย่างชัดเจนเป็นไปตามอนุสัญญาการตั้งชื่อและอนุกรมวิธานไวรัสที่จัดตั้งขึ้นโดยคณะกรรมการระหว่างประเทศว่าด้วยอนุกรมวิธานของไวรัส (ICTV) นอกเหนือจากการปรับโครงสร้างภายในแล้วการพับโปรตีนนี้ได้ถูกรวมเข้ากับนิวคลีโอพลาสมิน/PNGase F‐แบบพับเดิม

**โดเมนแอนติบอดี**

แอนติบอดีและชิ้นส่วนของพวกมันเป็นกลุ่มโปรตีนที่มีโครงสร้างคล้ายคลึงกันที่ใหญ่ที่สุด ใน SCOP มีโดเมนแอนติบอดี้มากกว่า 2,000 โดเมนที่จัดระเบียบไว้ก่อนหน้านี้ในชุดค่าผสมโดเมนผันแปร 228 ชนิดและชุดค่าคงที่ 185 สปีชีส์ ใน SCOP ปล่อย 1.65 โดเมนตัวแปรและค่าคงที่ทั้งหมดถูกจัดประเภทใหม่ตามห่วงโซ่และสิ่งมีชีวิตต้นทาง โดเมนคงที่ได้รับการจัดเรียงเพิ่มเติมตามลำดับห่วงโซ่ของพวกเขา เป้าหมายหลักของเราคือการจัดเตรียมลักษณะที่เป็นระบบที่ครอบคลุมมากขึ้นของรายการโครงสร้างของโดเมนตัวแปร งานที่ไม่ง่ายโดยคำนึงถึงจำนวนของโครงสร้างแอนติบอดีที่ได้รับการออกแบบทางวิศวกรรมใน PDB ในการวิเคราะห์ของเราเราได้แยกโดเมนไฮบริดและเทียมตัวแปร 51 ตัวออกจากชุดโดเมนและจำแนกเป็นสปีชีส์ 'ที่ออกแบบ' แยกจากกัน เพื่อระบุกลุ่มต่าง ๆ ในโดเมนตัวแปรแอนติบอดีเราได้ทำการจัดกลุ่มลำดับเฟสสองเฟส ขั้นแรกลำดับที่สอดคล้องกับส่วนของเชื้อโรคจะถูกรวมเข้าด้วยกันโดยใช้เกณฑ์ 85% สำหรับการรวมลำดับโปรตีนเข้ากับชุดคลัสเตอร์ จากนั้นส่วนจะถูกจัดเรียงตามขนาดของภูมิภาค CDR1 และ CDR2 เราคาดหวังว่ากลุ่มผลลัพธ์ที่ได้อาจสอดคล้องกับครอบครัวตระกูลเชื้อโรคในจีโนมชนิด

**ตั้งค่าโดเมน**

โดเมน E-set ถูกสันนิษฐานว่าเป็นโดเมนแรกเริ่มของอิมมูโนโกลบูลินแบบพับและอาจเป็นลิงก์วิวัฒนาการระหว่างอิมมูโนโกลบูลินกับไฟโบรเนกตินประเภทที่ 3 superfamilies ในรุ่น 1.63 ตระกูลโดเมน E-set เดิมถูกนำออกมาจากอิมมูโนโกลบูลิน superfamily ใน SCOP และเปลี่ยนเป็น superfamily โดเมนที่เป็นส่วนประกอบถูกจัดระเบียบใหม่เป็น 15 ตระกูลใหม่ เครื่อง C-terminalโดเมนของ mollusc haemocyanin แบ่งปันเพียงบางส่วนความคล้ายคลึงกันของโครงสร้างด้วยโดเมนที่เหมือนอิมมูโนโกลบูลินของ arthropod haemocyanin ได้รับการจัดประเภทใหม่เป็นรอยพับใหม่

**ไคเนสโปรตีน**

ใน SCOP ปล่อย 1.65 โดเมนตัวเร่งปฏิกิริยาที่เกี่ยวข้องของไคเนสไทร์และ Thr/Ser ถูกรวมเข้ากับไคเนสโปรตีนตระกูลเดียว ในความเป็นจริงความสัมพันธ์ที่ใกล้ชิดของพวกเขาได้รับการยืนยันโดยการกำหนดโครงสร้างของตัวรับชนิด TGF ββ R4 ซึ่งเป็นไคเนสของ Thr/Ser ที่มีความคล้ายคลึงกันมากกว่าในลำดับของไคเนส Tyr มากกว่าไคเนสของ Thr/Ser อื่น ๆ แม้ว่าโดเมนตัวเร่งปฏิกิริยาของไคเนสโปรตีนจะคล้ายกันมาก แต่มีลวดลายบางอย่างที่สามารถระบุได้ในลำดับของพวกเขาและใช้เพื่อระบุคุณสมบัติการทำงานของไคเนสแต่ละชนิด เราใช้คุณสมบัติเฉพาะเหล่านี้เพื่อกำหนดโดเมนโปรตีนไคเนสทั้งหมดของโครงสร้างที่รู้จักให้กับกลุ่มหลักที่กำหนดโดยความจำเพาะของสารตั้งต้นและ หรือโหมดของการควบคุมและจากการทำงานของครอบครัวย่อย สำหรับโปรตีนไคเนสแต่ละรายการตอนนี้ SCOP จะให้คำอธิบายโดยละเอียดในฟิลด์คำอธิบายประกอบ ฟิลด์นี้สามารถค้นหาได้และช่วยให้ผู้ใช้สามารถแยกชุดโปรตีนที่สนใจเป็นพิเศษได้

**รายการที่ไม่ได้ประสานงาน**

SCOP ที่ออกก่อนให้การจำแนกโครงสร้างโปรตีนนับสิบที่ตีพิมพ์ในวรรณคดี แต่ไม่สามารถใช้งานได้ในเวลาจาก PDB จัดเป็นเอกสารอ้างอิงโครงสร้างเหล่านี้เป็นตัวแทนของครอบครัวโปรตีนของพวกเขาในเวลาเดียว หลังจากนำนโยบายการเชื่อมโยงการตีพิมพ์โครงสร้างกับการบังคับใช้พิกัดโดยวารสารทางวิทยาศาสตร์ส่วนใหญ่การจัดหมวดหมู่ของรายการที่ไม่ประสานงานใหม่ใน SCOP ถูกยกเลิกและจำนวนของพวกเขาก็ค่อยๆลดลง พบโปรตีน 28 ตัวที่เหลือ 28 จาก 28 โปรตีนใน SCOP 1.63 โดยการตรวจสอบล่าสุดเพื่อให้มีโครงสร้างตัวแทนที่เกี่ยวข้องอย่างใกล้ชิดใน PDB และถูกทำให้ล้าสมัย ในรุ่นล่าสุดนั้นมี

เพียงหนึ่งการอ้างอิงวรรณกรรมแสดงถึงเอกลักษณ์

**การพัฒนาทางเทคนิค**

เป็นส่วนหนึ่งของโครงการบูรณาการฐานข้อมูลเราได้เริ่มปรับปรุงความสามารถของอินเตอร์เฟสของ SCOP และเชื่อมโยงฐานข้อมูลแบบไดนามิก ขั้นตอนเริ่มต้นที่แนะนำโดย MSD คือการใช้เซิร์ฟเวอร์ตามความต้องการของคำจำกัดความของโดเมน SCOP สิ่งนี้มีวัตถุประสงค์เพื่อหลีกเลี่ยงปัญหาการซิงโครไนซ์ที่เกิดขึ้นจากกำหนดการวางจำหน่ายที่แตกต่างกันของฐานข้อมูลต่าง ๆ มันใช้เทคโนโลยี Simple Object Access Protocol (SOAP) และปัจจุบัน Pfam ใช้งานอยู่ทีมเพื่อแสดงการเปรียบเทียบโดเมนใน CATH, Pfam และฐานข้อมูล SCOP คาดว่าจะมีการพัฒนาต่อไปและจะมีการเปิดเผยต่อผู้มีส่วนได้เสียอื่น ๆ

**สรุปข้อสังเกต**

จากจุดเริ่มต้นจุดสนใจหลักของ SCOP คือความสัมพันธ์เชิงวิวัฒนาการที่เป็นไปได้ระหว่างโปรตีนที่ตรวจไม่พบโดยวิธีการเปรียบเทียบลำดับ ข้อมูลโครงสร้างที่จัดเรียงตามลำดับชั้นมีส่วนสำคัญในการพัฒนาวิธีการตามลำดับแบบร่วมสมัยพร้อมความไวที่ดีขึ้น วิธีการเหล่านี้อนุญาตให้มีการรวมกลุ่มของโปรตีนที่รู้จักและสมมุติจำนวนมากในฐานข้อมูลลำดับในครอบครัวลำดับโปรตีนที่ค่อนข้างน้อย ความพร้อมใช้ของลำดับจีโนมที่สมบูรณ์ช่วยให้การสำรวจของวิวัฒนาการและโครงสร้างของสิ่งมีชีวิตที่แตกต่างกันและการปรับแต่งของสายเลือดของพวกเขา ส่วนใหญ่ของตระกูลโปรตีนที่มีโครงสร้างที่ไม่รู้จักสามารถมอบหมายด้วยความมั่นใจใน SCOP superfamilies ที่มีอยู่ การสะสมของโครงสร้างอย่างต่อเนื่อง

**กิตติกรรมประกาศ**

เรารับทราบการมีส่วนร่วมของ Dr Loredana Lo Conte ในการบำรุงรักษาและพัฒนาฐานข้อมูล SCOP งานนี้ได้รับการสนับสนุนโดย MRC ให้ทุนเชิงกลยุทธ์ G0100305

**อ้างอิง**

1. Murzin, A., Brenner, SE, Hubbard, TJP และ Chothia, C ( 1995 ) SCOP: การจำแนกโครงสร้างของฐานข้อมูลโปรตีนสำหรับการตรวจสอบลำดับและโครงสร้าง *J. Mol Biol* , 247, 536 – 540
2. เบรนเนอร์, SE, Chothia, C., Hubbard, TJP และ Murzin, A. ( 1996 ) ทำความเข้าใจโครงสร้างโปรตีน: ใช้ SCOP เพื่อการตีความแบบพับได้ *วิธีการ Enzymol*  , 266 , 635 – 643
3. Bernstein, FC, Koetzle, TF, Williams, GJB, Meyer, EF, Brice, MD, Rodgers, JR, Kennard, O, Shimanouchi, T และ Tasumi, M ( 1977 ) The Protein Data Bank: ไฟล์เก็บถาวรที่ใช้คอมพิวเตอร์สำหรับโครงสร้างโมเลกุลขนาดใหญ่ J. Mol Biol , 112 , 535 -542
4. Westbrook, J., Feng, Z, Jain, S. , Bhat, TN, Thanki, N., Ravichandran, V., Gilliland, GL, Bluhm, W. , Weissig, H., Greer, DS และคณะ ( 2002 ) ธนาคารข้อมูลโปรตีน: การรวมที่เก็บถาวร กรดนิวคลีอิก , 30 , 245 -248
5. Lo Conte, L., Brenner, SE, Hubbard, TJP, Chothia, C และ Murzin, AG ( 2002 ) ฐานข้อมูล SCOP ในปี 2545: การปรับแต่งรองรับโครงสร้างฟังก์ชั่น กรดนิวคลีอิก , 30 , 264 -267
6. Chandonia, J. ‐ M., Hon, G, Walker, NS, Lo Conte, L., Koehl, P., Levitt, M. และเบรนเนอร์, SE ( 2004 ) บทสรุปของ ASTRAL ในปี 2004 กรดนิวคลีอิก , 32 , D189 -D192
7. Gough, J., Karplus, K., Hughey, R. และ Chothia, C. ( 2001 ) การกำหนด homology ให้กับลำดับจีโนมโดยใช้ห้องสมุดของแบบจำลองมาร์คอฟที่ซ่อนอยู่ซึ่งแสดงถึงโปรตีนทั้งหมดของโครงสร้างที่รู้จัก J. Mol Biol , 313 , 903 -919
8. Mulder, NJ, Apweiler, R., Attwood, TK, Bairoch, A., Barrell, D., Bateman, A, Binns, D., Biswas, M., Bradley, P., Bork, P. et al. ( 2003 ) ฐานข้อมูล InterPro, 2003 นำเสนอความครอบคลุมที่เพิ่มขึ้นและคุณสมบัติใหม่ กรดนิวคลีอิก , วันที่ 31 , 315 -318
9. Bateman, A., Coin, L., Durbin, R., Finn, RD, Hollich, V., Griffith ‐ Jones, S. , Khanna, A., Marshall, M. Moxon, S., Sonnhammer, ELL et อัล ( 2004 ) ฐานข้อมูลตระกูลโปรตีน Pfam กรดนิวคลีอิก , 32 , D138 -D141
10. Orengo, CA, Michie, AD, Jones, S. , Jones, DT, Swindells, MB และ Thornton, JM ( 1997 ) CATH: การจำแนกลำดับชั้นของโครงสร้างโดเมนโปรตีน โครงสร้าง , 5 , 1093 -1108
11. โกโลวิน, A. โอลด์, TJ, เทต JG, Velankar, S. บาร์ตัน, จีเจ, Boutselakis, H. Dimitropoulos, D. Fillon เจ. ฮุสเซน, A. Ionides, JMC et al, ( 2004 ) E ‐ MSD: ความช่วยเหลือแบบบูรณาการสำหรับชีวสารสนเทศศาสตร์ กรดนิวคลีอิก , 32 , D211 -D214
12. Iwata, S., Lee, JW, Okada, K., Lee, JK, Iwata, M., Rasmussen, B., ลิงก์, TA, Ramaswamy, S. และ Jap, BK ( 1998 ) โครงสร้างที่สมบูรณ์ของ 11 "subunit bovine mitochondrial cytochrome bc1 complex วิทยาศาสตร์ , 281 , 64 -71
13. Jormakka, M., Tornroth, S. ., Byrne, B. และอิวาตะ, S. ( 2002 ) พื้นฐานระดับโมเลกุลของการสร้างแรงจูงใจของโปรตอน: โครงสร้างของรูปแบบดีไฮโดรจีเนส ‐ N วิทยาศาสตร์ , 295 , 1863 -1868
14. Lancaster, CR, Kroger, A., Auer, M และ Michel, H. ( 1999 ) โครงสร้างของ fumarate reductase จาก Wolinella succinogenes ที่ความละเอียด 2.2 Å ธรรมชาติ , 402 , 377 -385
15. Liljas, L., Tate, J., Lin, T., Christian, P. และจอห์นสัน, JE ( 2002 ) ความหมายเชิงวิวัฒนาการและอนุกรมวิธานของลวดลายโครงสร้างที่อนุรักษ์ไว้ระหว่าง Picornaviruses และ Picorna แมลงเช่นไวรัส โค้ง. Virol , 147 , 59 -84
16. Chandrasekar, V และจอห์นสัน, JE ( 1998 ) โครงสร้างของไวรัสริงค็อตยาสูบ: ลิงค์ในวิวัฒนาการของแคปซูลไอโซซาฮีดอลในรูปพิคานาไวรัส โครงสร้าง , 6 , 157 -171
17. van Regenmortel, MHV, Fauquet, CM, Bishop, DHL, Carstens, EB, Estes, MK, มะนาว, SM, Maniloff, J., Mayo, MA, McGeoch, ดีเจ, Pringle, CR และคณะ (บรรณาธิการ) ( 2000 ) อนุกรมวิธานไวรัส: เจ็ดรายงานของคณะกรรมการระหว่างประเทศเกี่ยวกับอนุกรมวิธานของไวรัส Academic Press, San Diego, CA.
18. Huse, M. , Chen, YG, Massague, J. และ Kuriyan, J. ( 1999 ) โครงสร้างผลึกของโดเมนไซโตพลาสซึมของตัวรับชนิด TGF in ในคอมเพล็กซ์ด้วย FKBP12 เซลล์ , 96 , 425 -436
19. Hanks, SK ( 2003 ) การวิเคราะห์จีโนมของไคเนสโปรตีนยูคาริโอต superfamily: มุมมอง จีโนมไบโอ , 4 , 111 .
20. Hoess, A., Watson, S., Siber, GR และ Liddington, R. ( 1993 ) โครงสร้างผลึกของเอนโดท็อกซินโปรตีนที่ทำให้เป็นกลางจากปูเกือกม้า Limulus ต่อต้าน factor LPS factor ที่ความละเอียด 1.5 Å EMBO J. , 12 , 3351 -3356