

C2150

Zpracování informací a vizualizace v chemii a biochemii

1. lekce (Výzkum)

Petr Kulhánek

kulhanek@chemi.muni.cz

Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta,
Masarykova univerzita, Kamenice 5, CZ-62500 Brno

➤ Výzkum

základní a aplikovaný výzkum, formulace problému, literární rešerše, experimenty a simulace, analýza dat, zveřejnění výsledků

➤ Primární a sekundární data

nezpracované (surové) data, analyzované data, reprodukovatelnost, archivování

➤ Zveřejnění výsledků

reporty, prezentace, plakátové sdělení (postery), bakalářské/diplomové/disertační práce, články

Výzkum

Výzkum

Výzkum je často popisován jako aktivní, vytrvalý a systematický proces bádání s cílem objevit, interpretovat nebo přepracovat fakta. Tento intelektuální proces produkuje velké množství teorií, zákonů, popisů chování a umožňuje jejich praktické využití.

Základní výzkum je experimentální a teoretická práce vynakládaná zásadně za účelem získání nových vědomostí o **základních principech** jevů nebo pozorovatelných skutečností, která není primárně zaměřena na uplatnění nebo využití v praxi.

Aplikovaný výzkum (někdy též cílený výzkum) je experimentální a teoretické práce k získání nových poznatků, ale zcela jednoznačně zaměřených na **specifické, konkrétní předem stanovené cíle** využití v praxi.

<https://cs.wikipedia.org/wiki/V%C3%BDzkum>

https://cs.wikipedia.org/wiki/Z%C3%A1kladn%C3%AD_v%C3%BDzkum

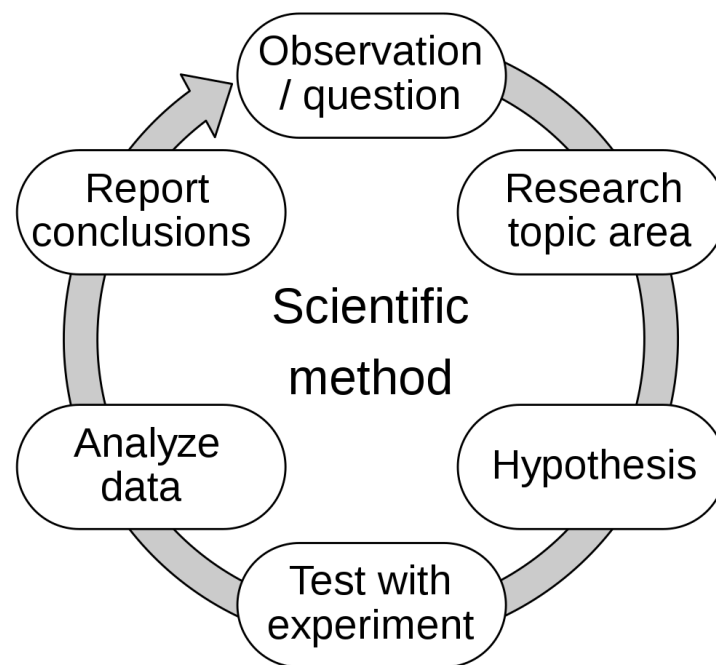
https://cs.wikipedia.org/wiki/Aplikovan%C3%BD_v%C3%BDzkum

Vědecká metoda

Vědecká metoda je posloupnost nebo sada procesů, používaných při vědeckém výzkumu. Cílem je získat znalosti a vědomosti pomocí pozorování a dedukce na základě dosud známých poznatků. Přijímání nových vědeckých poznatků je založeno na konkrétních důkazech.

Vědecká metoda je založena na předpokladu, že kritériem pravdivosti vědecké hypotézy je **souhlas předpovědí s výsledky výzkumu**. Tento přístup udržuje vědecké hypotézy v neustálém kontaktu s realitou a umožňuje jejich falzifikaci, neboť hypotéza, jejíž důsledky jsou v rozporu s výzkumnými zjištěními, bude falzifikována (tedy vyvrácena).

Mnohokrát ověřená hypotéza, kterou se zatím nepovedlo vyvrátit, se stává **vědeckou teorií**.



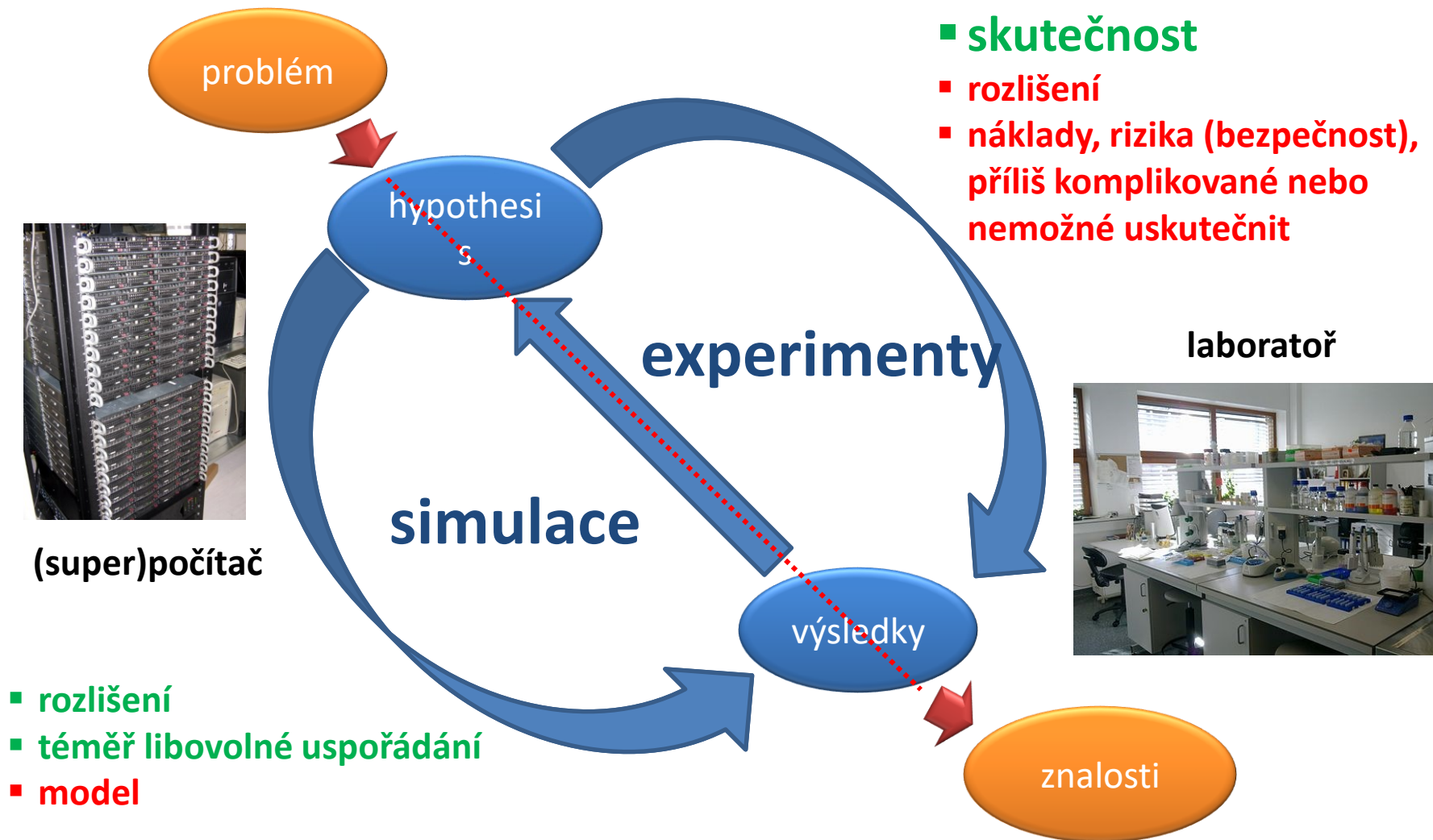
https://cs.wikipedia.org/wiki/V%C4%9Bdeck%C3%A1_metoda

Vědecká metoda

- **Observation/question (pozorování/otázka):** definice (vytvoření, formulace) předmětu (cíle) pozorování (sledování), otázky
- **Research topic area (tematická oblast výzkumu):** shromažďování informací (zdrojů, pramenů) o pozorování (sledování), otázce
→ literární rešerše (viz samostatná lekce)
- **Hypothesis (hypotéza):** vytvoření (formulace, definice) vysvětlující hypotézy (predikce)
- **Test with experiment (testování pomocí experimentu):** ověření hypotézy (predikce) provedením experimentu a sběru údajů (dat) opakovatelným (reprodukovatelným) způsobem
- **Analyze data (analýza údajů):** analýza a interpretace údajů (dat) a vyvození závěrů, která pak slouží jako východiska (vstupy, podklady) pro nové pozorování, otázky a potažmo hypotézy (predikce)
- **Report conclusions (zveřejnění výsledků):** uvedení závěrů ve všeobecnou známost (uveřejnění), obvykle v renomovaném periodiku – ve věhlasném médiu s dobrou pověstí, nejčastěji proto primárně (hlavně, prvotně) v recenzovaném časopise

https://cs.wikipedia.org/wiki/V%C4%9Bdeck%C3%A1_metoda

Experiment vs simulace



V moderním výzkumu je nutné k získání uceleného závěru použít jak experimenty tak i simulace.

Superpočítání - historie

1800 počátky děrných štítků

1946 ENIAC

1947 objev tranzistoru

1971 Intel 4004 (4 bit)

1974 Intel 8080 (8 bit)

1976 Intel 8086 (16 bit)

1985 Intel 80386 (32 bit)

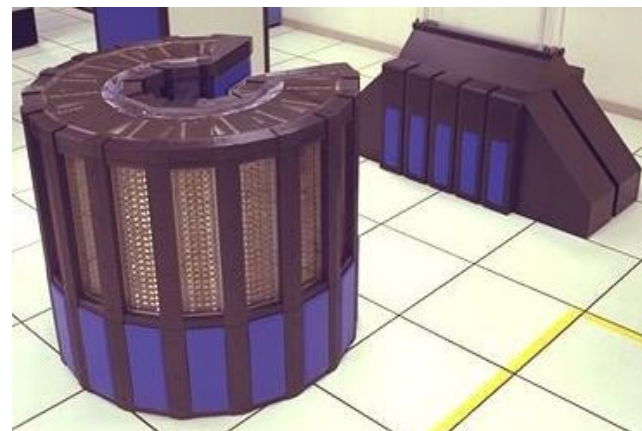
2001 IA-64 (64 bit)

2003 AMD64/EM64T (64 bit)

**2010 Intel Core i7 980X: @3,33 GHz
(6C/12T, Turbo@3,46 GHz): 109 GFlop/s**



VS



zdroj: wikipedia.org, intel.com

Přehled superpočítačů: TOP500 (<https://www.top500.org/>)

2023 Superpočítač Frontier, USA

1,6 EFlop/s (1 600 000 000 GFlop/s)

exa P T
 peta tera giga = 10^9

1985 Superpočítač Cray 2

1,9 GFlop/s

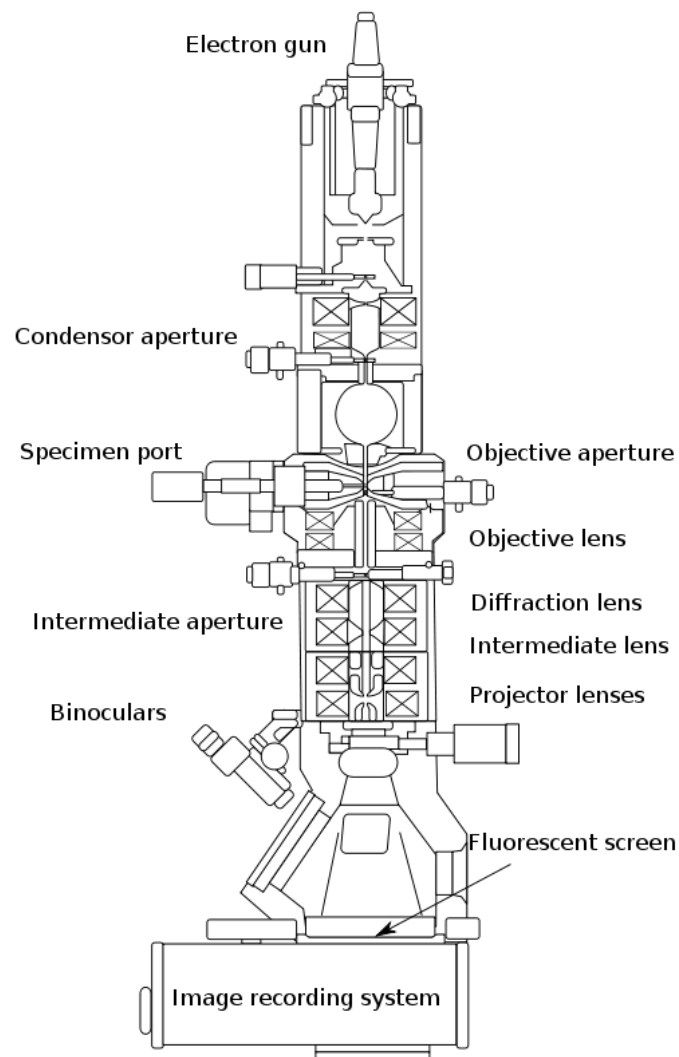
Experimenty - CryoEM

Elektronová kryogenní mikroskopie je forma transmisní elektronové mikroskopie, při níž se vzorek studuje při nízkých teplotách (obvykle při teplotě kapalného dusíku). Tato technika se používá ve **strukturní biologii**.



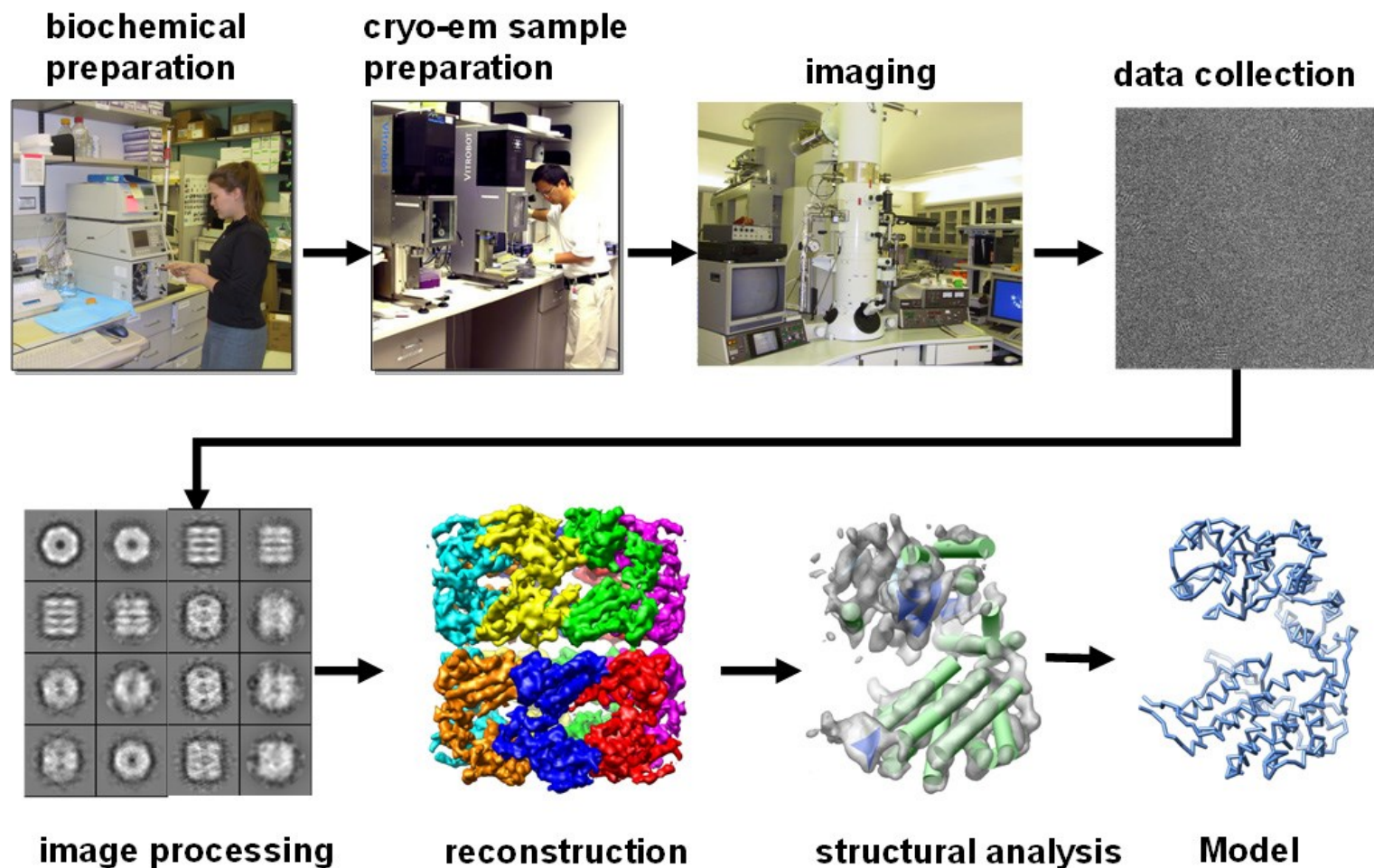
Urychlovací napětí: 300 kV

Pavilon E35/ CEITEC



Experimenty - CryoEM

Pipeline in Biological Cryo-EM



<http://proj.ncku.edu.tw/research/commentary/e/20080919/2.html>

Filozofické přístupy

Occamova břitva je princip logické úspornosti, od 19. století nazývaný podle anglického logika, františkána Williama z Ockhamu (1287–1347).

Několik různých formulací (např.):

Pokud pro nějaký jev existuje vícero vysvětlení, je lépe upřednostňovat to nejméně komplikované.

Falzifikovatelnost (častěji vyvratitelnost nebo zpochybnitelnost) je ve filosofii vědy vlastnost takového tvrzení, hypotézy nebo teorie, které je principiálně možné vyvrátit, například experimentem. (Karl R. Popper, 1934).

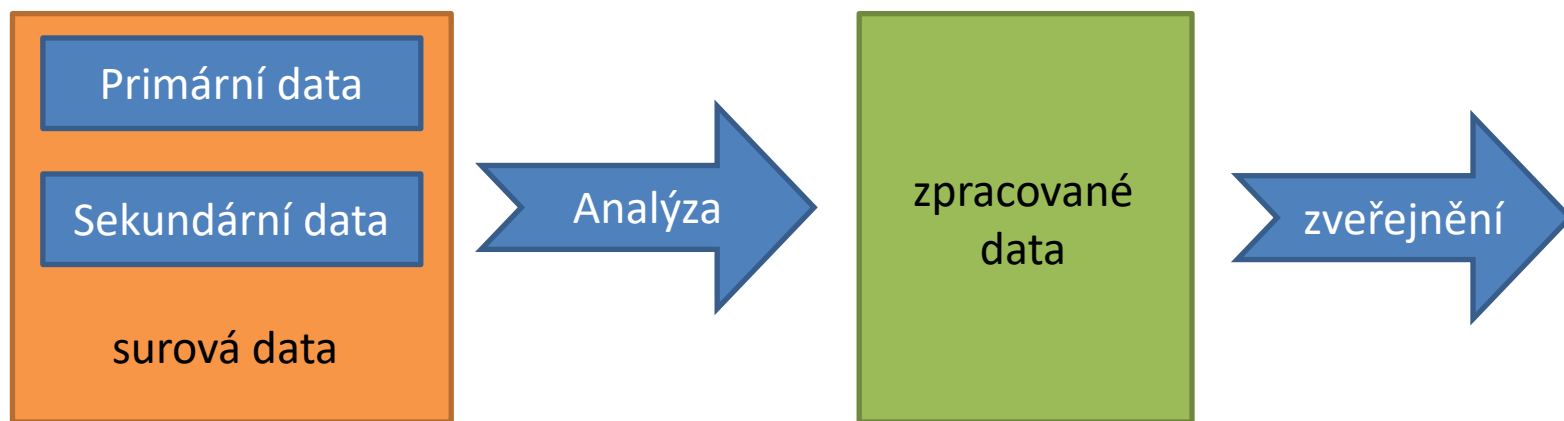
Popperova břitva: – „Vědecké teorie jsou ověřitelné. Ověřitelné teorie je možné na základě ověřovacího postupu zamítnout (a nahradit teoriemi jinými).“

https://cs.wikipedia.org/wiki/Occamova_b%C5%99itva
<https://cs.wikipedia.org/wiki/Falzifikovatelnost>

Primární a sekundární data

Data

- **Primární data** jsou data shromážděna výzkumníkem během výzkumu. Tyto data jsou výsledkem experimentů, simulací, či teoretických úvah. Návrh experimentů musí zajišťovat vysokou míru **reprodukovatelnosti** získaných dat. Je tedy nutné evidovat, kdy byly experimenty provedeny, za jakých podmínek probíhaly, kdo je realizoval, a jakým způsobem byla data zaznamenávána. Evidenci je nutné vést formou **laboratorního deníku** či obdobným způsobem.
- **Sekundární data** jsou data shromážděna buď stejným výzkumníkem v předchozím výzkumu či vytvořena jinými výzkumníky. **Užitečnost sekundárních dat je omezena** s ohledem na relevanci či přesnost, protože byly primárně určeny k jiným účelům. Typicky mezi sekundární data patří data z literární rešerše.



Archivace dat

Správně provedený výzkum poskytuje závěry postavené na experimentech či jiných výstupech, které jsou **zcela reprodukovatelné a to i nezávislými výzkumníky**.

Takto ideální výstup je však v současné době ne zcela dosažitelný většinou z důvodu složitosti prováděných experimentů.

Proto je nutné veškeré data archivovat tak, aby mohla být podrobena pozdější analýze/kritice v případě zpochybnění zveřejněných závěrů.

Z důvodů nárustu počtu provinění vůči správné vědecké praxi či dokonce úmyslných podvodů při zveřejňování vědeckých výsledků, je nutné **data archivovat minimálně 5 let**.*

* v současné době probíhá široká debata jak archivovat výzkumná data

K archivaci je možné použít datové úložiště CESNETu a MetaCentra.

podrobněji v C2115 Praktický úvod do superpočítání

Zveřejnění výsledků

Formy

- Reporty, souhrnné zprávy, protokoly
- Prezentace
- Plakátové sdělení (postery)
- Bakalářské/Diplomové/Disertační práce
- Články

poster

článek

prezentace

Gaussian Process and its Application in Molecular Simulations

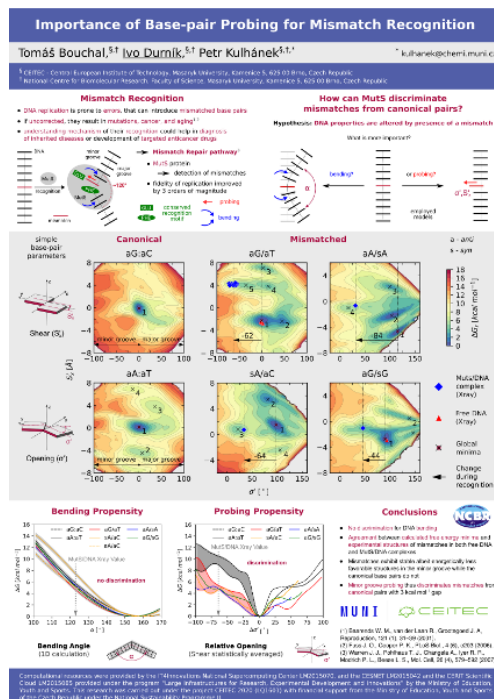
P. Kulhánek¹

¹National Centre for Biomolecular Research, Faculty of Science, Masaryk University, 625 00 Brno, Czech Republic

kulhanek@chemi.muni.cz

LCC+SB Meeting, October 21st, 2021

-1-



11222-11234 Nucleic Acids Research, 2020, Vol. 48, No. 20
doi:10.1093/nar/nkz995

Published online 20 October 2020

Importance of base-pair opening for mismatch recognition

Tomáš Bouchal^{1,2}, Ivo Durník^{1,2}, Viktor Illík², Kamila Réblová² and Petr Kulhánek^{1,2,*}

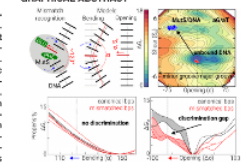
¹CEITEC - Central European Institute of Technology, Masaryk University, Kamencova 5, 625 00 Brno, Czech Republic and ²National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamencova 5, 625 00 Brno, Czech Republic

Received February 24, 2020; Revised September 09, 2020; Editorial Decision September 27, 2020; Accepted September 30, 2020

ABSTRACT

Mismatch repair is a highly conserved cellular pathway responsible for repairing mismatched dsDNA. Errors are detected by the MutS enzyme, which most likely senses altered mechanical property of damaged dsDNA rather than a specific molecular pattern. While the curved shape of dsDNA in crystallographic MutS/DNA structures suggests the role of DNA bending, the theoretical support is not fully convincing. Here, we present a computational study focused on a base-pair opening into the minor groove, a specific base-pair motion observed upon interaction with MutS. Propensities for the opening were evaluated in terms of two base-pair parameters: Opening and Shear. We tested all possible base pairs in anti/syn, anti/syn and syn/anti orientations and found clear discrimination between mismatches and canonical base-pairs only for the opening into the minor groove. Besides, the discrimination gap was also confirmed in hotspot and coldspot sequences, indicating that the opening could play a more significant role in the mismatch recognition than previously recognized. Our findings can be helpful for a better understanding of sequence-dependent mutability. Further, detailed structural characterization of mismatches can serve for designing anti-cancer drugs targeting mismatched base pairs.

GRAPHICAL ABSTRACT



INTRODUCTION

Watson-Crick (canonical) nucleobase pairs are essential for the integrity of double-stranded DNA (dsDNA) and fidelity of genetic information. The appearance of mismatched base pairs (mismatches) can result in the development of inherited genetic diseases, cancer and aging (1,2). Therefore, genetic information is continuously maintained by repair pathways that scan dsDNA and resolve inappropriate base pairing. Proteins responsible for mismatch recognition are very effective despite structural similarities between some mismatches and canonical base pairs. Understanding how such delicate discriminations is achieved on the molecular level can be beneficial for biomedical applications providing early diagnosis of inherited diseases (3,4). Further, the molecular origin of the recognition can be helpful in the development of more efficient anti-cancer drugs (5-8) targeting damaged DNA.

The mismatch repair pathway (MMR) is specialized in the repair of mismatches and short insertion/deletion loops. Experimental studies revealed that a mismatch is recognized

* To whom correspondence should be addressed. Tel: +420 549 855 499; Email: kulhanek@chemi.muni.cz

© The Author(s) 2021. Published by Oxford University Press on behalf of National Academies Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2020 Kulhánek et al. All rights reserved. This article is published with open access at <https://doi.org/10.1093/nar/nkz995>

Obecná osnova

- Název práce
- Autoři
- Afilace
- Klíčová slova
- Abstrakt
- Úvod
- Metody
- Výsledky
- Diskuze
- Závěr
- Poděkování
- Doprovodné materiály
- Citace

Pořadí sekcí a jejich obsah se liší podle formy presentování. Zde uvedený seznam je častý při zveřejňování výsledků ve formě článků.

Obecná osnova

- **Název práce** Název práce stručně vyjadřuje obsah sdělení.
- Autoři
- Afilace
- Klíčová slova Je vhodné se vyvarovat použití spojení, které evokují novost nebo unikátnost práce, např. nový (new, novel), první, atd. (novost posuzuje čtenatel)
- Abstrakt
- Úvod
- Metody
- Výsledky
- Diskuze
- Závěr
- Poděkování
- Doprovodné materiály
- Citace

Obecná osnova

- Název práce
 - **Autoři**
 - Afilace
 - Klíčová slova
 - Abstrakt
 - Úvod
 - Metody
 - Výsledky
 - Diskuze
 - Závěr
 - Poděkování
 - Doprovodné materiály
 - Citace
- Seznam autorů, kteří se podíleli na vypracování práce. Pořadí by mělo určovat zásluhu autorů na vypracování publikace.
- Korespondující autor (*) je zodpovědný za konečnou podobu publikace, dále komunikuje s redakcí a editorem časopisu.
- Vyhněte se zásluhovému autorství (honorary authorship), tj. když je vaše jméno uvedeno na aplikaci, na které jste se prakticky nepodíleli (při návrhu a realizaci experimentů, analýze a interpretaci dat, sepisování výsledků, apod.)

Obecná osnova

- Název práce
 - Autoři
 - **Afilace**
 - Klíčová slova
 - Abstrakt
 - Úvod
 - Metody
 - Výsledky
 - Diskuze
 - Závěr
 - Poděkování
 - Doprovodné materiály
 - Citace
- Přiřazení autorů k institucím.
Uvádí kontaktní informace: adresy, e-mailové a telefonní kontakty.

Obecná osnova

- Název práce
- Autoři
- Afilace
- **Klíčová slova**
- Abstrakt
- Úvod
- Metody
- Výsledky
- Diskuze
- Závěr
- Poděkování
- Doprovodné materiály
- Citace

Vhodně volená klíčová slova zlepšují indexaci v bibliografických databázích a pozdější vyhledávání publikovaných výsledků.

Obecná osnova

- Název práce
- Autoři
- Afilace
- Klíčová slova
- **Abstrakt**
- Úvod
- Metody
- Výsledky
- Diskuze
- Závěr
- Poděkování
- Doprovodné materiály
- Citace

Abstrakt stručně shrnuje studovaný problém a dosažené výsledky. Abstrakt je publikován a indexován v bibliografických databázích.

U prezentací a plakátových sdělení je abstrakt publikován v konferenčním sborníku (Book of Abstracts, Proceedings). V ojedinělých situacích může být publikován a indexován v bibliografických databázích.

Obecná osnova

- Název práce
 - Autoři
 - Afilace
 - Klíčová slova
 - Abstrakt
 - **Úvod**
 - Metody
 - Výsledky
 - Diskuze
 - Závěr
 - Poděkování
 - Doprovodné materiály
 - Citace
- Úvod do studované problematiky. Souhrn již publikovaných výsledků.

Obecná osnova

- Název práce
 - Autoři
 - Afilace
 - Klíčová slova
 - Abstrakt
 - Úvod
 - **Metody**
 - Výsledky
 - Diskuze
 - Závěr
 - Poděkování
 - Doprovodné materiály
 - Citace
- Popis použitých metod.
- Popis musí zajistit jednoznačnou reprodukovatelnost prezentovaných experimentů, simulací, či analýz.
- V případě, že je popis příliš dlouhý, může být v publikaci uvedena stručná forma, která je rozvedena v doprovodných materiálech (Supporting Information, Supplementary Materials).

Obecná osnova

- Název práce
 - Autoři
 - Afilace
 - Klíčová slova
 - Abstrakt
 - Úvod
 - Metody
 - **Výsledky**
 - **Diskuze**
 - Závěr
 - Poděkování
 - Doprovodné materiály
 - Citace
- Souhrn dosažených výsledků.
Interpretace a diskuze získaných dat.
- Může být uvedeno v samostatných kapitolách.
- } popř. **Výsledky a diskuze**

Obecná osnova

- Název práce Stručné shrnutí výsledků.
- Autoři
- Afilace
- Klíčová slova
- Abstrakt
- Úvod
- Metody
- Výsledky
- Diskuze
- **Závěr**
- Poděkování
- Doprovodné materiály
- Citace

Obecná osnova

- Název práce
 - Autoři
 - Afilace
 - Klíčová slova
 - Abstrakt
 - Úvod
 - Metody
 - Výsledky
 - Diskuze
 - Závěr
 - **Poděkování**
 - Doprovodné materiály
 - Citace
- Poděkování pracovníkům, kteří se podíleli na diskuzi k výsledkům, ale ne významnou rolí.
- Poděkování za poskytnuté financování či jiné služby (integrované laboratoře, superpočítačové centra, apod.)

Obecná osnova

- Název práce
 - Autoři
 - Afilace
 - Klíčová slova
 - Abstrakt
 - Úvod
 - Metody
 - Výsledky
 - Diskuze
 - Závěr
 - Poděkování
 - **Doprovodné materiály**
 - Citace
- Stručný popis doprovodných materiálů (Supporting Information, Supplementary Materials) obsahující informace, které nejsou klíčové pro pochopení prezentovaných dat v hlavním textu publikace.
- Jedná se například o úplná spektra (NMR), termogramy (ITC), absolutní energie (modelování), geometrie (modelování), apod.

Obecná osnova

- Název práce
 - Autoři
 - Afilace
 - Klíčová slova
 - Abstrakt
 - Úvod
 - Metody
 - Výsledky
 - Diskuze
 - Závěr
 - Poděkování
 - Doprovodné materiály
 - **Citace**
- Použitá literatura.

Jak psát abstrakt/souhrn?

Annotated example taken from *Nature* 435, 114–118 (5 May 2005).

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

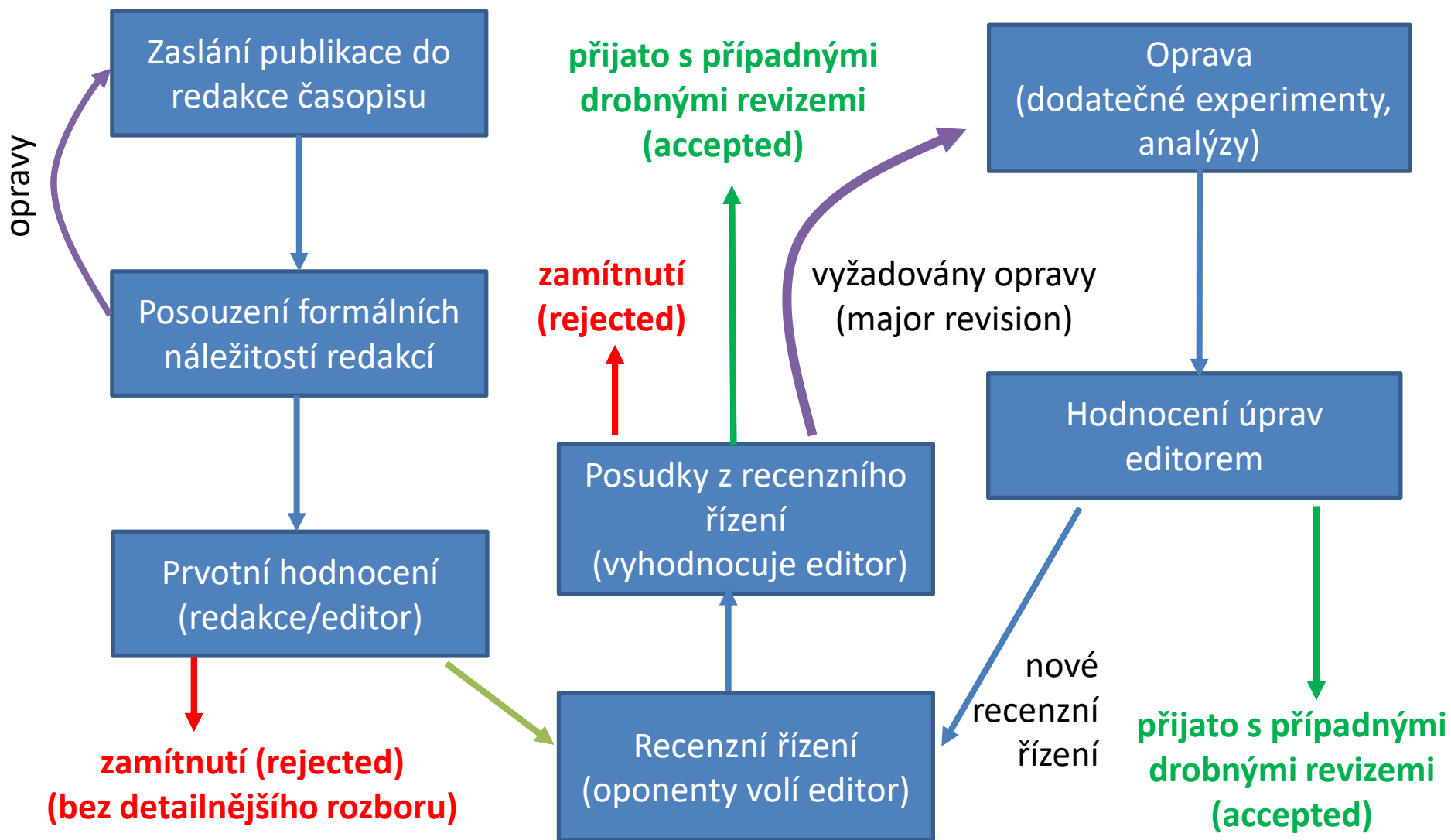
One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline, may be included in the first paragraph if the editor considers that the accessibility of the paper is significantly enhanced by their inclusion. Under these circumstances, the length of the paragraph can be up to 300 words. (This example is 190 words without the final section, and 250 words with it).

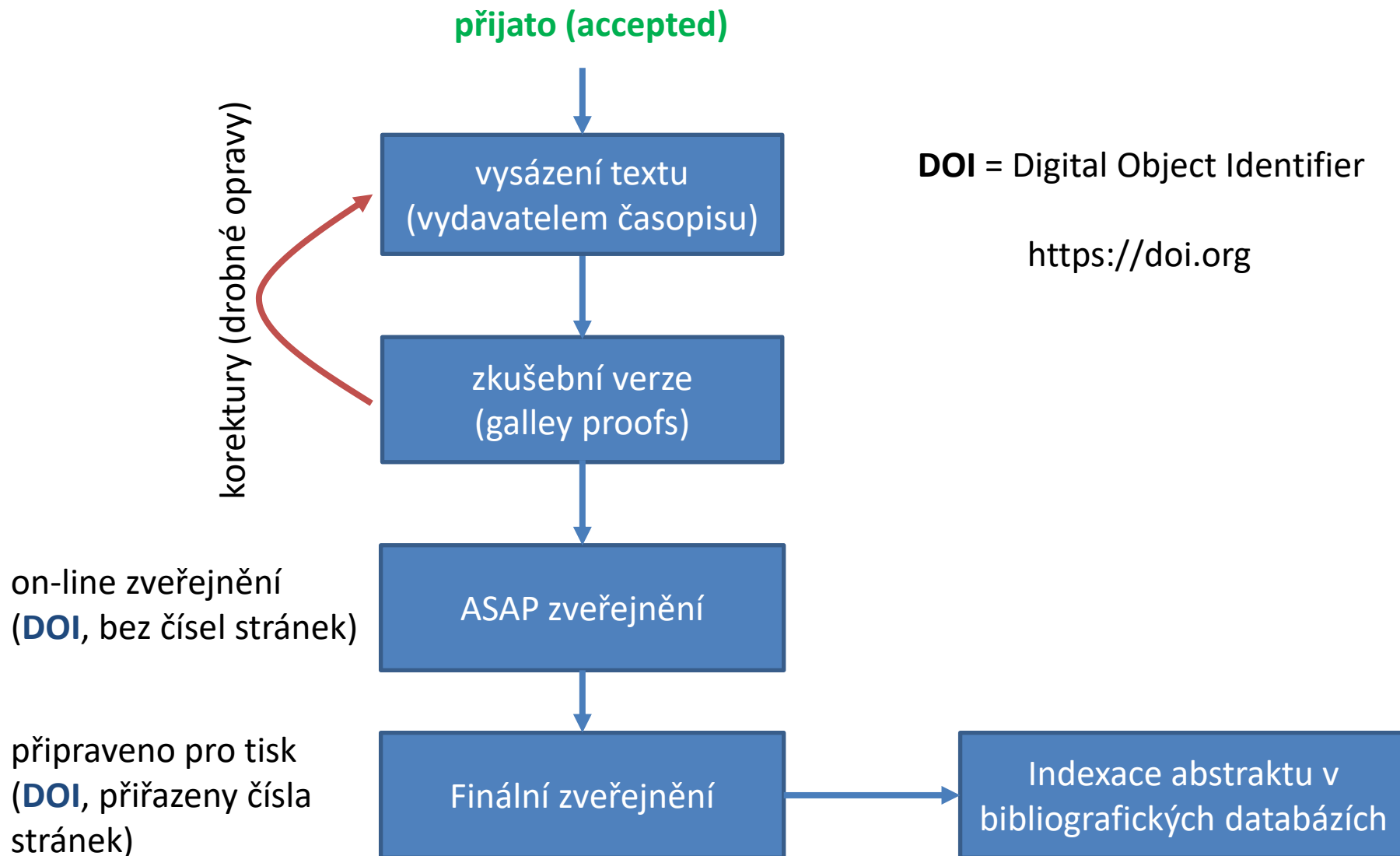
During cell division, mitotic spindles are assembled by microtubule-based motor proteins^{1,2}. The bipolar organization of spindles is essential for proper segregation of chromosomes, and requires plus-end-directed homotetrameric motor proteins of the widely conserved kinesin-5 (BimC) family³. Hypotheses for bipolar spindle formation include the ‘push–pull mitotic muscle’ model, in which kinesin-5 and opposing motor proteins act between overlapping microtubules^{2,4,5}. However, the precise roles of kinesin-5 during this process are unknown. Here we show that the vertebrate kinesin-5 Eg5 drives the sliding of microtubules depending on their relative orientation. We found in controlled *in vitro* assays that Eg5 has the remarkable capability of simultaneously moving at $\sim 20 \text{ nm s}^{-1}$ towards the plus-ends of each of the two microtubules it crosslinks. For anti-parallel microtubules, this results in relative sliding at $\sim 40 \text{ nm s}^{-1}$, comparable to spindle pole separation rates *in vivo*⁶. Furthermore, we found that Eg5 can tether microtubule plus-ends, suggesting an additional microtubule-binding mode for Eg5. Our results demonstrate how members of the kinesin-5 family are likely to function in mitosis, pushing apart interpolar microtubules as well as recruiting microtubules into bundles that are subsequently polarized by relative sliding. We anticipate our assay to be a starting point for more sophisticated *in vitro* models of mitotic spindles. For example, the individual and combined action of multiple mitotic motors could be tested, including minus-end-directed motors opposing Eg5 motility. Furthermore, Eg5 inhibition is a major target of anti-cancer drug development, and a well-defined and quantitative assay for motor function will be relevant for such developments.

<https://www.nature.com/documents/nature-summary-paragraph.pdf>

Publikační proces



Publikační proces, pokr.



Závěr

- Výzkum vyžaduje dobrou organizaci práce, preciznost, a trpělivost.
- Výzkum je typicky iterativní proces. Experimenty a simulace je nutné opakovat pro ověření jejich robustnosti a reprodukovatelnosti.
- Zveřejňování vědeckých výsledků je časově náročný proces vyžadující vysokou preciznost již při přípravě publikace.
- V kurzu C2150 Zpracování informací a vizualizace v chemii a biochemii se seznámíte s částí publikačního procesu, který zahrnuje přípravou vizuálních dat (obrázky, grafy, apod).
- Podrobnější informace k některým formám zveřejňování výsledků (bakalářské/diplomové práce) získáte od svého školitele, na oborových seminářích, či specializovaných seminářích, které pořádá Knihovna Univerzitého kampusu.

Cvičení 1

1. Jaký je autorský podíl autorů publikace DOI:10.1021/acs.jcim.1c00731?
2. Kdo je korespondenčním autorem uvedené publikace?
3. Jak dlouho trval publikační proces u publikace DOI:10.1093/nar/gkaa896?
4. Kolik příloh má uvedená publikace?
5. O čem je publikace DOI:10.1038/s41586-021-03819-2?
6. Jakou roli má redakční rada a editoři časopisu?
7. Kdo je editorem v časopise Nucleic Acid Research?
8. Který vydavatel vydává časopis Nucleic Acid Research?