

Delavnica NRRP

Primer ravnanja s podatki od načrtovanja do objave s področja naravoslovja

Matic KISOVEC, Odsek za molekularno biologijo in nanobiotehnologijo, Kemijski inštitut, Ljubljana

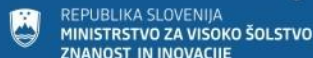
Andreja KEŽAR, Odsek za molekularno biologijo in nanobiotehnologijo, Kemijski inštitut, Ljubljana

Uroš KUNAVER, Centralna tehniška knjižnica Univerze v Ljubljani

Petra DURINI, Centralna tehniška knjižnica Univerze v Ljubljani

Miro PUŠNIK, Centralna tehniška knjižnica Univerze v Ljubljani

Ljubljana in online, Centralna tehniška knjižnica Univerze v Ljubljani,
Usposabljanje podatkovnih strokovnjakov, 18. – 21. 11. 2024





Predstavitev delavnice

- Delavnica ima za osnovo resnično raziskavo, ki je že zaključena.
- V okviru delavnice bomo najprej „razmislili“, kaj bomo raziskovali, s kakšnimi metodami in inštrumenti ter kakšne podatke in rezultate bomo dobili.
- Na podlagi tega bomo v drugem delu pripravili NRRP v skladu z zahtevami za izdelavo NRRP v okviru projektov ARIS. V okviru tega bomo tudi izbrali repozitorije in si ogledali pogoje za objavo.
- V tretjem delu bomo pa „pridobljene“ podatke pregledali, po potrebi FAIRficirali in jih pripravili za objavo v izbranih repozitorijih. Pri tem bo del podatkov objavljen v področno specifičnem repozitoriju, ki ima točno določene zahteve za objavo, del podatkov pa v splošnem repozitoriju.

Življenjski krog podatkov – kratka ponovitev

- **PONOVNA UPORABA:** Pregled in možnost uporabe obstoječih podatkov.
- **NAČRTOVANJE:** Pregled metodologij, meritev, inštrumentov... ter načrt ravnanja z raziskovalnimi podatki.
- **ZBIRANJE PODATKOV:** eksperimentiranje, meritve, beleženje, opazovanje in simulacije.
- **OBDELAVA PODATKOV:** pretvorba podatkov, priprava na analizo, preverjanje kakovosti, psevdonimizacija /anonimizacija.
- **ANALIZA:** metode, modeli, programska oprema, rezultati.
- **SHRANJEVANJE PODATKOV:** lokalno / v oblaku, varnostno kopiranje, politika dostopa
- **OBJAVA, DELJENJE PODATKOV:** izbira repozitorija, odprtost, omejitve...
- **PONOVNA UPORABA...**





Primer priprave NRRP na področju strukturne biologije



- Matic Kisovec
- Molekularni biolog, doktoriral na področju Biomedicine, Biokemija in molekularna biologija
- Kemijski inštitut, Odsek za molekularno biologijo in nanobiotehnologijo
- Skrbnik krioelektronskega mikroskopa
- Podatke, ki jih bomo pokazali na delavnici, so prispevali kolegi in kolegice iz Odseka (Marjetka, Andreja, Luka, Primož)
- Del podatkov izvira tudi iz sodelovanja z Nacionalnim inštitutom za biologijo

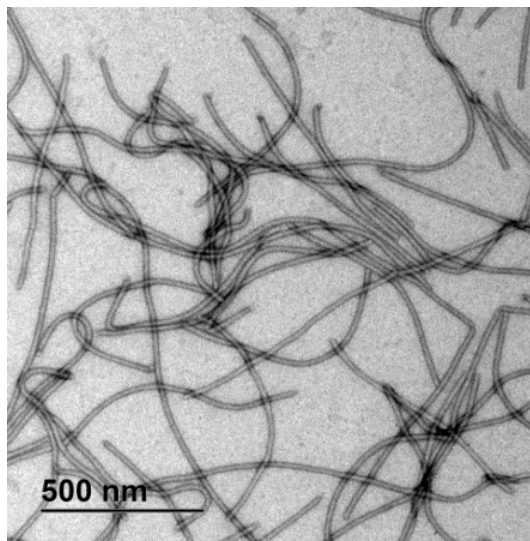


Strukturne raziskave krompirjevega virusa Y

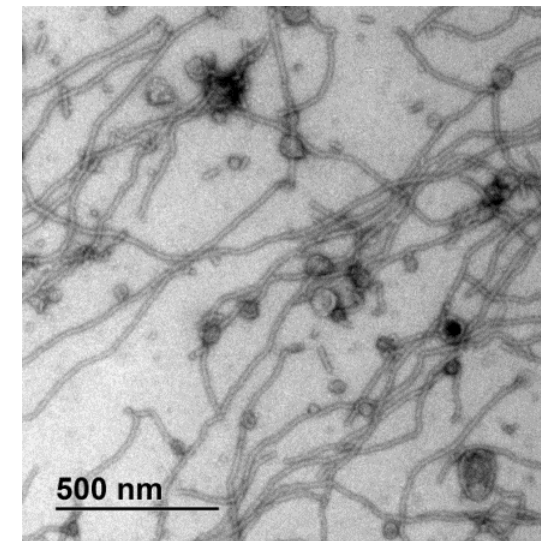


Foto: arhiv NIB

Virus (PVY)

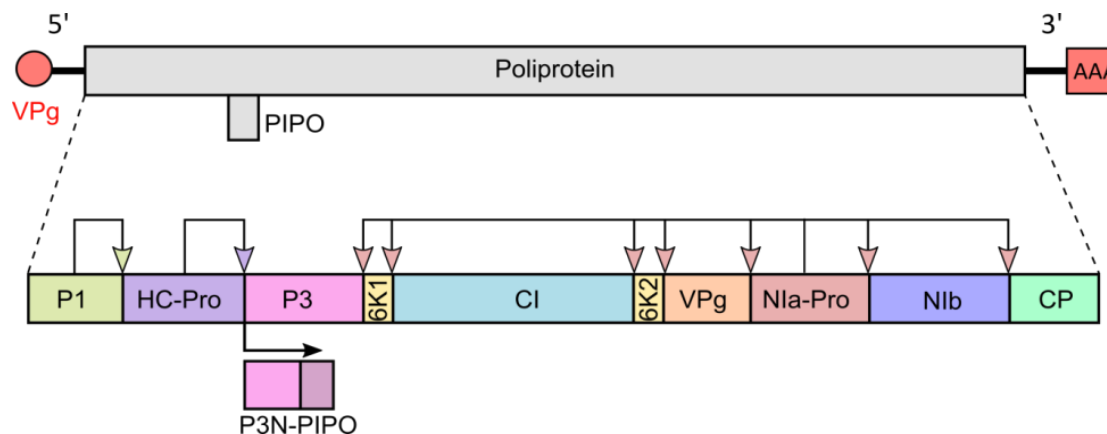


Virusom podobni delci (VLP)



Virusna ssRNA

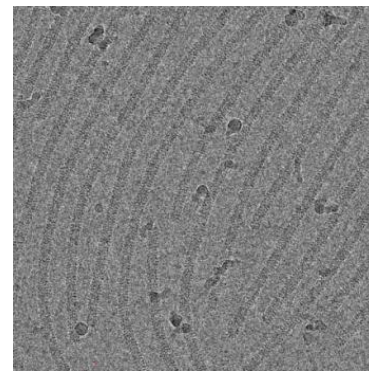
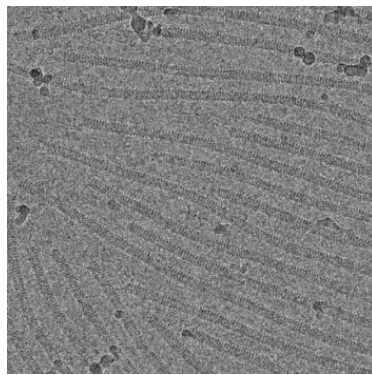
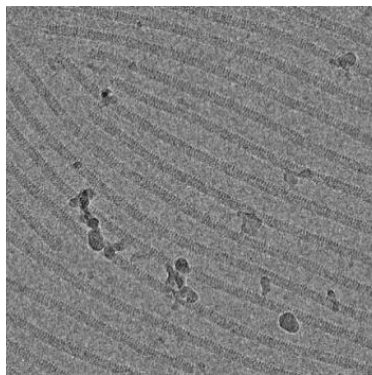
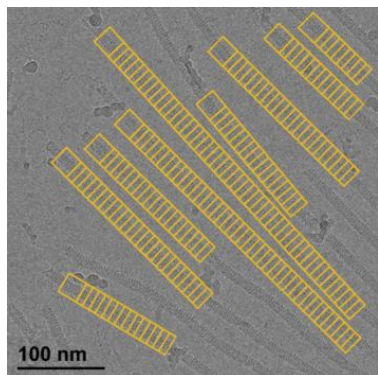
Proteini PVY



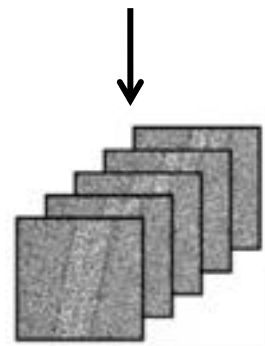


Strukturne raziskave krompirjevega virusa Y (PVY)

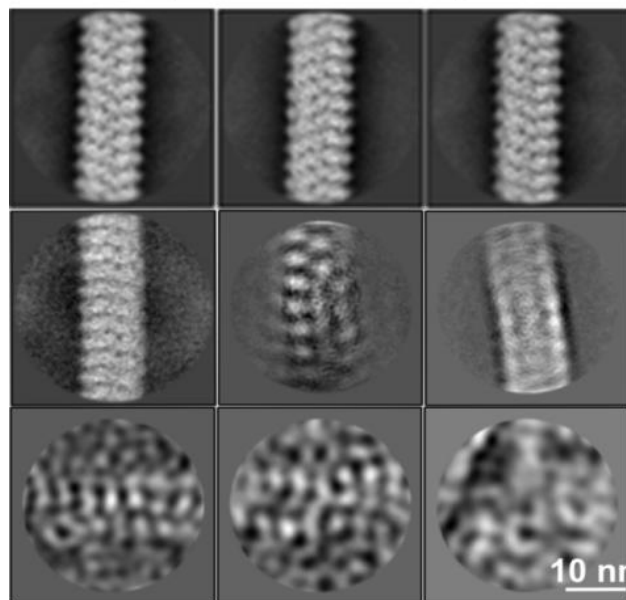
Virus: 726 posnetkov



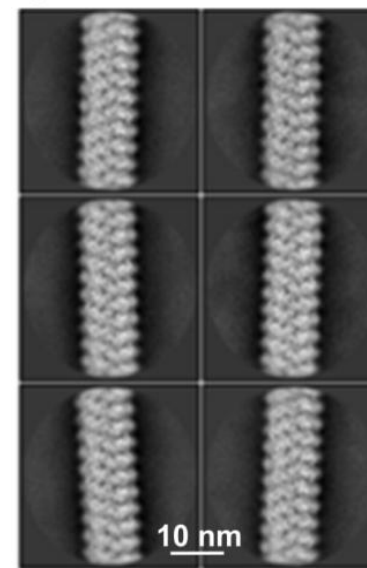
PVY – 726 posnetkov
VLP – 5344 posnetkov



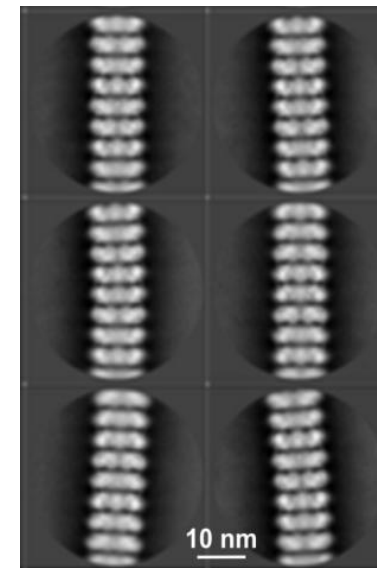
215123 segmentov



PVY

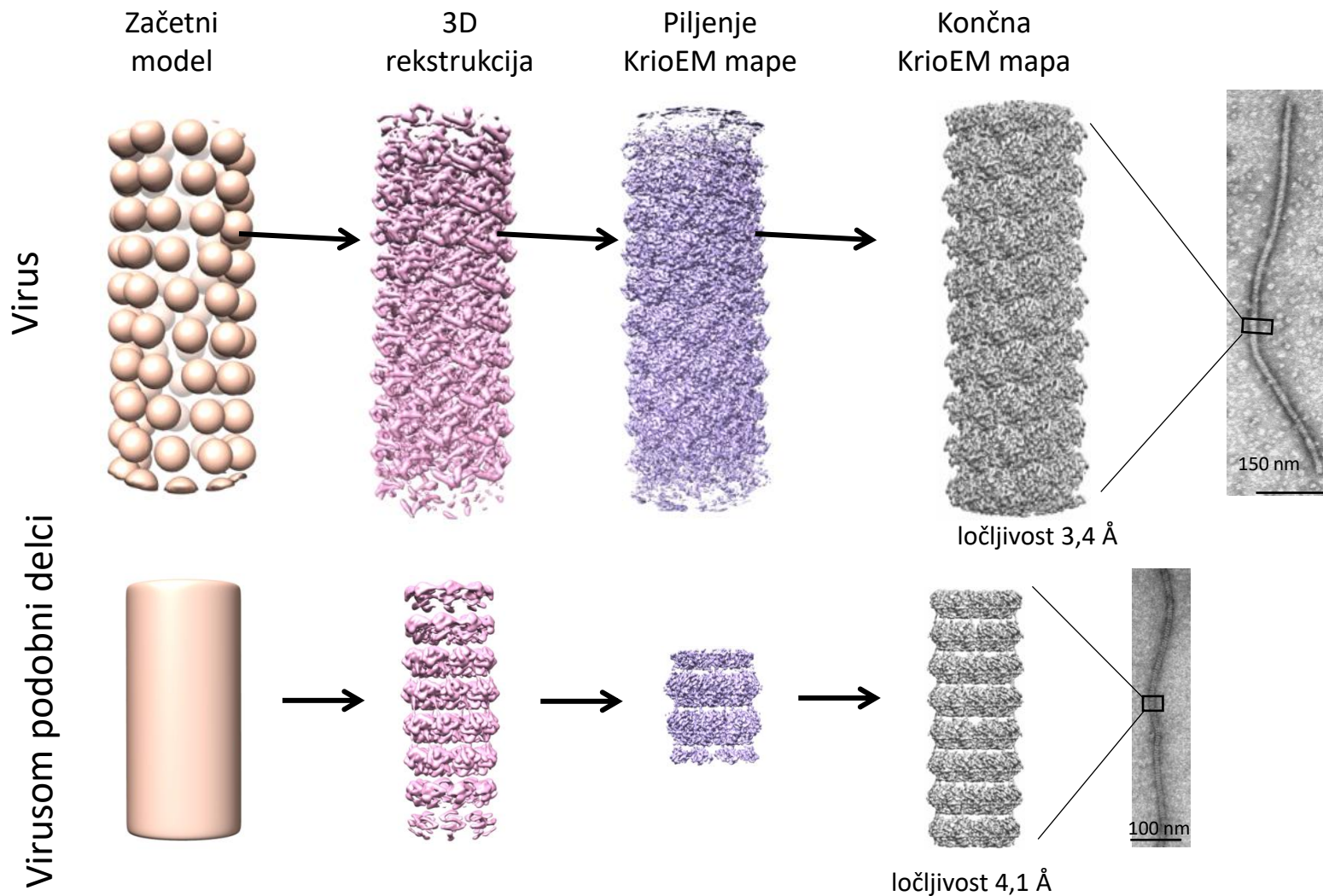


VLP





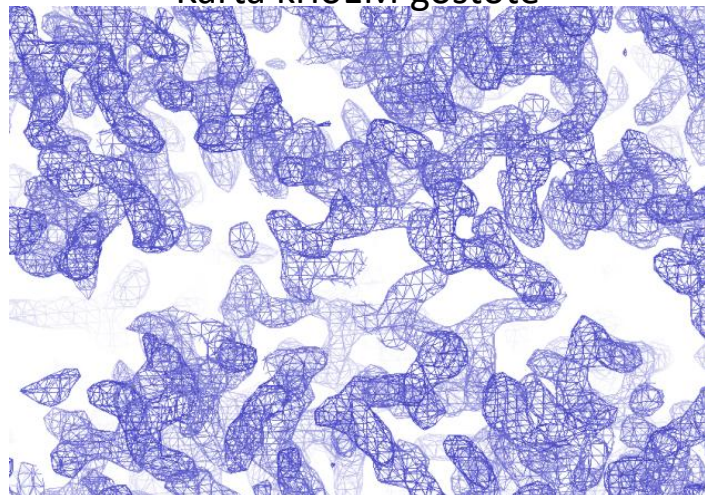
Strukturne raziskave krompirjevega virusa Y (PVY)



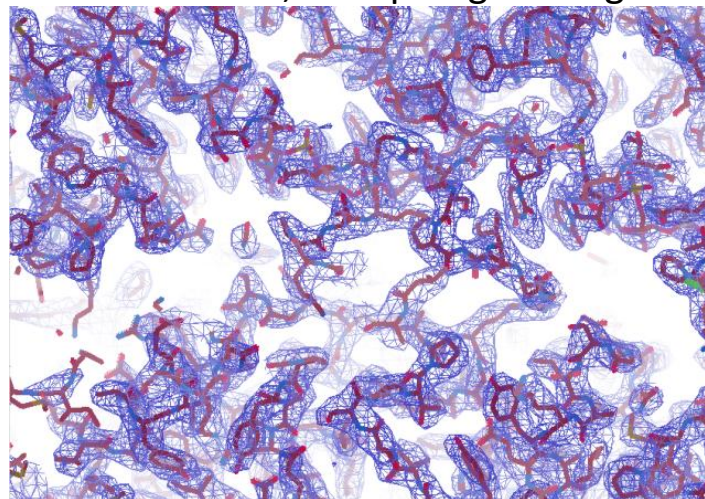


Strukturne raziskave krompirjevega virusa Y (PVY)

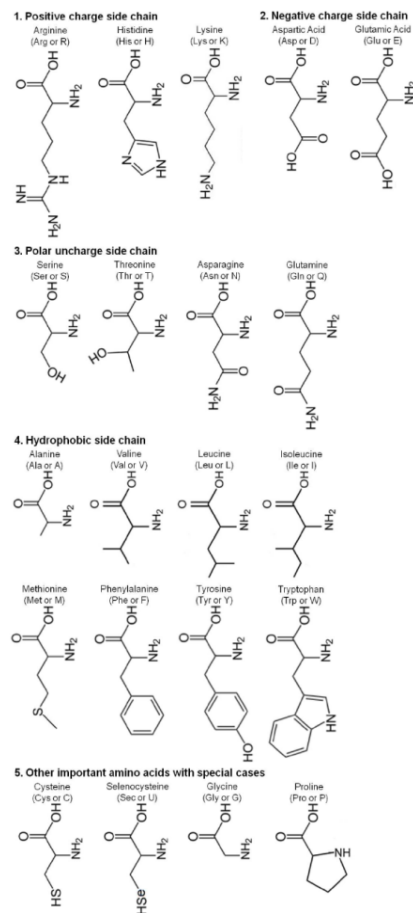
Karta krioEM gostote



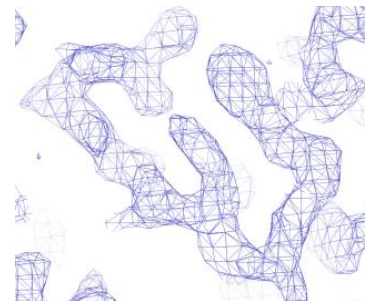
Atomski model, ki se prilega karti gostote



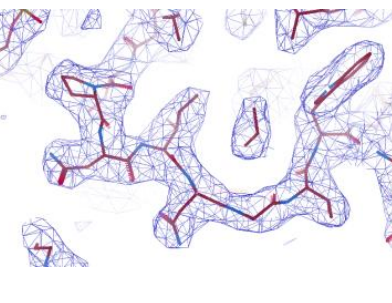
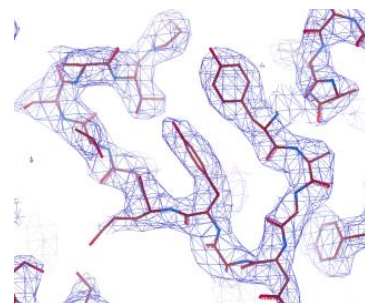
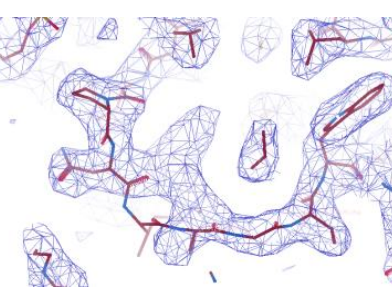
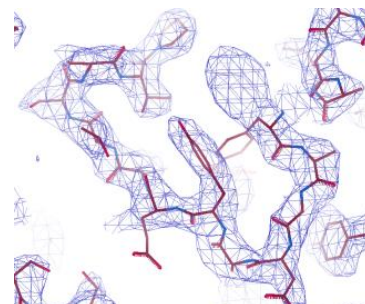
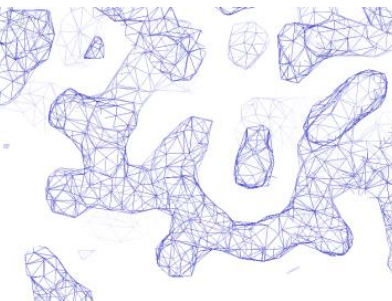
aminokisline



Primer 1

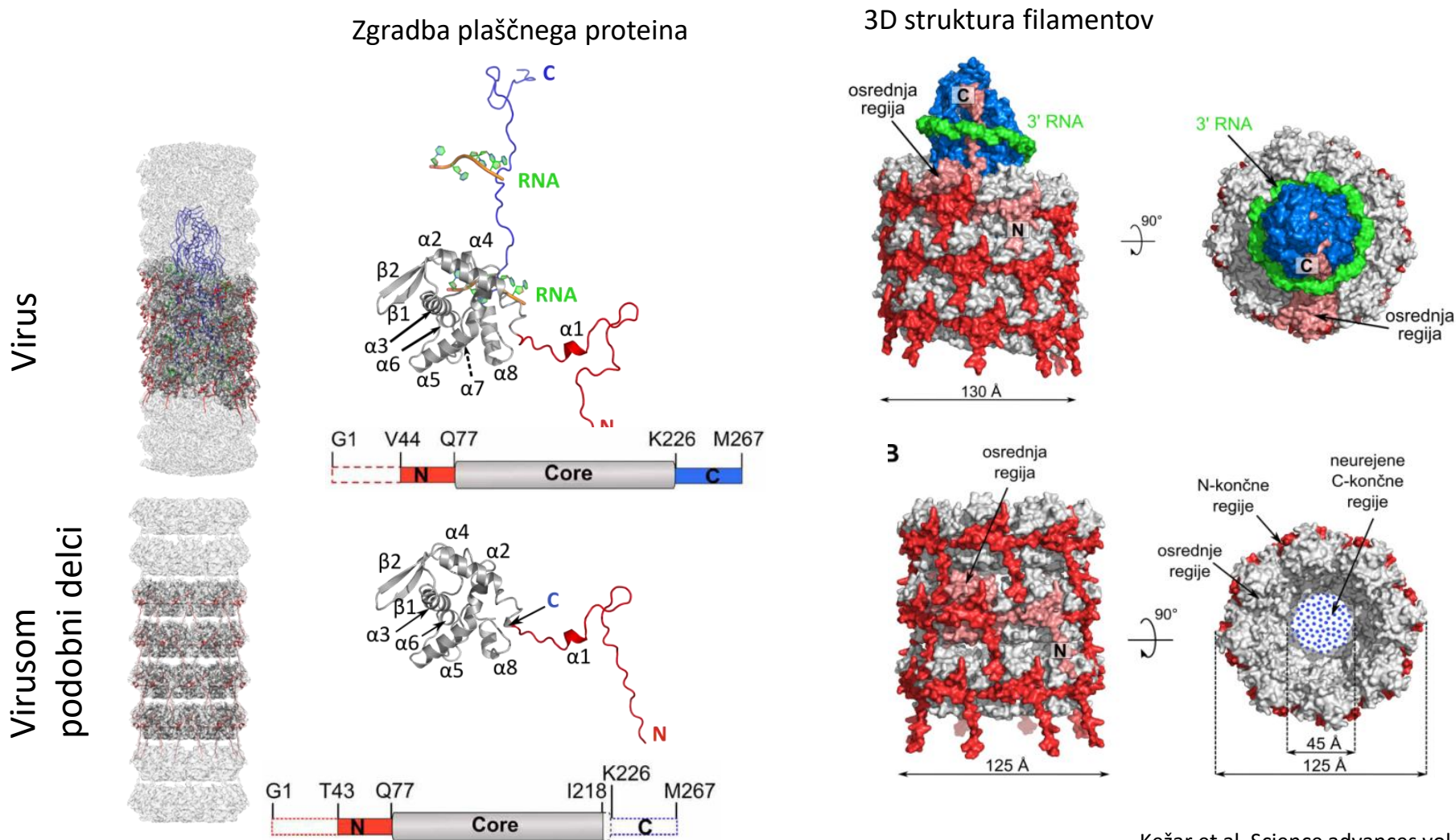


Primer 2

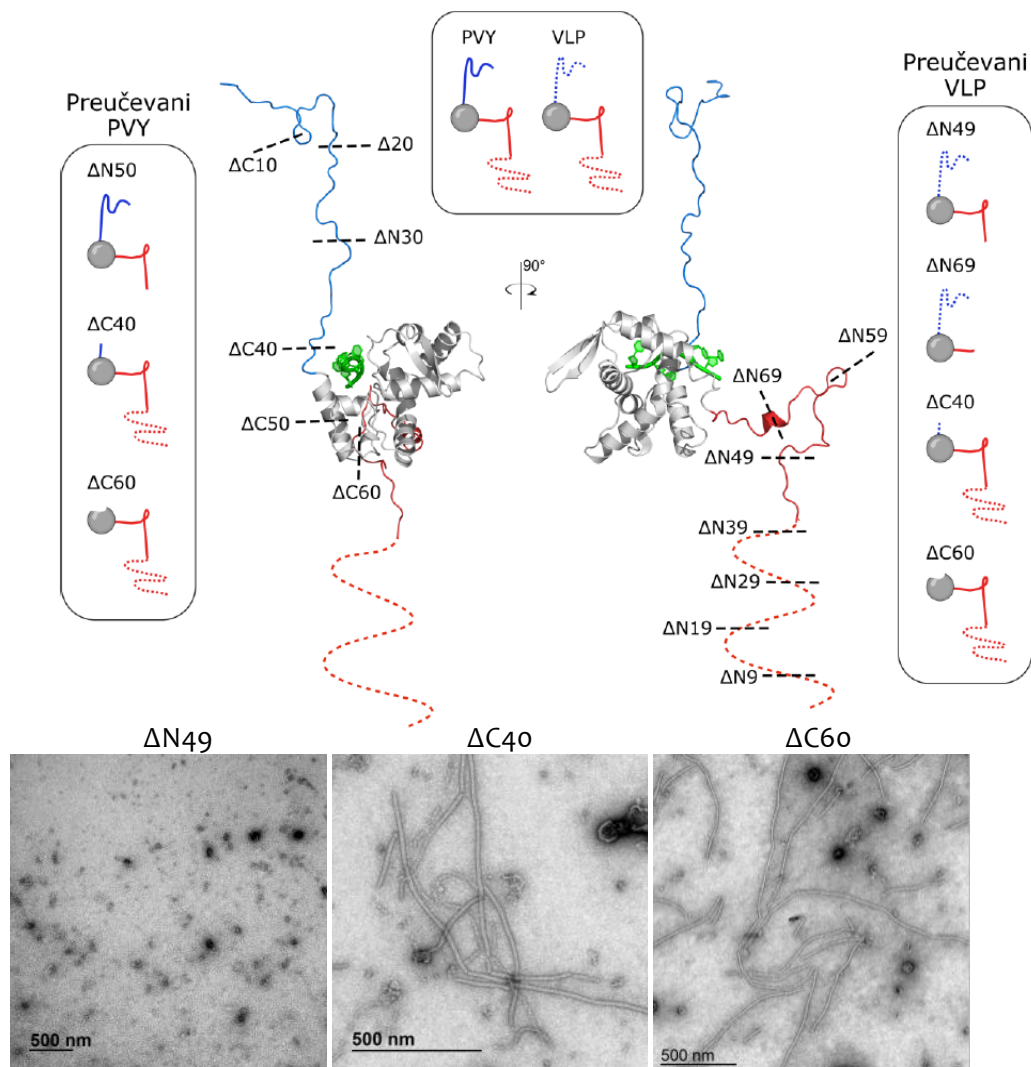




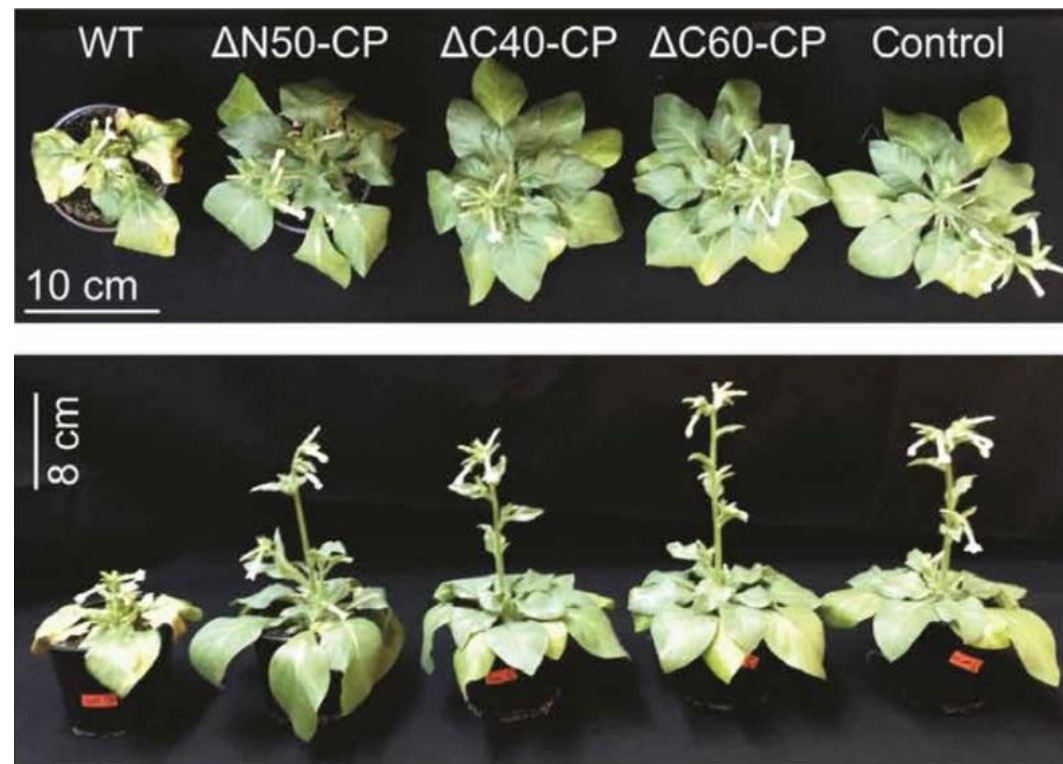
Strukturne raziskave krompirjevega virusa Y (PVY)



Strukturne raziskave krompirjevega virusa Y (PVY)



Raziskave na rastlinah



1. Ponovna uporaba obstoječih podatkov

- obstoječa zaporedja DNA (in aminokislin) – NCBI (končnica .fasta)
- obstoječi slikovni surovi podatki (slikovni – velikost več TB) - ni javno objavljeno (končnica .mrc)
- obstoječi strukturni podatki – PDB, EMDB (format .pdb ali .cif)

**Kvaliteta
podatkov?**

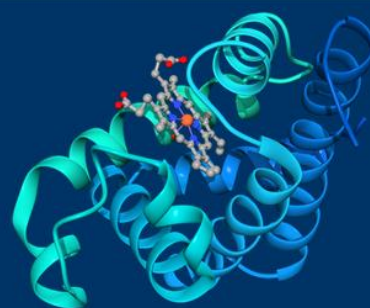


Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

Celebrating 50 Years of the PDB

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

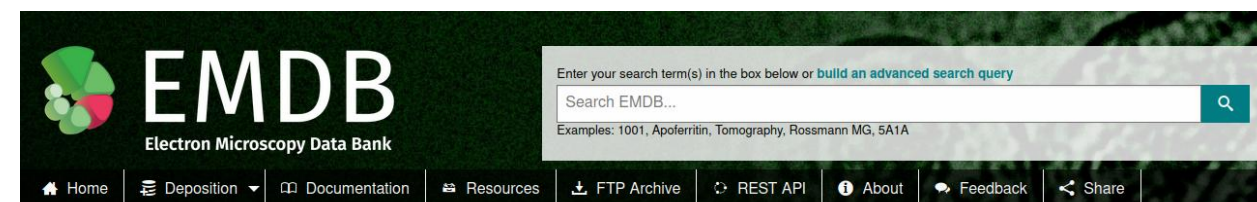
Celebrating 20 Years of the wwPDB Partnership



- Validate Structure**
or View validation reports
- Deposit Structure**
All Deposition Resources
- Download Archive**
Instructions

Začetek PDB v 1971

www.wwpdb.org
www.ebi.ac.uk/emdb
www.ebi.ac.uk/empiar
www.ncbi.nlm.nih.gov/nuccore

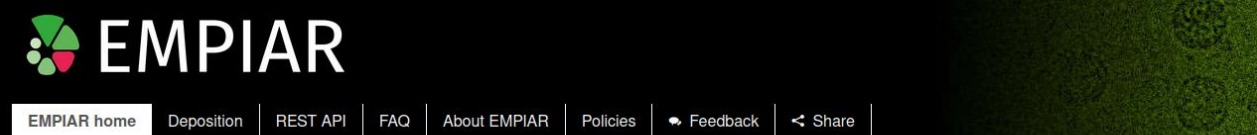


The Electron Microscopy Data Bank (EMDB) is a public repository for cryogenic-sample Electron Microscopy (cryoEM) volumes and representative tomograms of macromolecular complexes and subcellular structures. It covers a variety of techniques, including single-particle analysis, helical reconstruction, electron tomography, subtomogram averaging, and electron crystallography. [More...](#)

As of 13 November 2024, EMDB contains 40179 entries ([latest entries](#), [trends](#)).

Quick links

- [EMDB Policies](#)
- [Talks & Tutorials](#)



EMPIAR, the Electron Microscopy Public Image Archive, is a public resource for raw images underpinning 3D cryo-EM maps and tomograms (themselves archived in EMDB). EMPIAR also accommodates 3D datasets obtained with volume EM techniques and soft and hard X-ray tomography. [More ...](#)

As of 2024-11-19, EMPIAR contains 1891 entries, taking up 4.89 PB of storage.

Browse and download EMPIAR datasets using the table below. The EMDB search system also supports searches of all EMPIAR entries (video tutorial).

License:

Quick links

- [Browse EMPIAR](#)
- [Sample-Preparation Widget](#)

1. Ponovna uporaba obstoječih podatkov

- obstoječi strukturni podatki – PDB, EMDB (format .pdb ali .cif)

O tej vrsti podatkov smo že govorili v 2. sklopu usposabljanja

2. primer: področni repozotorij

Poglejmo zapis v RCSB PDB

(<https://www.rcsb.org/>)

uporabimo iskanje „6HXX“ in kliknemo na oznako 6hxx:

Dobimo podoben, le grafično nekoliko drugačen zapis.



The screenshot displays the RCSB PDB entry for 6HXX. The top navigation bar includes tabs for Structure Summary, Structure, Annotations, Experiment, Sequence, Genome, and Versions. Below the tabs, there are buttons for Display Files, Download Files, and Data API. The main content area shows the entry ID 6HXX, the name Potato virus Y, and the PDB DOI: <https://doi.org/10.2210/pdb6HXX/pdb>. It also lists the EM Map EMD-0297 and the EMDB EMDataResource NAKB: 6HXX. The classification is listed as VIRUS, the organism as Potato virus Y, and the mutation status as No. A 3D model of the virus capsid is shown on the left side of the entry.



1.1 Povzetek in opis raziskovalnih podatkov

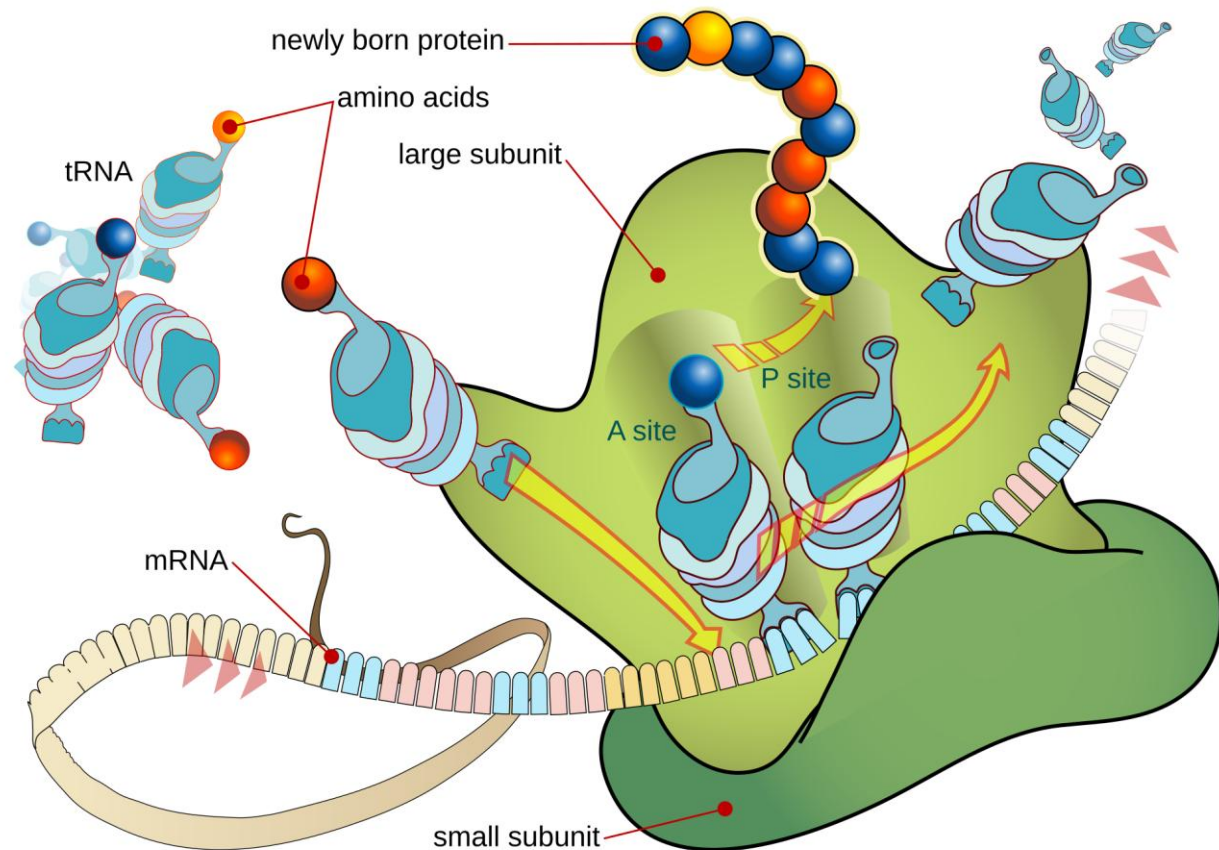
- Načrtovanje raziskave
- Priprava načrta ravnanja z raziskovalnimi podatki
- <https://nrrp.openscience.si>





1.2 Načrtovanje raziskave

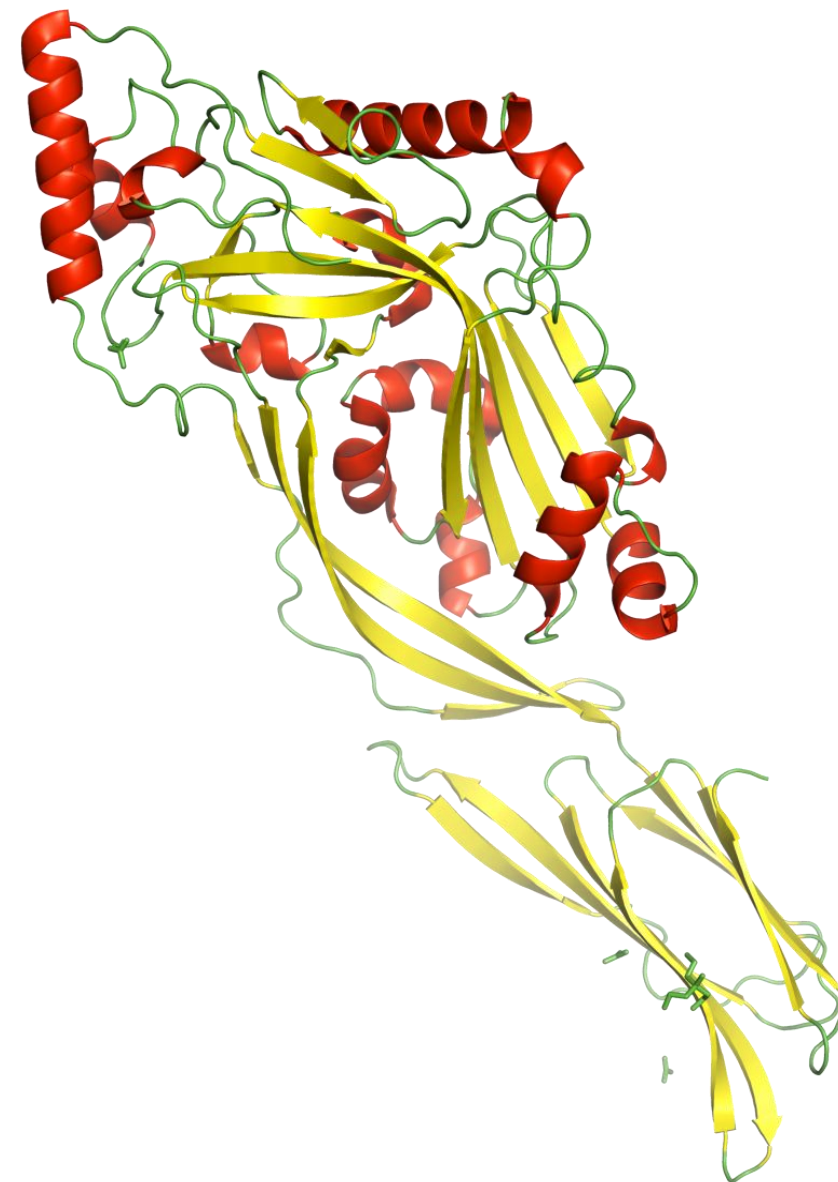
- Načrtovanje spremenjenih/mutiranih zaporedij DNK
- Eksperimenti v laboratoriju
 - Mutacije zaporedij DNK
 - Izražanje in čiščenje spremenjenih načrtovanih proteinov
 - Karakterizacija izraženih proteinov
 - kromatografija z ločevanjem po velikosti
 - diferenčna dinamična fluorimetrija
 - Karakterizacija interakcij
 - Izotermalna titracijska kalorimetrija





1.2 Načrtovanje raziskave

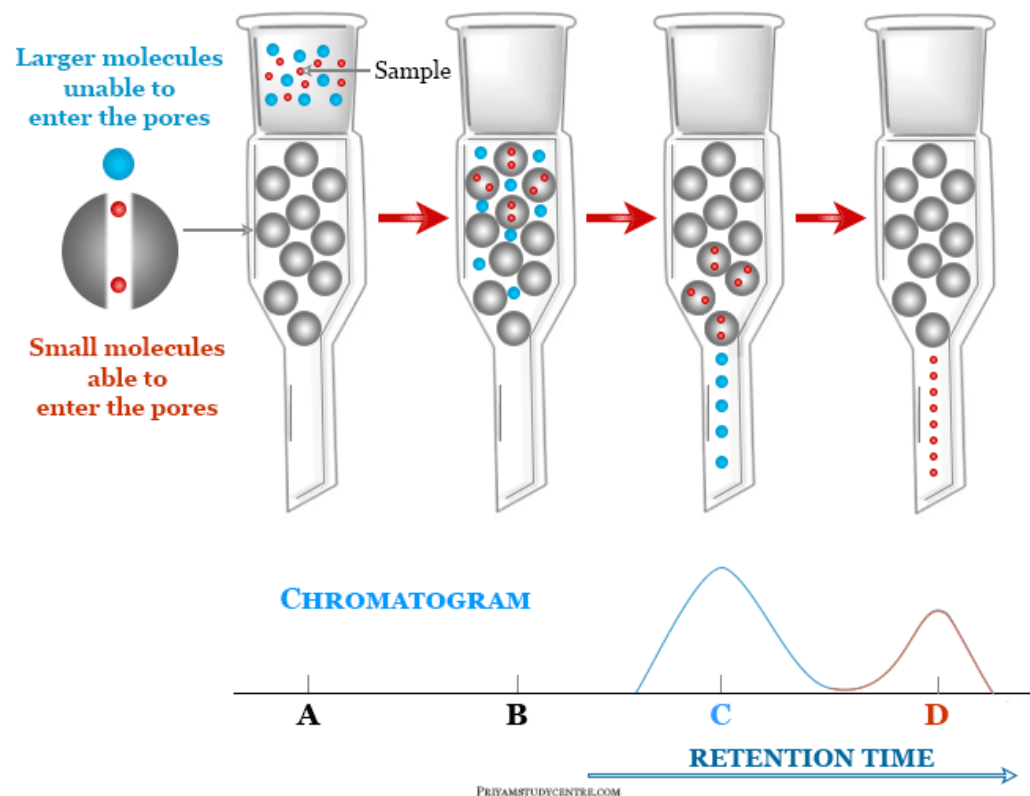
- Eksperimenti v laboratoriju (nadaljevanje)
 - Določitev 3D zgradbe spremenjenih proteinov



Kromatografija z ločevanjem po velikosti - SEC

S tekočinsko kromatografijo bomo ločili molekule po velikosti.

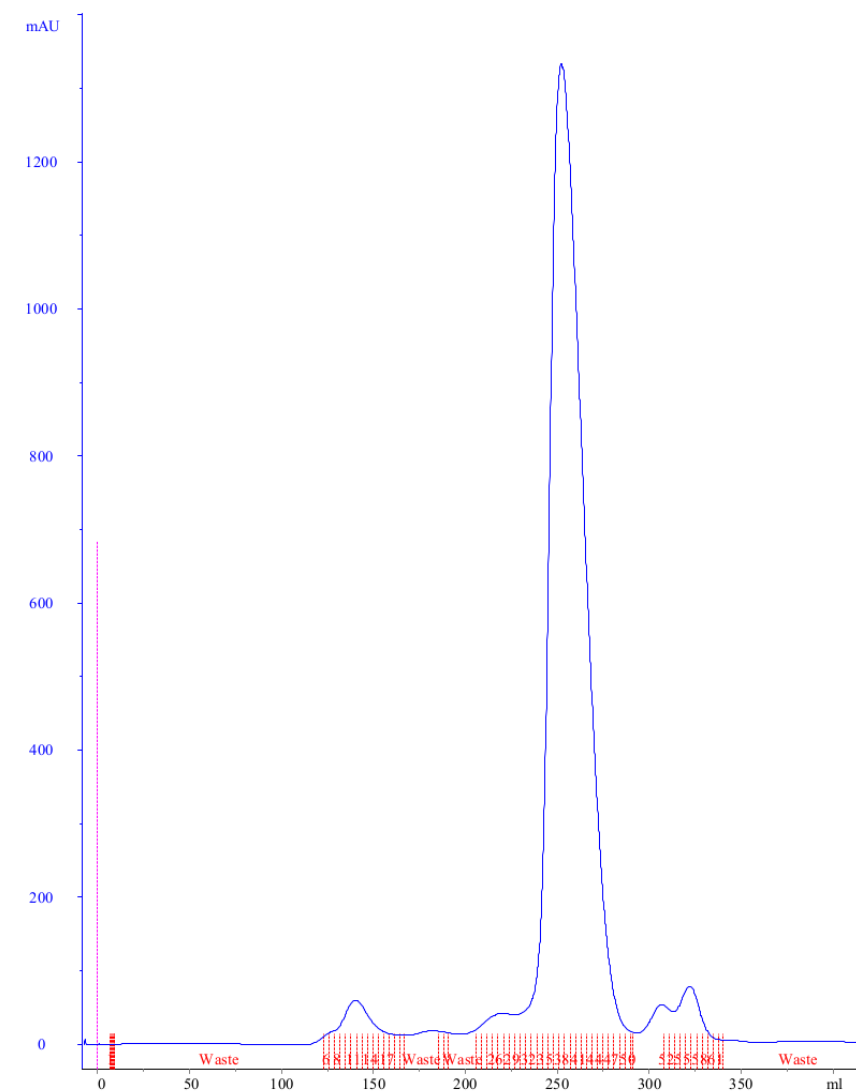
Size Exclusion Chromatography (SEC)



Kromatografija z ločevanjem po velikosti - SEC

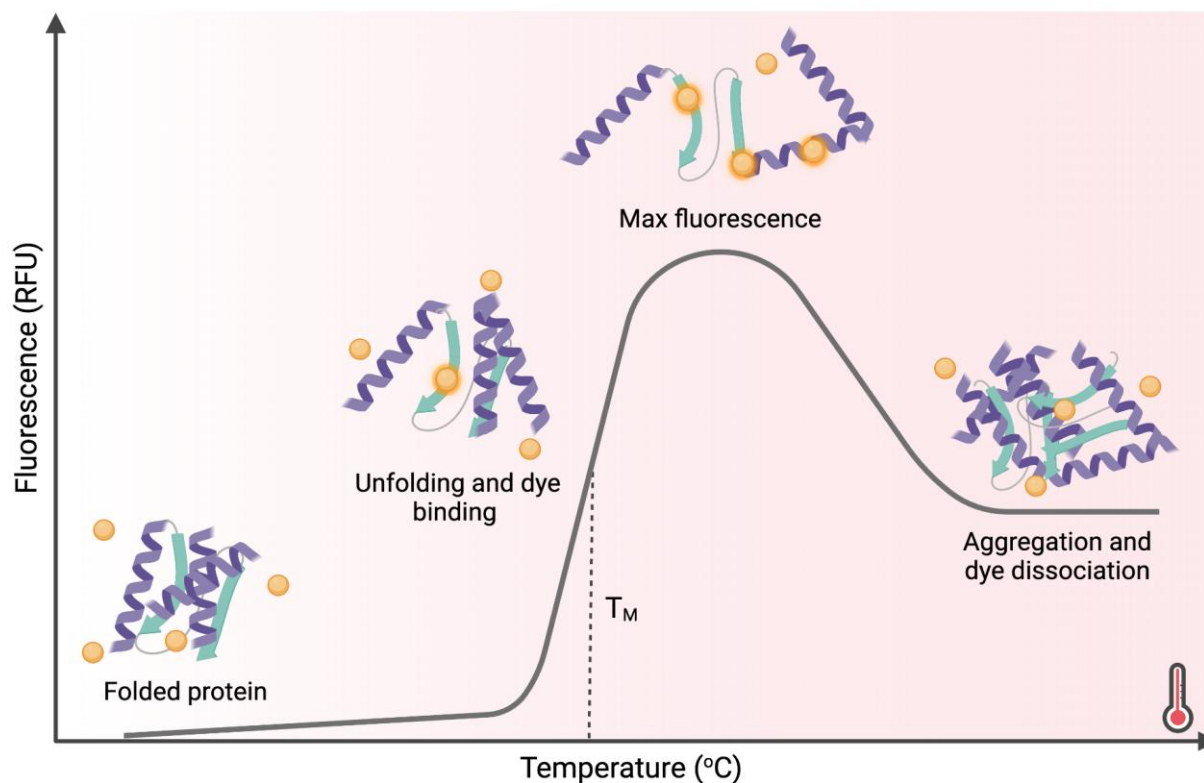
- Rezultat eksperimenta so ločene molekule in binarna datoteka .res
- V proizvajalčevi programski opremi UNICORN lahko pretvorimo rezultat v numerično obliko 'tab delimited' .csv
- Ni specifičnega repozitorija

Manual run 3.res									
4463	0000	0000	0000	0000	0000	0000	0000	0000	0000
4464	0000	0000	0000	0000	0000	0000	0000	0000	0000
4465	0000	0000	0000	0000	0000	0000	0000	0000	0000
4466	0000	0000	0000	0000	0000	0000	0000	0000	0000
4467	0000	0000	0000	0000	0000	0000	0000	0000	0000
4468	0000	0000	0000	0000	0000	0000	0000	0000	0000
4469	0000	0000	0000	0000	0000	0000	0000	0000	0000
4470	0000	0000	0000	0000	0000	0000	0000	0000	0000
4471	0000	0000	0000	0000	0000	0000	0000	0000	0000
4472	0000	0000	0000	0000	0000	0000	0000	0000	0000
4473	0000	0000	0000	0000	0000	0000	0000	0000	0000
4474	0000	0000	0000	0000	0000	0000	0000	0000	0000
4475	0000	0000	0000	0000	0000	0000	0000	0000	0000
4476	0000	0000	0000	0000	0000	0000	0000	0000	0000
4477	0000	0000	0000	0000	0000	0000	0000	0000	0000
4478	0000	0000	0000	0000	0000	0000	0000	0000	0000
4479	0000	0000	0000	0000	0000	0000	0000	0000	0000
4480	0000	0000	0000	0000	0000	0000	0000	0000	0000
4481	0000	0000	0000	0000	0000	0000	0000	0000	0000
4482	0000	0000	0000	0000	0000	0000	0000	0000	0000
4483	0000	0000	0000	0000	0000	0000	0000	0000	0000
4484	0000	0000	0000	0000	0000	0000	0000	0000	0000
4485	0000	0000	0000	0000	0000	0000	0000	0000	0000



Diferenčna dinamična fluorimetrija - DSF

Differential Scanning Fluorimetry



Preverili bomo temperaturno stabilnost proteinov.



Diferenčna dinamična fluorimetrija - DSF

- Eksperimentalni rezultat so mutante in okoljski pogoji, kjer so proteini najbolj temperaturno najbolj stabilni
- Rezultat je v obliki .ixo - vrsta xml formata, zaprt format, ki pa vsebuje nekaj metapodatkov (ime metode, datum in ura začetka in konca meritve, uporabniško ime avtorja, identifikacija naprave, datum zadnje kalibracije...)
- Krivulje v neberljivi obliki

```

1 <objectstream signature="IX0S" version="1">
2   <obj name="root" class="HTCExperiment" version="5">
3     <prop name="name">20130612 DSF A145 Tea</prop>
4     <prop name="UID">F345F79645B545A8834EB80004AC937B</prop>
5     <prop name="OwnerUID">7C7D339333FC41869C2084E33B1A7AE3</prop>
6     <prop name="Created">2013-06-12T15:18:43.578</prop>
7     <prop name="CreatedBy">7C7D339333FC41869C2084E33B1A7AE3</prop>
8     <prop name="CreatedByName">L11</prop>
9     <prop name="LastModified">2013-06-12T16:28:39.359</prop>
10    <prop name="LastModifiedBy">7C7D339333FC41869C2084E33B1A7AE3</prop>
11    <prop name="LastModifiedByName">L11</prop>

```

```

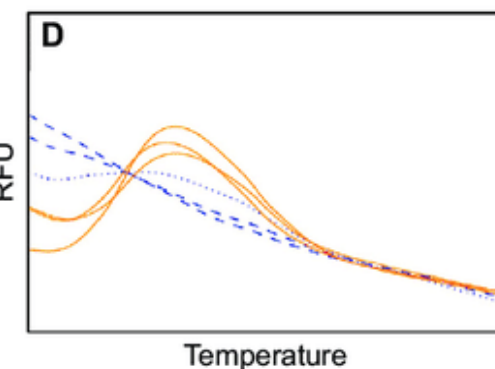
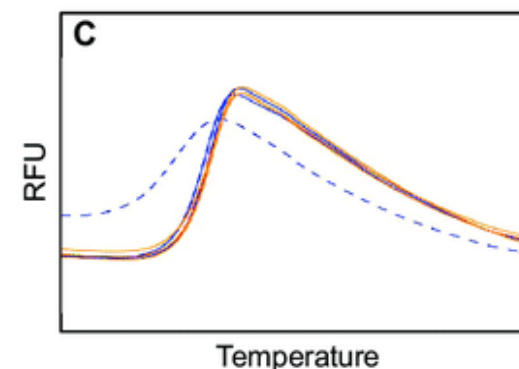
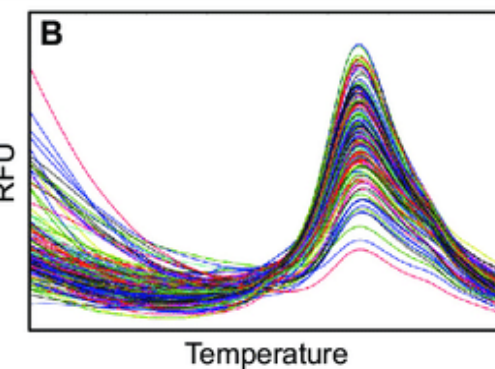
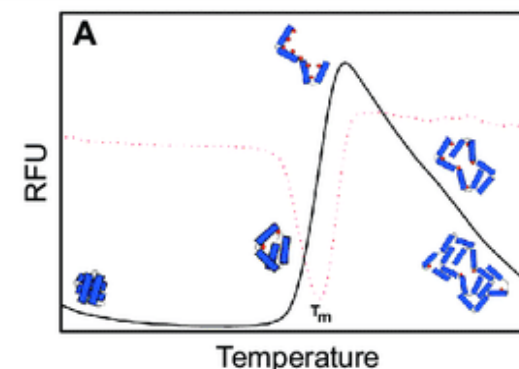
12    <list name="AppliedTemplates"
13      <item>DSFMarta-new</item>
14    </list>
15    <list name="TemplateOwners" co
16      <item>L11</item>
17    </list>
18    <prop name="SWVersion">LCS480
19    <obj name="AppProperty" class=
20      <prop name="Created">2013-
21      <prop name="LastModified">
22    </obj>
23    <obj name="ConProperty" class=
24      <prop name="Created">2013-
25      <prop name="LastModified">
26    </obj>
27    <obj name="run" class="HTCRun"
28      <prop name="name">Run</pro
29      <prop name="Created">2013-
30      <prop name="LastModified">
31      <obj name="AppProperty" cl
32        <prop name="Created">2
33        <prop name="LastModifi

```

```

        <prop name="StartTime">2013-06-12T15:34:1
        <prop name="EndTime">2013-06-12T16:20:00.
        <prop name="InstrumentID">21081</prop>
        <prop name="InstrumentVersion">HTC_VER_10
        <prop name="InstrumentName">LC_480</prop>
        <prop name="InstrCalibrationDate">1899.12
        <prop name="Technician">L11</prop>
        <prop name="AcqStoreClass">HTCAcquisition
        <prop name="AcquisitionStore">eJzcVvM48q
        pf/vN/1ddz/cdXjrb56bQ6jM6P0/0v/6n/ryZ/7P/
        wPPEaysP5H77yH55z+/98zuHuiAc2tK72z550uDrW
        IvdVCqP4iNmTeaEllobomdCLXGuRaDZSDJrzqRipZW
        5eUVRZtYhibzSU31ucUZnHYy2JzNCVv/NF0Cw47QV
        ig0Xj3csMovWqTHQ4l5rl0l7qSrKqXJUvIXDXQ5Wl
        LTcD0K4nStb33PxLVafwLMfrrX/+Mif/fiBp2fi1R
        xbsuV38kpIb8NohK3AaIi9cFP1JwGstFDI3nUuiBq
        ttrJVbN4UpsW4irQhiGo2QFTw7IKoUI8i+WKibCaKhsXZFPURFE/8FVSM1HW0x1R0bCGVYJIncLL/VeUDUFUEl33urSG8Y2G1/0hChYf1X
        ORLmfjMad8UcLOsN8JuLJ9ED+eSuZ0BDV0Q00CtfWsoqLe7uWZmEIEG2AqBYTtaajTZNXBaIWlv33EdVuGf+zozptq6cN/vjRX0nU4BcTpe
        fG7Gg5F7AUXJaRbI3n19oce3stHC3TlxxjuGofLWaaMmgC0eBkQfRMyejvtU0k+cTiFrPd3CU3le6XEUN00rZ0ougCXRqkYc0oqUuQZR2x
        0mbyNQdTw0Iajlnpqif7Uf5PqTn0VbS54PPszs8jv1iBqz0StUxDlmwQ0GgWWLq8L+0rWF0RRXruZGDRBlNrs4Si3M8D3RyDKPw2EigVvFM
        SLFV0LL61uJ4vKpxY01bhJyl/H1Ed49+GK03XEmXujx6lvYtbZd2TTnJ/7Id0KldwFBZiLXUmKqir+pKciWqwo6ypB6JOV6zoWw2iqM0p75
        VHFxIXd6dh2SVD8WZHZN9AgcyUZcKRK1GbgjHzaelv07eY0pLn4k6CV2KdD6lccM+C//5wu0PZ9sRBdoAK/ZRgdokMdcUCFWr+F5ihc/

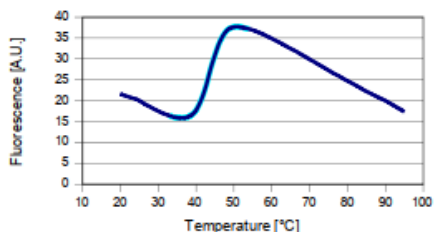
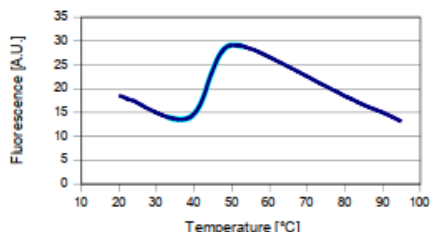
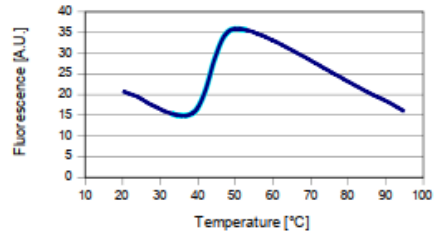
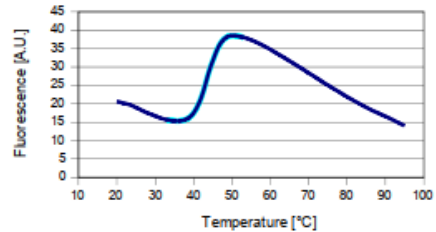
```



Diferenčna dinamična fluorimetrija - DSF

- V originalni programski opremi LC480 lahko pretvorimo rezultat v numerično obliko v obliko 'tab delimited' .csv
- Ni specifičnega repozitorija

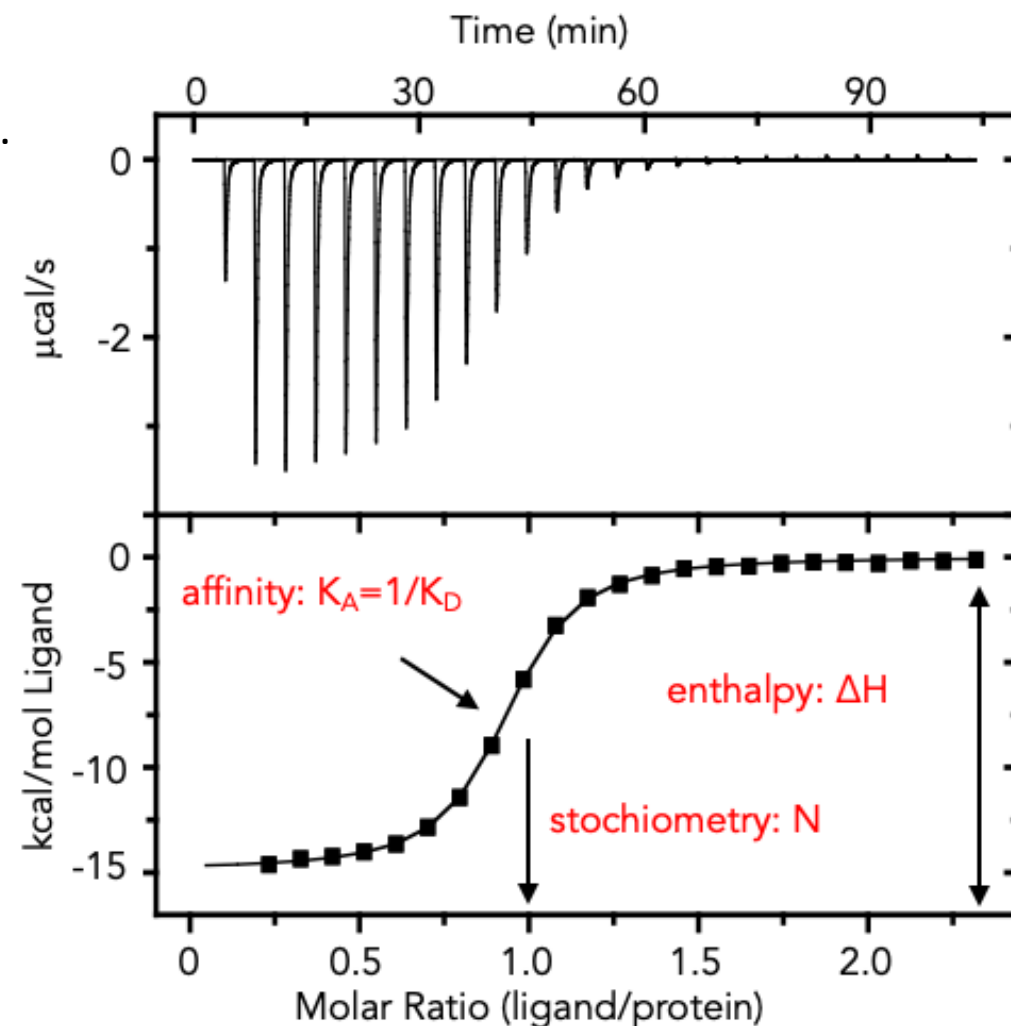
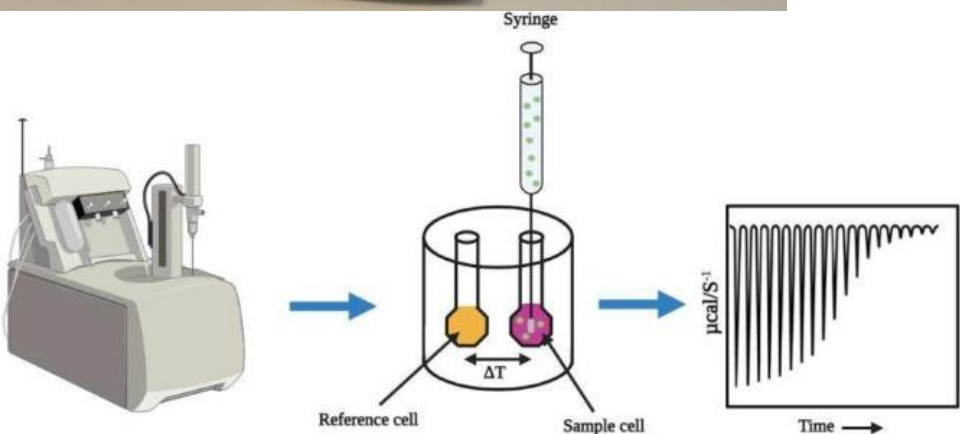
1	X	A1: 5	A2: 5	A3: 5	A4: 5	B1: 5.5	B2: 5.5	B3: 5.5	B4: 5.5	C1: 6	C2:
2	20.01	7.066084323	6.879413849	6.806880602	6.356087165	7.167008708	8.051673332				
3	20.16	7.037383548	6.850489507	6.780587469	6.339131846	7.122101581	8.014161923				
4	20.16	7.049481613	6.852626507	6.788673369	6.342631105	7.125967459	8.031745261				
5	20.16	7.063209179	6.856327855	6.802805049	6.353496319	7.151096633	8.046199941				
6	20.16	7.056652747	6.862598919	6.79826652	6.347605277	7.151097334	8.038376535				
7	20.24	7.028170679	6.852636324	6.785654495	6.335269401	7.140699983	8.017685935				
8	20.24	7.037377937	6.850484597	6.780580455	6.345018968	7.136816581	8.023531243				
9	20.31	7.031467247	6.830706889	6.785667118	6.339693576	7.109639595	8.002088772				
10	20.31	7.042300004	6.845794678	6.765476206	6.327351289	7.136827102	8.000102684				
11	20.31	7.03145252	6.851075346	6.781125834	6.335275712	7.132531142	8.00832357				
12	20.32	7.021621721	6.838543338	6.781130042	6.338223059	7.13744549	7.9989584				
13	20.39	7.020991623	6.842661051	6.780593781	6.328820615	7.128659894	8.003234908				
14	20.39	7.029825625	6.830712499	6.77055592	6.333808701	7.119458473	7.997405135				
15	20.39	7.033484061	6.83267099	6.783077107	6.348881963	7.116599515	7.991887155				
16	20.47	7.010534035	6.824846321	6.777043939	6.328260049	7.116611439	7.984088439				
17	20.47	7.014440323	6.834828387	6.776063067	6.319982216	7.117212478	8.004808735				
18	20.47	7.017086275	6.837380522	6.777033418	6.334152225	7.119876687	7.991887856				
19	20.55	7.01381261	6.831115876	6.769478105	6.326789109	7.110067617	7.988772018				
20	20.55	7.019737898	6.830529109	6.767432386	6.324752811	7.115998302	7.989924289				
21	20.63	7.005614069	6.821713647	6.772509068	6.328264959	7.123160172	7.982527013				
22	20.63	7.008897554	6.834259773	6.761918583	6.332684091	7.100254338	7.97627359				
23	20.63	7.01974351	6.828963249	6.776502729	6.330642325	7.101272763	7.977419157				
24	20.71	7.004993418	6.816430055	6.773488168	6.321807002	7.091474212	7.966494617				
25	20.71	6.996159765	6.814280624	6.762367578	6.315350699	7.094133385	7.969209239				
26	20.77	7.000072745	6.8148663	6.758363447	6.326230977	7.096382257	7.974311991				
27	20.77	7.000270223	6.825251637	6.766808128	6.325664067	7.084208168	7.970772517				

well	A01	A02
	0,1 mg/ml 406A; 2 x Sypro Orange; 5,7 in (20mM Na-fosfat, 0mM NaCl, pH=5,7)	0,1 mg/ml 406A; 2 x Sypro Orange; 5,7 in (20mM Na-fosfat, 0mM NaCl, pH=5,7)
Row A		
Min before transition (change if necessary)	36	36
Max after transition (change if necessary)	51	50
slope [%/K]	14,20	14,50
T _m (change if appropriate)	43,64591	43,82276
comment / irregularity		
well	B01	B02
	0,1 mg/ml 406A; 2 x Sypro Orange; 5,7 in (20mM Na-fosfat, 10mM NaCl, pH=5,7)	0,1 mg/ml 406A; 2 x Sypro Orange; 5,7 in (20mM Na-fosfat, 10mM NaCl, pH=5,7)
Row B		
Min before transition (change if necessary)	36	36
Max after transition (change if necessary)	50	50
slope [%/K]	15,00	14,40
T _m (change if appropriate)	43,4659	43,54906
comment / irregularity		

Izotermalna titracijska kalorimetrija - ITC



Izmerili bomo termodinamske parametre interakcij med proteini.



www.sciencedirect.com/science/article/pii/S0378517323004830

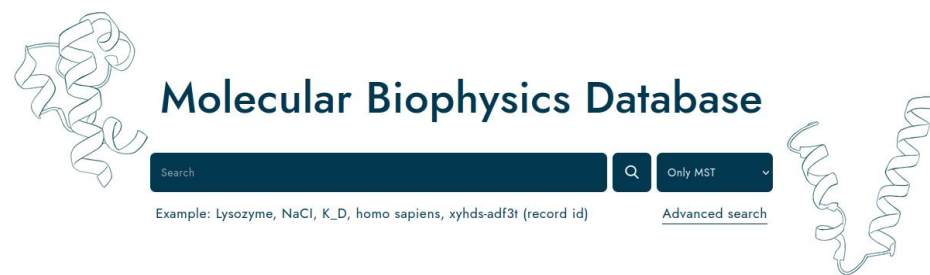
<https://cmi.hms.harvard.edu/isothermal-titration-calorimetry>

<https://spwindustrial.com/microcal-malvern-vp-itc>

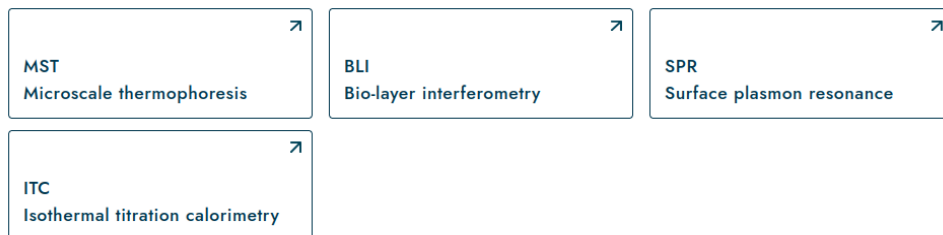
www.ki.si/odseki/d11-odsek-za-molekularno-biologijo-in-nanobiotehnologijo/oprema/oprema-aris/izotermalni-titracijski-kalorimeter-vp-itc

Izotermalna titracijska kalorimetrija - ITC

- Dobimo podatke o jakosti vezave, stehiometrije in spremembi entalpije
- Rezultat je v obliki .itc - numerična oblika podatkov, zaprt format, ki pa vsebuje nekaj neanotiranih metapodatkov
V bistvu gre za csv format.
- Krivulje v berljivi obliki
- Obstaja alternativna programska oprema za analizo, npr. NITPIC.
- Ni (še) specifičnega repozitorija.
- V beta fazi testiranja je Molecular Biophysics Database, kjer bo mogoča depozicija ITC podatkov.
- Predvidoma bodo avtomatsko prebrani tudi metapodatki v datotekah .itc.



Currently supported techniques



Trenutno podatkovna zbirka 'Molecular Biophysics Database' še ni javno dostopna.

```

1  $ITC
2  $ 38
3  $NOT
4  $ 25
5  $ 900
6  $ 307
7  $ 5
8  $ 2
9  $ADCGainCode: 1
10 $False,True,True
11 $ 2 , 4 , 300 , 2
12 $ 2 , 4 , 300 , 2
13 $ 8 , 16 , 300 , 2
14 $ 8 , 16 , 300 , 2
15 $ 8 , 16 , 300 , 2
16 $ 8 , 16 , 300 , 2
17 $ 8 , 16 , 300 , 2
18 $ 8 , 16 , 300 , 2
19 $ 8 , 16 , 300 , 2
20 $ 8 , 16 , 300 , 2
21 $ 8 , 16 , 300 , 2
22 $ 8 , 16 , 300 , 2
23 $ 8 , 16 , 300 , 2
24 $ 8 , 16 , 300 , 2
25 $ 8 , 16 , 300 , 2
26 $ 8 , 16 , 300 , 2

```

```

$ 8 , 16 , 300 , 2
$ 8 , 16 , 300 , 2
# 0
# 0.4
# 0.0125
# 1.4366
# 25
# 28.7662
# 28.7662
?
% VPITC09.05.709
% 1.4366
% 28.7662
%-0.774283 , 100.0328
%-0.117156 , 97.591924
% 0.2232
% 121.71 , 121.7
% 1 , 80
% 0.018102 , 0.054272
% 0 , 8
% 0 , 18
% 101600 319645
% 101.8
% 260 , 300 , 450 , 550
% 0.079109607877 , -0.002430676307 , -4.151377187071E-05 , 1.379618135893E-06 , -9.482611389214E-09
% VPViewer2000 Ver: 1.4.12
@0
2.00,4.646363,25.01637
4.00,4.645543,25.01758
6.00,4.645385,25.01633
8.00,4.642471,25.01677

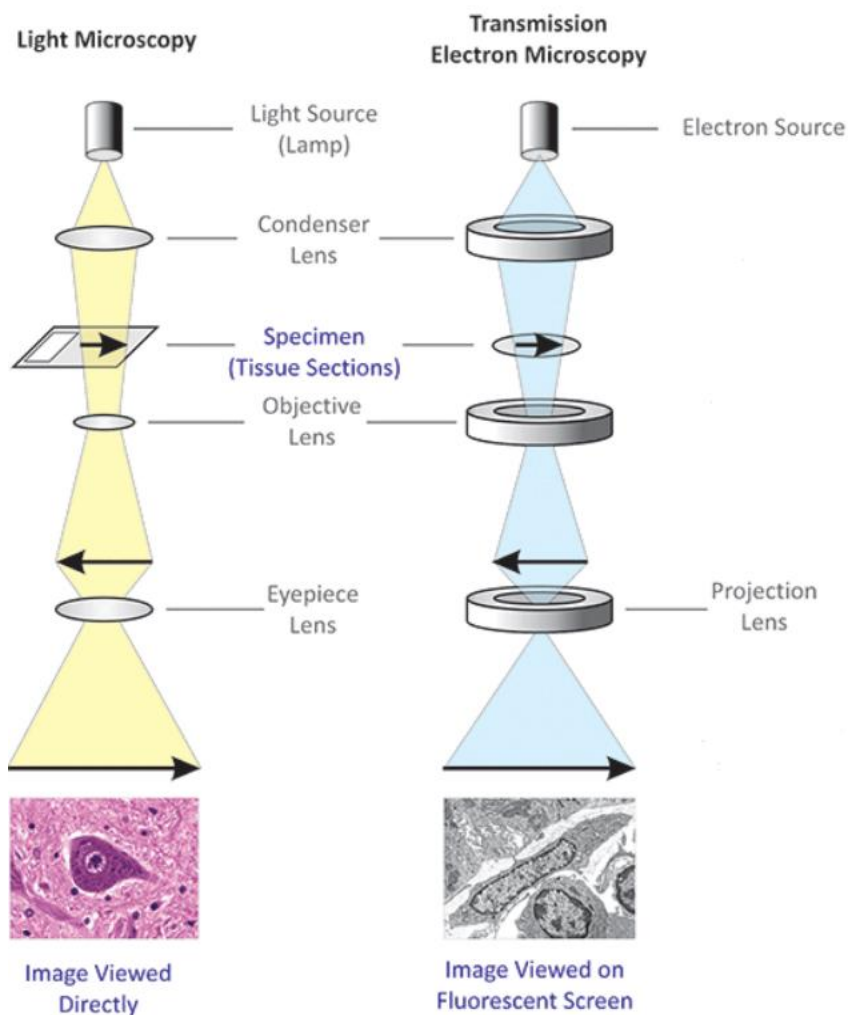
```

Svetlobna in elektronska mikroskopija



Debelina vzorca > 5 μm
Povečava 500X - 1500X
Valovna dolžina: > 400 nm

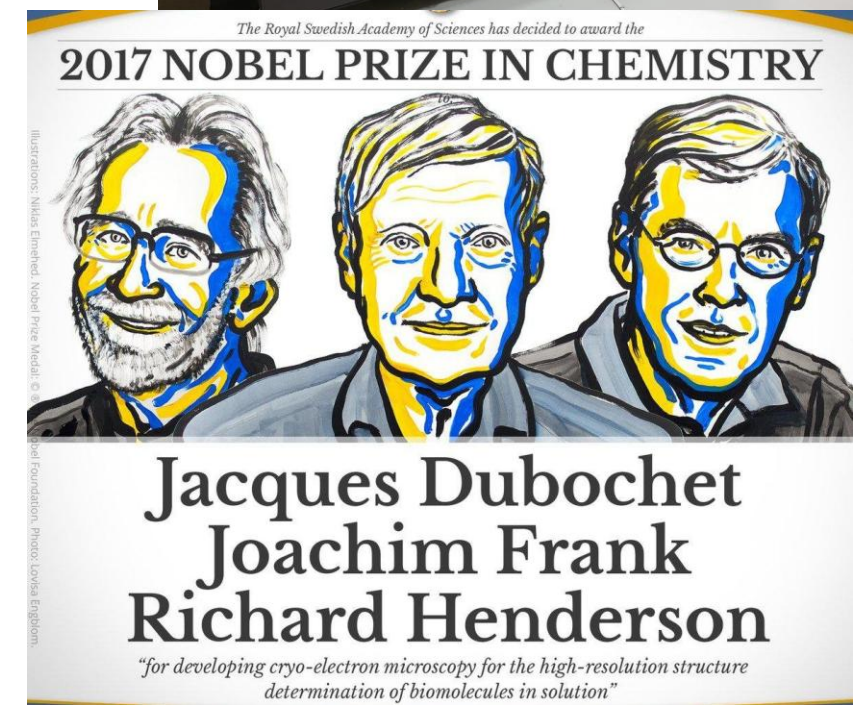
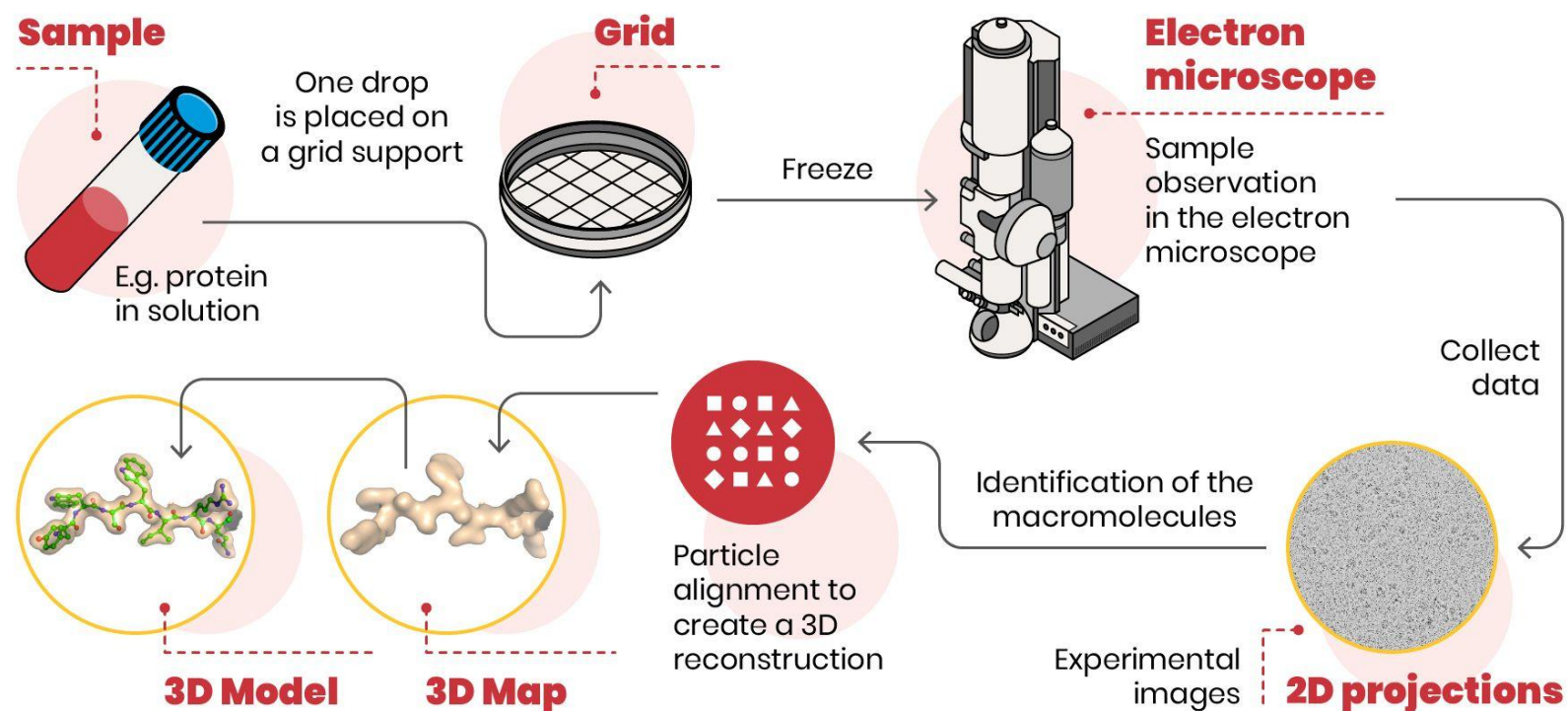
<https://microbiologyinfo.com/>



Debelina vzorca < 0,1 μm
Povečava 100 000X – 300 000X
Valovna dolžina: 2 pm

Krioelektronska mikroskopija

Zajeli bomo posnetke s krioelektronskim mikroskopom in tako določili 3D strukturo proteinov.





Krioelektronska mikroskopija



- Dobimo 16-bitne posnetke, ločljivosti 4096*4096, kjer so vidni proteinski kompleksi
- Vsak posnetek je opremljen z zelo podrobno metapodatkovno datoteko .xml.
- Posnetki so v obliki .mrc. - odprt binarni format, kasneje jih med analizo pretvorimo v obliko .tif, ki zasede ~20% manj diskovnega prostora brez izgube informacij
- Proizvajalec opreme omogoča dostop do slikovnega kodeka, ki omogoči hiter predogled .mrc posnetkov v Windows okolju.
- Obstaja področno specifični repozitorij za surove posnetke krioelektronske mikroskopije - EMPIAR
- Na dan lahko mikroskop proizvede do 1,5 TB slikovnih podatkov.

```
1 <MicroscopeImage xmlns="http://schemas.datacontract.org/2004/07/Fei.SharedObjects" xmlns:i="http://www.w3.org/2001/XMLSchema-instance"><name>Empty</name><uniqueID>
6592869d-5b37-416d-bdd3-150e35b5bcfa</uniqueID><CustomData xmlns:a="http://schemas.microsoft.com/2003/10/Serialization/Arrays"><a:KeyValueOfstringanyType><a:Key>DoseOnCamera</a:Key><a:Value i:type="b:double"
xmlns:b="http://www.w3.org/2001/XMLSchema">38.835283529204389</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>StemMagnification</a:Key><a:Value i:type="b:boolean"
xmlns:b="http://www.w3.org/2001/XMLSchema">false</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Aperture[C2].Name</a:Key><a:Value i:type="b:string" xmlns:b="http://www.w3.org/2001/
XMLSchema">50</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Aperture[OB].Name</a:Key><a:Value i:type="b:string" xmlns:b="http://www.w3.org/2001/
XMLSchema">100</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Aperture[SA].Name</a:Key><a:Value i:type="b:string" xmlns:b="http://www.w3.org/2001/
XMLSchema">None</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].TimeStamp</a:Key><a:Value i:type="b:long" xmlns:b="http://www.w3.org/2001/
XMLSchema">1731556423517936</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].PixelValueToCameraCounts</a:Key><a:Value i:type="b:int" xmlns:b="http://www.w3.org/2001/
XMLSchema">1</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].ExposureTime</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
XMLSchema">42.14479</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].CountsToElectrons</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
XMLSchema">0.009686576</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].AlignIntegratedImage</a:Key><a:Value i:type="b:boolean" xmlns:b="http://www.w3.org/2001/
XMLSchema">true</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].CommercialName</a:Key><a:Value i:type="b:string" xmlns:b="http://www.w3.org/2001/XMLSchema">Falcon
3EC</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].DriftCorrected</a:Key><a:Value i:type="b:boolean" xmlns:b="http://www.w3.org/2001/
XMLSchema">true</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].DriftCorrectionConfidence</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
XMLSchema">0.8042934</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].DriftCorrectionClipping</a:Key><a:Value i:type="b:boolean" xmlns:b="http://www.w3.org/2001/
XMLSchema">false</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].DriftCorrectionVectorX</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
XMLSchema">46.12346</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Detectors[BM-Falcon].DriftCorrectionVectorY</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
XMLSchema">-25.18686</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>IlluminationIntensity</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
XMLSchema">0.448559999999998</a:Value></a:KeyValueOfstringanyType><a:KeyValueOfstringanyType><a:Key>Dose</a:Key><a:Value i:type="b:double" xmlns:b="http://www.w3.org/2001/
```



1.2 - 1.4 Povzetek in opis raziskovalnih podatkov

- Načrtovanje raziskave
- Priprava načrta ravnanja z raziskovalnimi podatki
- <https://nrrp.openscience.si>





2. Shranjevanje podatkov pred objavo

- Podatki, z izjemo posnetkov krioelektronske mikroskopije, se hranijo na delovnih postajah pri znanstvenih napravah in/ali na osebnih računalnikih zaposlenih in/ali na mrežnih diskih znotraj inštituta. Vse lokacije so avtomatsko varnostno kopirane.
- Dostop ni posebej omejen, vsaka oseba potrebuje uporabniško ime in geslo za vstop v operacijski sistem in za dostop do mrežnih lokacij.
- Ne bomo pridobivali občutljivih podatkov.



2. Shranjevanje podatkov pred objavo

- Posnetki krioelektronske mikroskopije se hranijo na različnih HPC gručah: Ažmanov računski center na Kemijskem inštitutu, HPC Arnes, HPC VEGA, Elixir Slovenija.
- Večina podatkov krioelektronske mikroskopije ni varnostno kopiranih.
- Podatki so dostopni samo iz določenih uporabniških računov.
- Med analizo se lahko količina podatkov podvoji (zagotavljanje zadostne količine diskovnega prostora).



2. Shranjevanje in varnostno kopiranje podatkov

- Priprava načrta ravnanja z raziskovalnimi podatki
- <https://nrrp.openscience.si>





Obdelava in analiza podatkov

- Obdelava podatkov kromatografije, fluorimetrije in kalorimetrije
- Pretvorba v bolj odprt format .csv
- Analiza in priprava končnih slik (npr. za objavo v članku)
- Priprava 'readme' datoteke
- Objava v splošnem repozitoriju Zenodo
- Objava s področno specifičnih repozitorijih



Zbiranje podatkov

- Pridobili smo naslednje podatke:
 - Tekočinska kromatografija - numerični podatki v obliki .csv (in zaprtem formatu .res)
 - Temperaturna stabilnost proteinov - numerični podatki v obliki .ixo (vrsta .xml formata, zaprt format), rezultate izvozimo v .csv
 - Izotermalna titracijska kalorimetrija - numerični podatki v obliki itc (berljiv tekst, zaprt format)
 - Surove posnetki iz krioelektronskega mikroskopa - slikovni podatki v obliki .mrc ali .tif, poleg posnetkov so še metapodatkovne datoteke v obliki .xml
 - Rezultati iz krioelektronske analize bodo karte gostote - volumetrični podatki v obliki .mrc in strukturni modeli - tekstovni podatki v obliki .pdb/.cif (.mrc in .pdb/.cif formata odprtokodna)



Zbiranje podatkov

- Nadaljnja analiza ni potrebna vendar je treba poskrbeti za FAIR:
 - Tekočinska kromatografija - numerični podatki v obliki .csv (in zaprtem formatu .res)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Chrom.1		Chrom.1		Chrom.1		Chrom.1		Chrom.1		Chrom.1	
2	UV 1_280		Cond		Conc B		Injection		Fraction		UV 1_280_CUT_TEMP@100,BASEM	
3	ml	mAU	ml	mS/cm	ml	%	ml	Injection	ml	Fraction	ml	mAU
4	-13.006014	0	-13.006014	0.54665	-13.006014	0	0.000737		0.996644	1.A.2		0
5	-13.002701	-149.524139	-13.002701	0.543528	-12.997731	0	44.924721		1.997123	1.A.3	0.003313	-4.724196
6	-12.999388	-148.84967	-12.999388	0.487834	-12.989449	0	57.451088		2.996918	1.A.4	0.006626	-4.724196
7	-12.996075	-148.425171	-12.996075	0.384545	-12.981167	0			3.996939	1.A.5	0.009939	-4.724196
8	-12.992762	-148.536652	-12.992762	0.293179	-12.972884	0			4.996994	1.A.6	0.013252	-4.724196
9	-12.989449	-148.990387	-12.989449	0.22938	-12.964602	0			5.997616	1.A.7	0.016565	-4.724196
10	-12.986136	-150.192307	-12.986136	0.188666	-12.956319	0			6.997641	1.A.8	0.019877	-4.724196
11	-12.982823	-151.123047	-12.982823	0.163311	-12.948037	0			7.997698	1.A.9	0.02319	-4.724196
12	-12.97951	-151.64415	-12.979509	0.145174	-12.939754	0			8.997841	1.A.10	0.026504	-4.724196
13	-12.976197	-152.012878	-12.976196	0.131845	-12.931472	0			9.997671	1.A.11	0.029817	-4.724196
14	-12.972884	-152.256607	-12.972883	0.121378	-12.923189	0			10.997728	1.A.12	0.03313	-4.724196
15	-12.969571	-152.412994	-12.96957	0.112233	-12.914907	0			11.997749	1.B.12	0.036443	-4.724196
16	-12.966258	-152.471893	-12.966257	0.104292	-12.906625	0			12.998375	1.B.11	0.039756	-4.724196
17	-12.962945	-152.275757	-12.962944	0.097551	-12.898342	0			13.99843	1.B.10	0.043069	-4.724196
18	-12.959632	-151.884628	-12.959631	0.091949	-12.890059	0			14.998447	1.B.9	0.046382	-4.724196



Obdelava in analiza podatkov

- Priprava 'readme' datoteke

ReadMe Template en.txt

```
1 This (ReadMe) file was created on [YYYY-MM-DD] by [NAME AND SURNAME] <if multiple authors contributed to the creation of the file, please list all authors>
2 <Auxiliary text in square brackets should be deleted before completing your document>
3 <if you do not have the appropriate answers, please indicate with />
4
5 -----
6 GENERAL INFORMATION
7 -----
8
9 Name of the dataset:
10
11 <please enter the requested information>
12 Author/leading researcher details
13 Name and surname:
14 ORCID:
15 Institution:
16 Address:
17 Email:
18
19 Information about the author(s) <if you have data from other authors, please indicate them (create new input fields for each author)>
20 Name and surname:
21 ORCID:
22 Institution:
23 Address:
24 Email:
25
26 Date of data collection: <YYYY-MM-DD> <if the data collection took place over a longer period of time, please specify a time period <from YYYY-MM-DD to YYYY-MM-DD>
27
28 Geographical location of data collection: <specify latitude, longitude or city/region, country>
29
30 Information on the funders/programmes/projects that made the data collection possible:
31
32 -----
33 SHARING/ACCESSING INFORMATION
34 -----
35
36 Data licences/restrictions:
37
38 Links to publications that cite or use the data:
39
40 Links to other publicly available data sites:
41
42 Links to ancillary databases:
43
44 Was the data obtained from another source?
45 If yes, please specify the source(s):
46
47 -----
48 VIEWING DATA AND FILES
49 -----
50
51 File list: <list of all files or folders, with a short description>
```



Zbiranje podatkov

- Nadaljnja analiza je potrebna in poskrbeti je treba za FAIR:
 - Temperaturna stabilnost proteinov - numerični podatki v obliki .ixo (vrsta xml formata, zaprt format), rezultate izvozimo v .csv
 - Izotermalna titracijska kalorimetrija - numerični podatki v obliki .itc (berljiv tekst, zaprt format), rezultate izvozimo v .csv



Zbiranje podatkov

- Nadaljnja analiza je potrebna in poskrbeti je treba za FAIR:
 - Surovi posnetki iz krioelektronskega mikroskopa - slikovni podatki v obliki .mrc ali .tif, poleg posnetkov so še metapodatkovne datoteke v obliki .xml

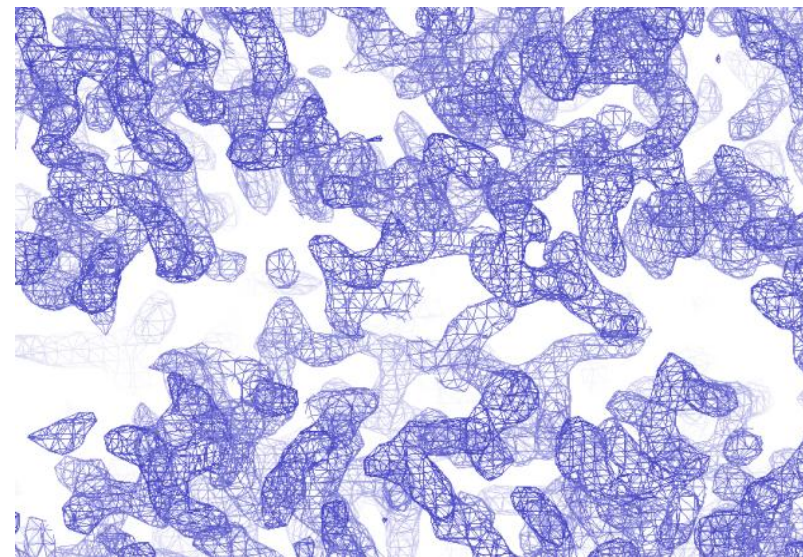


Obdelava in analiza podatkov

Obdelava podatkov krioelektronske mikroskopije s programsko opremo:

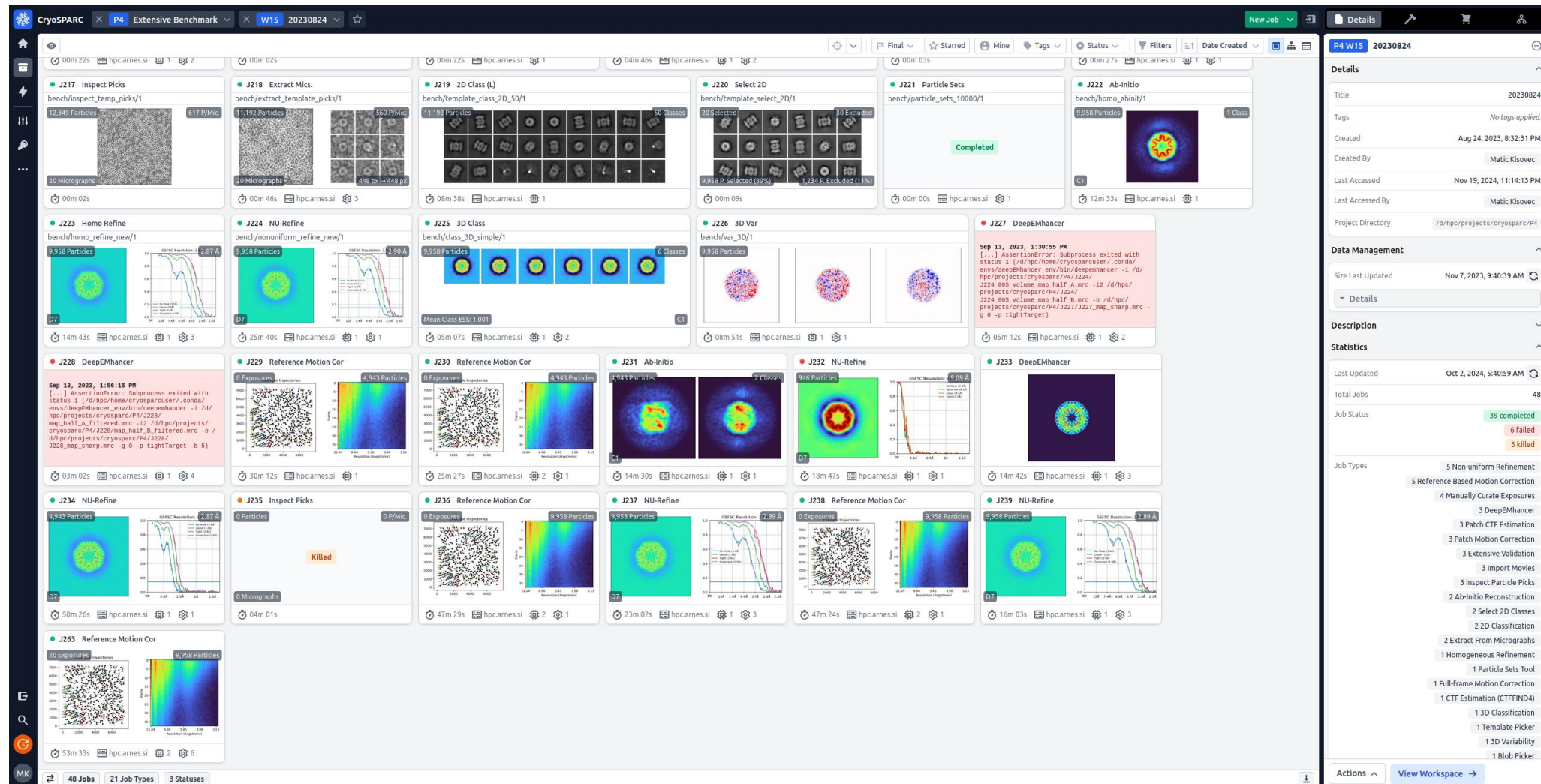
- naziv programske opreme: cryoSPARC
- Programska oprema izboljša kvaliteto posnetkov, poišče proteine, integrira/sešteje enako orientirane delce, preko Fourierjeve transformacije pridemo iz 2D v 3D (karta gostote krioEM, piljenje končne karte gostote krioEM)
- Ko imamo izpiljeno karto gostote, vanjo zgradimo 3D model, ki predstavlja položaj vsakega atoma v prostoru.
- Programska oprema (npr. Phenix) v datoteko s koordinatami dodaja tudi že prve metapodatke neposredno v .pdb/.cif
- Preostanek metapodatkov se obvezno vnese ročno pri vnosu v repozitorij

Karta krioEM gostote



Obdelava in analiza podatkov - krioelektronska mikroskopija

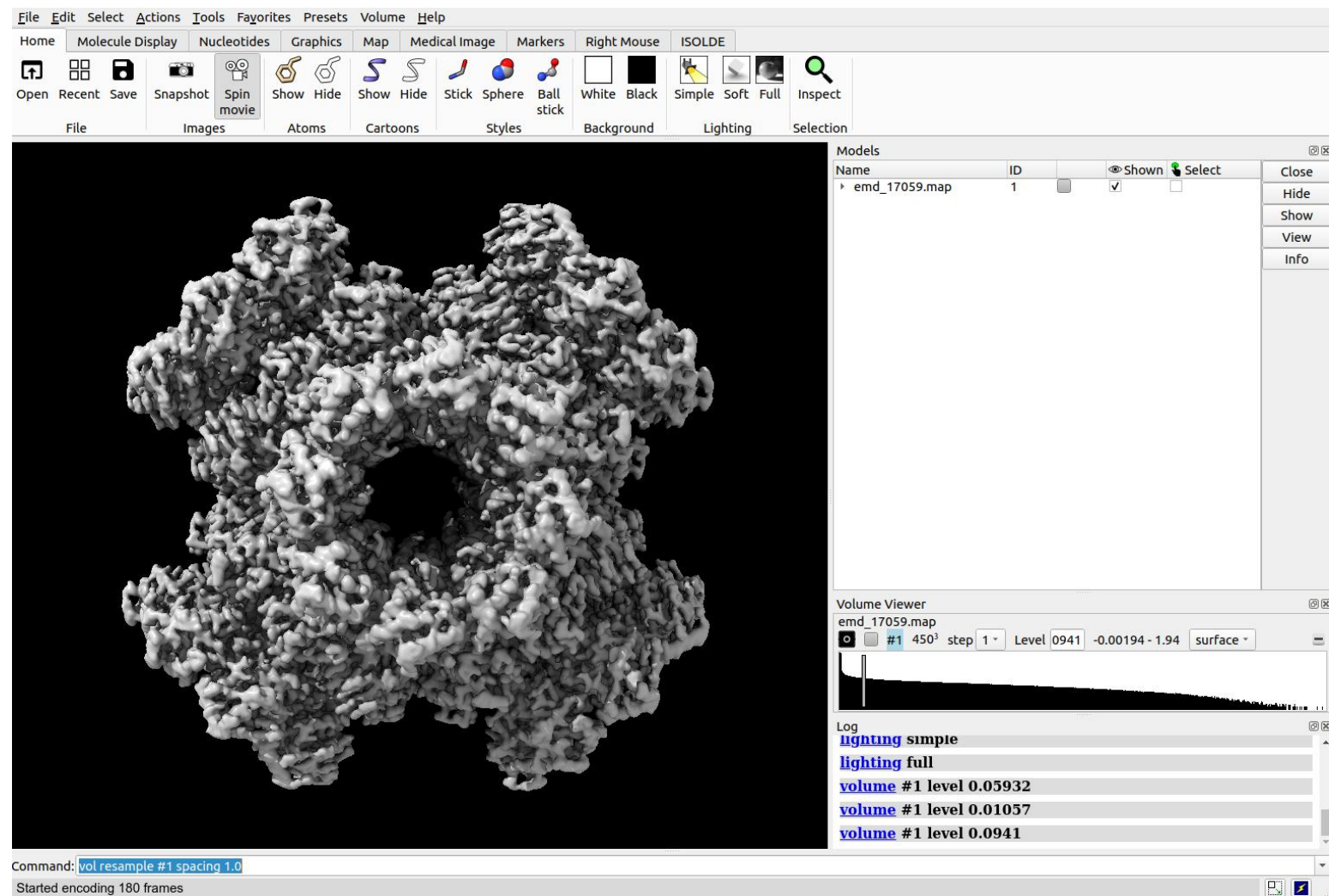
- CryoSPARC
- Uporabljamo HPC infrastrukturo



Obdelava in analiza podatkov - krioelektronska mikroskopija

Rezultati analize krioelektronskih posnetkov
so karte krioEM gostote - volumetrični podatki
v obliki .mrc

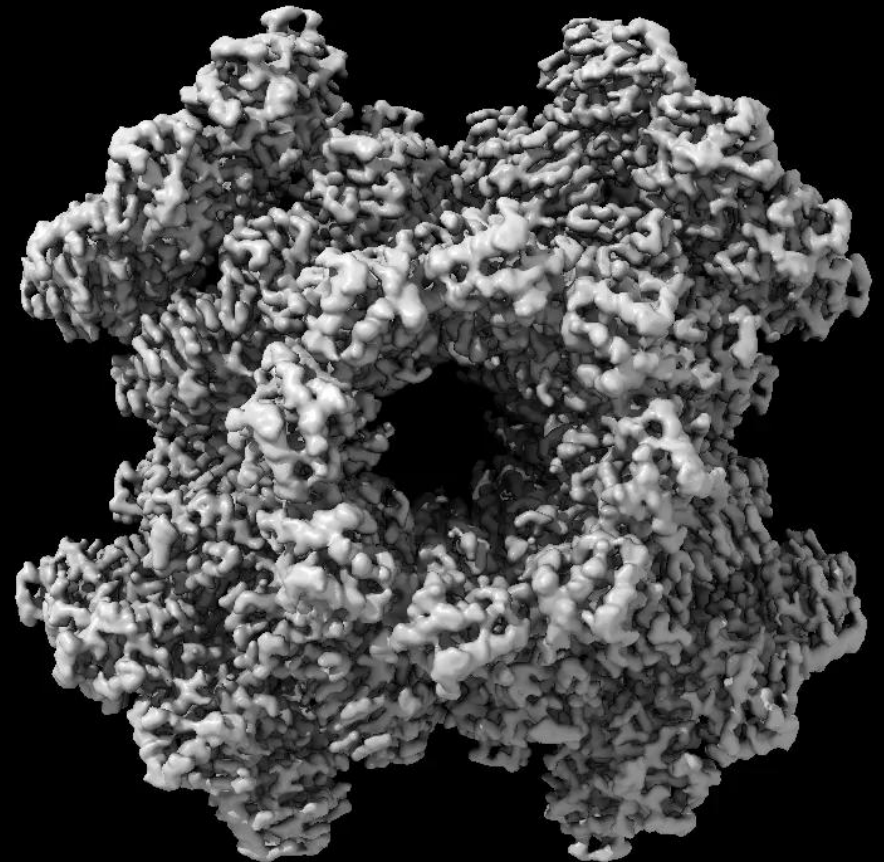
- Pričnemo z datotekami .mrc, ki vsebujejo 3D karto krioEM gostote, brez metapodatkov
- Za ogled kart uporabimo programsko opremo ChimeraX



Obdelava in analiza podatkov - krioelektronska mikroskopija

Rezultati analize krioelektronskih posnetkov so karte
krioEM gostote - volumetrični podatki v obliki .mrc

VIDEO



Obdelava in analiza podatkov - krioelektronska mikroskopija

Rezultati analize krioelektronskih posnetkov
so karte krioEM gostote - volumetrični podatki
v obliki .mrc

- Obstaja področno specifični repozitorij za surove posnetke krioelektronske mikroskopije - EMDB
- V EMDB ročno vnesemo metapodatke med depozicijo in povežemo z EMPIAR in PDB

The screenshot displays the EMDB (Electron Microscopy Data Bank) website interface. At the top, the EMDB logo is visible, along with a search bar and navigation links. The main content area shows the details for EMD-17063, including a 3D visualization of the protein structure, its classification as a single-particle map with a resolution of 3.16 Å, and a list of deposition authors. The right sidebar contains a list of links to related resources, such as the 3D volume, experimental metadata, and validation reports. The bottom section of the page provides additional information, including the PubMed ID, DOI, and ISSN.

EMDB
Electron Microscopy Data Bank

Enter your search term(s) in the box below or [build an advanced search query](#)

Search EMDB...

Examples: 1001, Apoferritin, Tomography, Rossmann MG, 5A1A

Home | Deposition | Documentation | Resources | FTP Archive | REST API | About | Feedback | Share

EMD-17063
Single-particle
3.16 Å

Overview | 3D View | Sample | Experiment | Validation | Volume Browser | Additional data | Links

EMD-17063 Download

Local refinement of cubic assembly from truncated PVY coat protein mutation

Sample Organism: [Potato virus Y strain NTN](#)
Sample: [truncated coat protein \(dN49C40\) with C-terminal His tag and K176C mutation](#)
Fitted models: [8opk](#) (Avg. Q-score: 0.428)
Raw data: [EMPIAR-11549](#)

Deposition Authors: [Kavcic L](#), [Kezar A](#), [Podobnik M](#)

From structural polymorphism to structural metamorphosis of the coat protein of nonsegmented filamentous potato virus Y

[Kavcic L](#), [Kezar A](#), [Koritnik N](#), [Znidaric MT](#), [Klobucar T](#), [Vivic Z](#), [Merzel F](#), [Holden E](#), [Benesch JLP](#), [Podobnik M](#)
(2024) Commun Chem, 7, 14 - 14

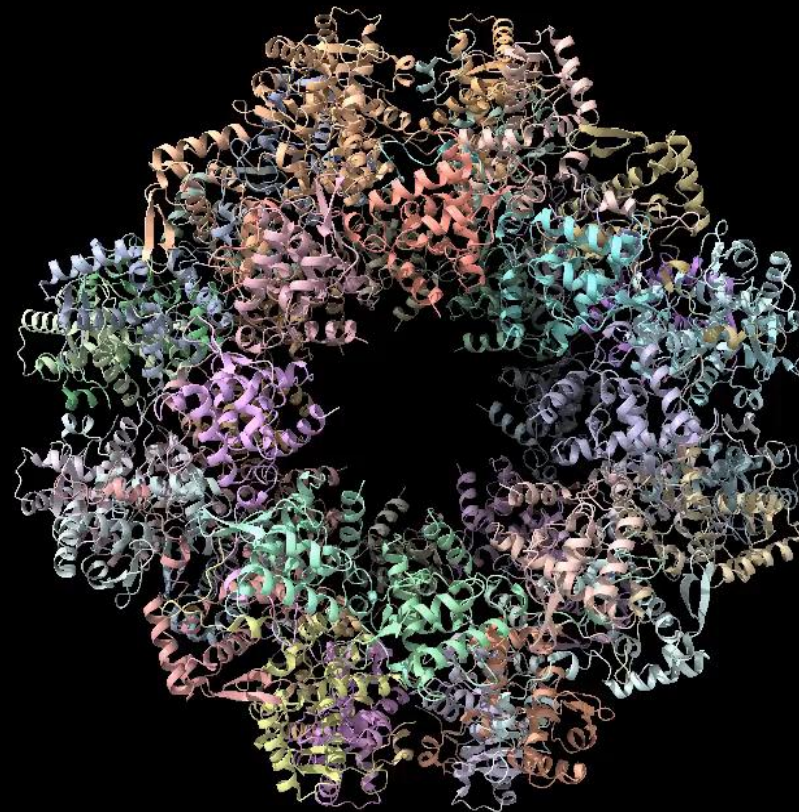
PUBMED: [38233506](#)
DOI: [doi:10.1038/s42004-024-01100-x](#)
ISSN: 2399-3669

3D volume (map.gz)
Experimental metadata (x)
Experimental Metadata (c)
Map-only Validation report
Map-only Validation report
Map-only Validation report
Map and PDB model 8opk
Map and 8opk FULL report
EMICSS annotations (xml)

Obdelava in analiza podatkov - krioelektronska mikroskopija

V karte krioEM gostote - volumetrični podatki v
obliki .mrc zgradimo strukturne modele -
tekstovni podatki v obliki .pdb/.cif

VIDEO



Obdelava in analiza podatkov - krioelektronska mikroskopija

V karte krioEM gostote - volumetrični podatki v obliki .mrc zgradimo strukturne modele - tekstovni podatki v obliki .pdb/.cif

- Obstaja področno specifični repozitorij za strukturne modele – PDB

RCSB PDB Deposit Search Visualize Analyze Download Learn About Careers COVID-19 Help Contact us MyPDB

RCSB PDB PROTEIN DATA BANK 227,344 Structures from the PDB 1,068,577 Computed Structure Models (CSM) Enter search term(s), Entry ID(s), or sequence Include CSM Help

Advanced Search | Browse Annotations

PDB-101 PDB EMDDataResource NAKB wwPDB Foundation PDB-Dev

Structure Summary Structure Annotations Experiment Sequence Genome Versions

Biological Assembly 1

8OPJ
Global refinement of cubic assembly from truncated PVY coat protein
PDB DOI: <https://doi.org/10.2210/pdb8OPJ/pdb> EM Map EMD-17062: EMD EMDDataResource

Classification: VIRUS LIKE PARTICLE
Organism(s): Potato virus Y strain NTN
Expression System: Escherichia coli BL21(DE3)
Mutation(s): Yes

Deposited: 2023-04-07 **Released:** 2024-01-24
Deposition Author(s): Kavcic, L., Kezar, A., Podobnik, M.
Funding Organization(s): Slovenian Research Agency, Ministry of Education, Youth and Sports

Experimental Data Snapshot
Method: ELECTRON MICROSCOPY
Resolution: 2.99 Å
Aggregation State: PARTICLE
Reconstruction Method: SINGLE PARTICLE

wwPDB Validation
Metric Clashscore Ramachandran outliers Sidechain outliers Validation Full PDF Validation (XML - gz) Validation (CIF - gz) Biological Assembly 1 (CIF - gz) Percentile relative to all EM structures

Display Files Download Files Data API

FASTA Sequence
PDBx/mmCIF Format
PDBx/mmCIF Format (gz)
BinaryCIF Format (gz)
PDB Format
PDB Format (gz)
PDBML/XML Format (gz)
EM Map EMD-17062 (map - gz)
PDB format-like files (tar.gz)

3D Report Full Report Value 6 0 0.3% Better

This is version 1.1 of the entry. See complete history.

Literature Download Primary Citation

From structural polymorphism to structural metamorphosis of the coat protein of flexuous filamentous potato virus Y.
Kavcic, L., Kezar, A., Koritnik, N., Znidaric, M.T., Klobucar, T., Vicio, Z., Merzel, F., Holden, E., Benesch, J.L.P., Podobnik, M. (2024) Commun Chem 7: 14-14
PubMed: 38233506 Search on PubMed Search on PubMed Central
DOI: <https://doi.org/10.1038/s42004-024-01100-x>
Primary Citation of Related Structures:
8OPA, 8OPB, 8OPC, 8OPD, 8OPE, 8OPF, 8OPG, 8OPH, 8OPJ, 8OPK, 8OPL

PubMed Abstract:
The structural diversity and tunability of the capsid proteins (CPs) of various icosahedral and rod-shaped viruses have been well studied and

Macromolecule Content
• Total Structure Weight: 1,073.3 kDa
• Atom Count: 65,352
• Modelled Residue Count: 8,112
• Deposited Residue Count: 9,312
• Unique protein chains: 1

Obdelava in analiza podatkov - krioelektronska mikroskopija

V karte krioEM gostote - volumetrični podatki v obliki .mrc zgradimo strukturne modele - tekstovni podatki v obliki .pdb/.cif

- Obstaja področno specifični repozitorij za strukturne modele – PDB
- Programska oprema za gradnjo modela v datoteko .pdb/.cif vključi že mnoge metapodatke (<https://mmcif.wwpdb.org>)
- Med depozicijo v zbirko PDB ročno vnesemo metapodatke, ki jih zahteva PDB
- Med depozicijo v zbirko PDB povežemo z EMPIAR in EMDB

```
1 data_80PJ
2 #
3 _entry.id      80PJ
4 #
5 _audit_conform.dict_name      mmcif_pdbx.dic
6 _audit_conform.dict_version   5.384
7 _audit_conform.dict_location  http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
8 #
9 loop_
10 _database_2.database_id
11 _database_2.database_code
12 _database_2.pdbx_database_accession
13 _database_2.pdbx_DOI
14 PDB      80PJ      pdb_000080pj  10.2210/pdb80pj/pdb
15 WWPDB D_1292128956 ?      ?
16 EMDB    EMD-17062  ?      ?
17 #
18 loop_
19 _pdbx_audit_revision_history.ordinal
20 _pdbx_audit_revision_history.data_content_type
21 _pdbx_audit_revision_history.major_revision
22 _pdbx_audit_revision_history.minor_revision
23 _pdbx_audit_revision_history.revision_date
24 1 'Structure model' 1 0 2024-01-24
25 2 'Structure model' 1 1 2024-01-31
26 #
27 _pdbx_audit_revision_details.ordinal      1
28 _pdbx_audit_revision_details.revision_ordinal 1
29 _pdbx_audit_revision_details.data_content_type 'Structure model'
30 _pdbx_audit_revision_details.provider      repository
31 _pdbx_audit_revision_details.type          'Initial release'
32 _pdbx_audit_revision_details.description   ?
33 _pdbx_audit_revision_details.details       ?
```

Obdelava in analiza podatkov - krioelektronska mikroskopija

V karte krioEM gostote - volumetrični podatki v obliki .mrc zgradimo strukturne modele - tekstovni podatki v obliki .pdb/.cif

- Obstaja področno specifični repozitorij za strukturne modele – PDB
- Programska oprema za gradnjo modela v datoteko .pdb/.cif vključi že mnoge metapodatke (<https://mmcif.wwpdb.org>)
- .pdb/.cif je zelo zanimiv format saj ena datoteka vsebuje metapodatke in raziskovalne podatke

```
13995 loop_
13996 _atom_site.group_PDB
13997 _atom_site.id
13998 _atom_site.type_symbol
13999 _atom_site.label_atom_id
14000 _atom_site.label_alt_id
14001 _atom_site.label_comp_id
14002 _atom_site.label_asym_id
14003 _atom_site.label_entity_id
14004 _atom_site.label_seq_id
14005 _atom_site.pdbx_PDB_ins_code
14006 _atom_site.Cartn_x
14007 _atom_site.Cartn_y
14008 _atom_site.Cartn_z
14009 _atom_site.occupancy
14010 _atom_site.B_iso_or_equiv
14011 _atom_site.pdbx_formal_charge
14012 _atom_site.auth_seq_id
14013 _atom_site.auth_comp_id
14014 _atom_site.auth_asym_id
14015 _atom_site.auth_atom_id
14016 _atom_site.pdbx_PDB_model_num
14017 ATOM 1 N N . LYS A 1 5 ? 295.206 260.190 297.603 1.00 48.62 ? 53 LYS Aa N 1
14018 ATOM 2 C CA . LYS A 1 5 ? 295.875 258.923 297.875 1.00 51.24 ? 53 LYS Aa CA 1
14019 ATOM 3 C C . LYS A 1 5 ? 295.155 258.158 298.977 1.00 52.87 ? 53 LYS Aa C 1
14020 ATOM 4 O O . LYS A 1 5 ? 294.362 258.729 299.724 1.00 55.49 ? 53 LYS Aa O 1
14021 ATOM 5 C CB . LYS A 1 5 ? 297.330 259.162 298.276 1.00 50.53 ? 53 LYS Aa CB 1
...
```

```
79365 ATOM 65349 C CA . ALA VA 1 174 ? 246.815 303.859 224.205 1.00 68.34 ? 222 ALA Bv CA 1
79366 ATOM 65350 C C . ALA VA 1 174 ? 248.024 303.034 224.633 1.00 66.53 ? 222 ALA Bv C 1
79367 ATOM 65351 O O . ALA VA 1 174 ? 249.102 303.575 224.876 1.00 64.55 ? 222 ALA Bv O 1
79368 ATOM 65352 C CB . ALA VA 1 174 ? 245.573 303.417 224.963 1.00 65.92 ? 222 ALA Bv CB 1
79369 #
79370
```

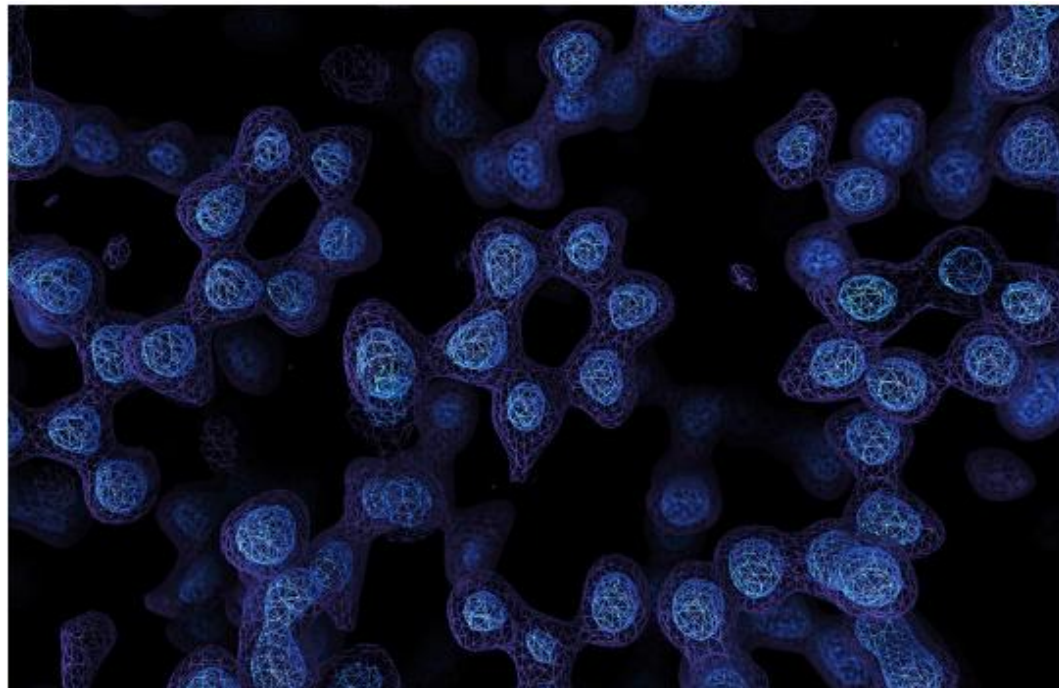


Krioelektronska mikroskopija

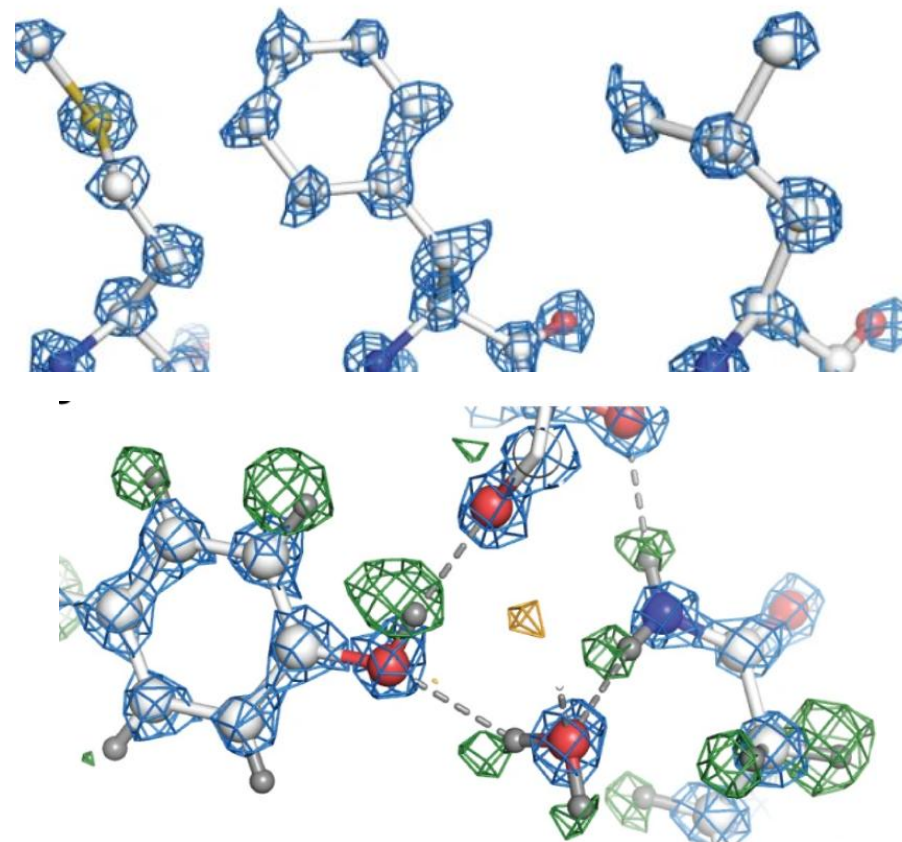
nature

NEWS • 03 JUNE 2020

**‘It opens up a whole new universe’:
Revolutionary microscopy technique sees
individual atoms for first time**



Leto 2020: atomska ločljivost



Nakane, T., Kotecha, A., Sente, A. *et al.* *Nature* **587**, 152–156 (2020).

Yip, K.M., Fischer, N., Paknia, E. *et al.* *Nature* **587**, 157–161 (2020).



3. Zagotovitev podatkov na način FAIR

- Priprava načrta ravnanja z raziskovalnimi podatki
- <https://nrrp.openscience.si>



Deljenje (objava v repozitoriju)

Surovi podatki so objavljeni v EMPIAR

www.ebi.ac.uk/empiar/EMPIAR-11549

www.ebi.ac.uk/empiar/deposition



EMPIAR home

Deposition

REST API

FAQ

About EMPIAR

Policies

Feedback

Share

EMPIAR
deposition
system

You are logged in as **matick**

Edit profile

Helpdesk

Deposition manual

Invite reviewers

Get empiar-depositor API token

Log out

Please make sure that you have read and understood the [EMPIAR Policies & Procedures](#) before you start your deposition

Create a new deposition

Create a new deposition from XML

Show 10 entries

Search:

ID	EMPIAR ID (accession code)	Title	Upload Token	Status	Permissions
47484908	EMPIAR-11550	Cryo electron microscopy of Virus-like Particle based on PVY coat protein with T43C and D136C mutation	wwfHsHdQeqvjeMpM	REL	View and Edit
47484907	EMPIAR-11549	Cryo electron microscopy of assemblies based on truncated PVY coat protein with K176C mutation	mmaAYUpLHOFAURdB	REL	View and Edit
47484906	EMPIAR-11548	Cryo electron microscopy of assemblies based on truncated PVY coat protein	46bbgOtbH1wPCjea	REL	View and Edit

TB!!

EMPIAR total size per year and cumulatively

Year	Annually added data storage	Cumulative data storage
2024	1128.7	5007.9
2023	1479	3879.2
2022	904.2	2400.2
2021	788.4	1496
2020	450.5	707.6
2019	150.1	257.1

Deljenje (objava v repozitoriju)

Obdelani podatki kriomikroskopije (PDB)

www.rcsb.org/structure/8OPJ

<https://deposit-2.wwpdb.org/deposition>



wwPDB Deposition: D_1292137787 -- IDs: 9EYO(HPUB), EMD-50059(HPUB)

Deposition locked

All items

Mandatory items

Navigation

- ✓ Instructions
- ✓ Communication
- ✓ Review
 - Assembly Review
 - Validation reports
- ✓ Re-upload files
- ✓ Admin
 - ✓ Contact information
 - ✓ Grant information
 - ✓ Release status
 - ✓ Entry title & author
 - ✓ Citation information
- ✓ Macromolecules
 - ✓ 1) Actinoporin
- ✓ EM sample
 - ✓ Overall sample description
- ✓ EM experiment
 - ✓ Specimen preparation
 - ✓ Microscopy
 - Image recording**
 - ✓ Reconstruction
 - ✓ Fitting interpretation
- ✓ Ligands
- ✓ Related entries
- ✓ Summary & conditions

Image recording

Microscopy settings*: (1) TFS GLACIOS (200kV, x150000) v

Film/CCD/Direct electron detector model*: FEI Falcon III (4k x 4k) v

Detector mode: COUNTING v

Width (pixel units): 4096

Height (pixel units): 4096

Sampling interval (μm):

Average electron dose per image (e⁻/Å²)*: 40

Average exposure time per image (seconds):

Number of grids imaged:

Number of images: 11470

Details:

Continue to next section

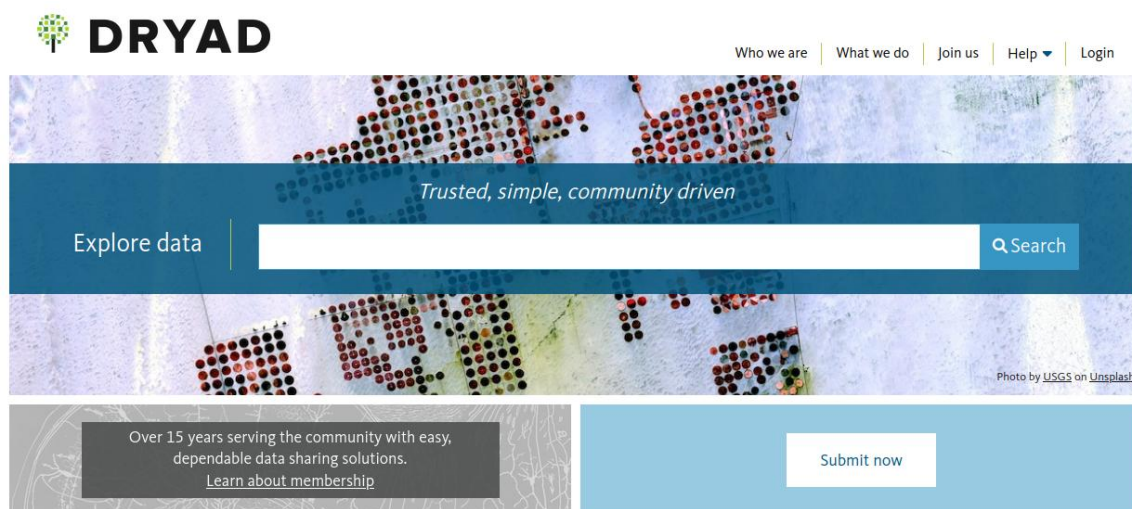
Data Storage Growth

PDB Core Archive and Related Holdings (single copy)

Year	Size of PDB legacy archive snapshot	Size of PDB versioned archive	Size of EMDB Core Archive
2023	1242 GB for 214,121 PDB structures	231 GB	13 TB for 32,033 EMDB entries
2022	1086 GB for 199,755 PDB structures	218 GB	7.5 TB for 25,319 EMDB entries
2021	923 GB for 185,541 PDB structures	174 GB	4.5 TB for 18,059 EMDB entries

Obdelava in analiza podatkov

- Kjer ni področno specifičnih repozitorijev - objava v splošnem repozitoriju npr. Zenodo
- Dryad npr. za objavo med recenzijo, recenzentski dostop do pomembnih datotek
- Drugi splošni repozitoriji





Deljenje (objava v repozitoriju)

Deljenje in objava v splošnih repozitorijih (Zenodo, Figshare, Dryad, Mendeley Data ...) ali institucionalnih repozitorijih (RUL, DKUM, RUP, RUNG, DiRROS, ReVIS)

Primer: Podatki iz krioelektronske mikroskopije v repozitoriju Zenodo

- <https://zenodo.org/records/12611974>
 - Preprint: <https://doi.org/10.1101/2024.04.10.588894>
 - Članek: <https://doi.org/10.1093/nar/gkae636>

Vnos podatkov tekočinske kromatografije - numerični podatki v obliki .csv (in zaprtem formatu .res)



3. Zagotovitev podatkov na način FAIR

- Priprava načrta ravnanja z raziskovalnimi podatki
- <https://nrrp.openscience.si>





Objava znanstvenega članka

[nature](#) > [communications chemistry](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 17 January 2024

From structural polymorphism to structural metamorphosis of the coat protein of flexuous filamentous potato virus Y

[Luka Kavčič](#), [Andreja Kežar](#), [Neža Koritnik](#), [Magda Tušek Žnidarič](#), [Tajda Klobučar](#), [Žiga Vičič](#), [Franci Merzel](#), [Ellie Holden](#), [Justin L. P. Benesch](#) & [Marjetka Podobnik](#) ✉

[Communications Chemistry](#) **7**, Article number: 14 (2024) | [Cite this article](#)

Download PDF



Sections

Figures

References

[Abstract](#)

[Introduction](#)

[Results](#)

[Discussion](#)

Supplementary information

[Peer Review File](#)

[Supplementary Information](#)

[Description of Additional Supplementary Files](#)

[Supplementary data 1](#)

[reporting summary](#)

Data availability

Cryo-EM maps and atomic models have been deposited in the Electron Microscopy Data Bank (EMDB) and wwPDB, respectively, with EMDB/PDB accession codes: EMD-17046/8OPA, EMD-17047/8OPB, EMD-17048/8OPC, EMD-17049/8OPD, EMD-17050/8OPE, EMD-17051/8OPF, EMD-17052/8OPG, EMD-17053/8OPH, EMD-17054, EMD-17055, EMD-17056, EMD-17057, EMD-17058, EMD-17059, EMD-17060, EMD-17061, EMD-17062/8OPJ, EMD-17063/8OPK, EMD-17064, EMD-17065, EMD-17066, EMD-17067, EMD-17068, EMD-17069, EMD-17070, EMD-17071, EMD-17072/8OPL, EMD-17073, EMD-17074 and EMD-17075, with corresponding structures and atomic models provided in Table 1 and Supplementary Tables 1 and 2. Raw cryo-EM datasets have been deposited to the Electron Microscopy Pilot Image Archive (EMPIAR) with accession codes EMPIAR-11545 (EMD-17046/8OPA, EMD-17047/8OPB, EMD-17048/8OPC), EMPIAR-11546 (EMD-17049/8OPD, EMD-17050/8OPE), EMPIAR-11547 (EMD-17052/8OPG), EMPIAR-11548 (EMD-17053/8OPH, EMD-17054, EMD-17055), EMPIAR-11549 (EMD-17062/8OPJ, EMD-17063/8OPK) and EMPIAR-11550 (EMD-17072/8OPL). RNA nanopore sequencing data have been deposited on European nucleotide archive (ENA) with accession code PRJEB61146. All data are available in the main text, figures and supplementary information. Source data are provided with this paper as Supplementary Data file. Additional data related to this paper may be requested from the authors.

Objava znanstvenega članka

- Revija zahtevala začasno depozicijo vseh pdb/cif in kart gostote v repozitorij Dryad.
- V repozitoriju PDB in EMDB so bile datoteke še zaklenjene in zato nedostopne.
- Datoteke v EMPIAR, EMDB in PDB so bile avtomatsko odklenjene po objavi članka.

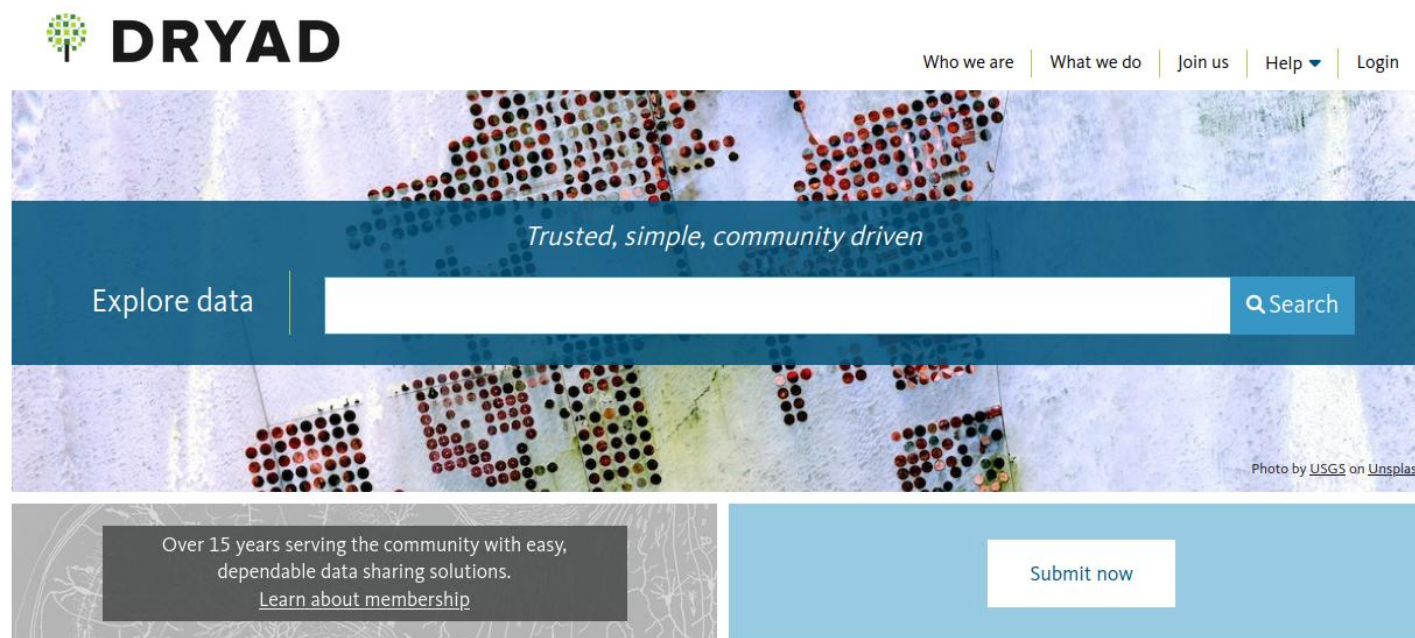


Figure 1 Figure 2 Figure 5 Figure 7 Figure 8 Suppl. Fig. 4 Suppl. Fig. 5 Suppl. Fig. 6 Suppl. Fig. 8 Suppl. Fig. 9 Suppl. Fig. 11 Suppl. Fig. 13 Suppl. Fig. 14 Suppl. Fig. 15 Suppl. Fig. 18 **Suppl. Fig. 19** Suppl. Fig. 22 Suppl. Fig. 25 +

- Supplementary data 1 (.xlsx format)

[Figure 1](#)
[Figure 2](#)
[Figure 5](#)
[Figure 7](#)
[Figure 8](#)
[Suppl. Fig. 4](#)
[Suppl. Fig. 5](#)
[Suppl. Fig. 6](#)
[Suppl. Fig. 8](#)
[Suppl. Fig. 9](#)
[Suppl. Fig. 11](#)
[Suppl. Fig. 13](#)
[Suppl. Fig. 14](#)
[Suppl. Fig. 15](#)
[Suppl. Fig. 18](#)
[Suppl. Fig. 19](#)
[Suppl. Fig. 22](#)
[Suppl. Fig. 25](#)
[+](#)



Objava znanstvenega članka

Figure 1 Figure 2 Figure 5 Figure 7 Figure 8 Suppl. Fig. 4 Suppl. Fig. 5 Suppl. Fig. 6 Suppl. Fig. 8 Suppl. Fig. 9 Suppl. Fig. 11 Suppl. Fig. 13 Suppl. Fig. 14 Suppl. Fig. 15 Suppl. Fig. 18 Suppl. Fig. 19 Suppl. Fig. 22 Suppl. Fig. 25 +

Figure 1a Raw data of fIDPnn prediction of structural disorder in CP

-----File Format-----

line 1: >protein ID (when the pssm matrix cannot be calculated for an input sequence)

line 2: sequence

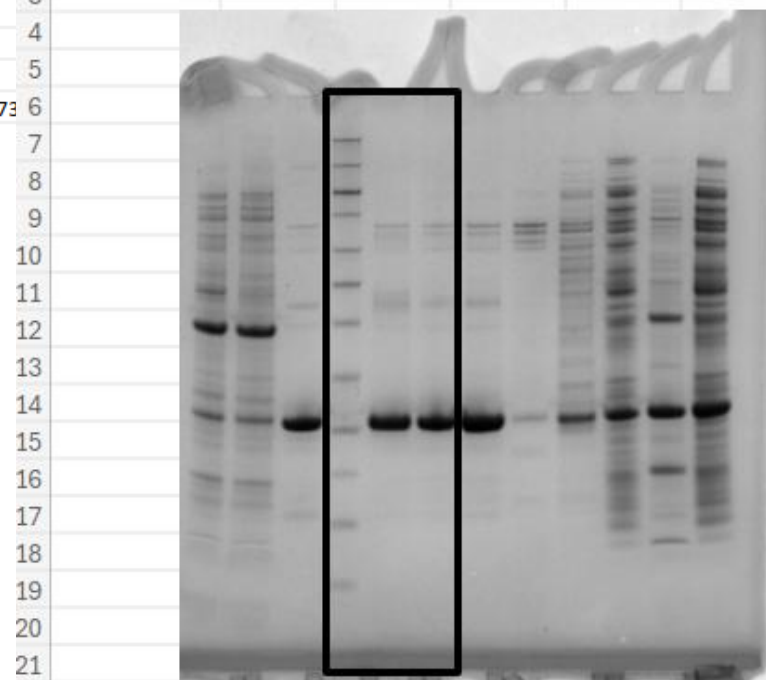
line 3: binary disorder prediction (1 = disordered residue/amino acid; 0 ordered residue)

line 4: disorder propensity (higher value denotes higher likelihood that a given residue is disordered)

>_PVYCP

	G	N	D	T	I	D	A
M	1	1	1	1	1	1	1
1	0.498	0.577	0.579	0.607	0.538	0.57	0.668

Suppl. Fig. 9a Uncropped SDS PAGE gel of purified trCPE150C under reducing (+DTT) or oxidizing (- DTT) conditions





Objava znanstvenega članka

V članku skupno:

- 11 molekulskih struktur .pdb/.cif
- 30 kart gostote
- 6 setov surovih podatkov (10 TB)
- 8 + 25 + 19 slik

Table 1 Cryo-EM data collection, refinement and validation statistics.

	VLP ^r	VLP ^h	VLP ^h +RNA	VLP ^{ΔC40:r}	VLP ^{ΔC40:h} +RNA	VLP ^{ΔC60:h}	VLP ^{ΔC79:h}	trCP (H2Tdouble ring)	trCP ^{K176C} (cubes, global)	trCP ^{K176C} (cubes, local)	VLP ^{T43C} +D136C ^h +RNA
	(EMD- 17046)	(EMD- 17047)	(EMD- 17048)	(EMD- 17049)	(EMD- 17050)	(EMD-17051)	(EMD-17052)	(EMD-17053)	(EMD-17062)	(EMD-17063)	(EMD-17072)
	(PDB 8OPA)	(PDB 8OPB)	(PDB 8OPC)	(PDB 8OPD)	(PDB 8OPE)	(PDB 8OPF)	(PDB 8OPG)	(PDB 8OPH)	(PDB 8OPJ)	(PDB 8OPK)	(PDB 8OPL)
Percentage of particles detected in the sample (%)	67	25	8	65	35	100	100	93	100		100
Data collection and processing											
Magnification	×150,000			×75,000		×150,000	×150,000	×150,000	×165,000		× 150,000
Voltage (kV)	200			300		200	200	200	300		200
Microscope	Glacios			Titan Krios		Glacios	Glacios	Glacios	Titan Krios		Glacios
Electron exposure (e ⁻ Å ⁻²)	40			84		40	42	43	32		40
Defocus range (μm)	−0.8 to −2.1 (step 0.3)			−1.3 and −0.4 (step 0.3)		−0.8 to −2.1 (step 0.3)	−0.8 to −2.1 (step 0.3)	−0.8 to −2.1 (step 0.3)	−0.3 to −3.6 (step 0.3)		−0.8 to −2.1 (step 0.3)
Pixel size (Å)	0.950			1.063		0.950	0.950	0.950	0.822		0.950
Symmetry imposed	helical, C8	helical, C1	helical, C1	helical, C8	helical, C1	helical, C1	helical, C1	C8	O	C1	helical, C1
Helical rise (Å)	43.24	5.88	3.97	43.29	4.05	4.65	4.70	39.30	N/A	N/A	4.01
Helical twist (°)	13.78	−48.11	−40.96	15.04	−41.05	−46.36	−46.47	−3.59	N/A	N/A	−41.00
Total movies	502			4805		2020	491	2862	9480		461
EMPIAR-ID	EMPIAR-11545			EMPIAR-11546		/	EMPIAR-11547	EMPIAR-11548	EMPIAR-11549		EMPIAR-11550
Initial particle/segment (no.)	109,796			2,966,850		788,376	102,608	1,406,546	1,580,128		195,881
Final particle/segment (no.)	49,140	27,500	10,266	399,617	152,296	222,692	66,934	174,238	210,779	843,116 ^a	141,112
Map resolution (Å)	3.3	3.6	3.0	3.5	3.1	2.8	3.2	2.9	3.0	3.2	2.4
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Refinement											
Initial model used (PDB-ID)	6HXZ	6HXX	6HXX	6HXZ	6HXX	6HXX	6HXX	8OPD	8OPK	8OPD	6HXX
Model resolution (Å)	3.5	3.8	3.1	3.6	3.1	2.9	3.4	3.1	3.5	3.2	2.6
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Model composition											
Number of protein chains	24	16	28	24	21	19	19	16	48	16	27
Number of RNA chains	0	0	28	0	21	0	0	0	0	0	27
Non-hydrogen atoms	34,992	23,328	53,088	66,672	33,411	22,401	22,059	21,824	65,352	21,784	50,679
Protein residues	4344	2896	6328	4344	3906	2793	2755	2704	8112	2704	6048
Nucleotide residues	0	0	140	0	105	0	0	0	0	0	135
Ligands	0	0	0	0	0	0	0	0	0	0	0
R.m.s. deviations											
Bond lengths (Å)	0.006	0.005	0.005	0.008	0.006	0.005	0.005	0.005	0.007	0.007	0.005
Bond angles (°)	1.175	1.157	1.033	1.215	0.969	0.991	1.048	1.033	1.148	1.147	0.969
Validation											
MolProbity score	1.76	1.87	1.87	1.99	1.31	1.47	1.70	1.72	1.73	1.70	1.20
Clashscore	7.55	9.17	8.83	11.53	5.64	7.88	7.48	6.64	7.12	6.64	4.21
Rotamers outliers (%)	0.00	0.63	1.08	0.03	0.24	0.53	0.78	0.60	0.72	0.68	0.71
Ramachandran plot											

SPŌZNAJ

PODPORA PRI UVAJANJU NAČEL ODPRTE ZNANOSTI V SLOVENIJI

Centralna tehniška knjižnica Univerze v Ljubljani
Central Technical Library at the University of Ljubljana
Trg republike 3, SI-1000 Ljubljana
Slovenija / *Slovenia*

