

크

▲□▶ ▲圖▶ ▲厘▶ ▲厘≯

Interpretation of structural models

CRC/RTG Spring School – X-Ray Crystallography

Alexander Minges Institute of Biochemical Plant Physiology



Where to get the slides

http://www.biochemplant.hhu.de

Will be uploaded today evening!



Software used

PyMOL https://pymol.org
 https://pymolwiki.org/index.php/Windows_Install
 Coot https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot



Where we left off before lunch

Diffraction experiments yield data in form of spot intensities



Where we left off before lunch

- Diffraction experiments yield data in form of spot intensities
- Intensity of each reflection is proportional to the squared amplitude of a corresponding structure factor



Where we left off before lunch

- Diffraction experiments yield data in form of spot intensities
- Intensity of each reflection is proportional to the squared amplitude of a corresponding structure factor
- Structure factors contain information of both the phase and amplitude of a wave



Where we left off before lunch

- Diffraction experiments yield data in form of spot intensities
- Intensity of each reflection is proportional to the squared amplitude of a corresponding structure factor
- Structure factors contain information of both the phase and amplitude of a wave
- Phase problem may be solved using various techniques (e.g. molecular replacement)



Where we left off before lunch

- Diffraction experiments yield data in form of spot intensities
- Intensity of each reflection is proportional to the squared amplitude of a corresponding structure factor
- Structure factors contain information of both the phase and amplitude of a wave
- Phase problem may be solved using various techniques (e.g. molecular replacement)

Structure factor amplitudes and phase information can be combined to reconstruct the electron density.





Figure: Scheme of X-Ray structure determination





Where to find structural data?





Where to find structural data?



https://www.rcsb.org





What is the PDB?

A database/collection of structural models...

- ...from various sources (e.g. X-Ray, NMR, Cryo-EM)...
- ...and associated experimental data.







What is the PDB?

A database/collection of structural models...

- ...from various sources (e.g. X-Ray, NMR, Cryo-EM)...
- ...and associated experimental data.







What is the PDB?

- A database/collection of structural models...
- ...from various sources (e.g. X-Ray, NMR, Cryo-EM)...
- ...and associated experimental data.







What is the PDB?

- A database/collection of structural models...
- ...from various sources (e.g. X-Ray, NMR, Cryo-EM)...
- ...and associated experimental data.









Figure: Growth of PDB (only X-Ray)

https://www.rcsb.org/stats



Classic PDB files are plain text files which are human-readable!

Header

Contains information about:

- authors/publication
- experimental procedure
- data collection
- refinement and model statistics
- biological assembly
- primary/secondary structure
- missing residues/atoms

...



Header

HEADER	HYDROLASE 17-JAN-16 5HMV									
TITLE	RE REFINEMENT OF 4MWK.									
COMPND	MOL_ID: 1;									
COMPND	2 MOLECULE: LYSOZYME C;									
COMPND	3 CHAIN: A;									
COMPND	4 SYNONYM: 1,4-BETA-N-ACETYLMURAMIDASE C,ALLERGEN GAL D IV;									
COMPND	5 EC: 3.2.1.17									
SOURCE	MOL_ID: 1;									
SOURCE	2 ORGANISM_SCIENTIFIC: GALLUS GALLUS;									
SOURCE	3 ORGANISM_COMMON: CHICKEN;									
SOURCE	4 ORGANISM_TAXID: 9031;									
SOURCE	5 TISSUE: EGG WHITE									
KEYWDS	STRUCTURAL DYNAMICS CISPLATIN HISTIDINE PROTEIN, HYDROLASE									
EXPDTA	X-RAY DIFFRACTION									
AUTHOR	J.R.HELLIWELL									
REVDAT	7 13-SEP-17 5HMV 1 REMARK									
REVDAT	6 17-AUG-16 5HMV 1									
REVDAT	5 10-AUG-16 5HMV 1 JRNL									



Body

ATOM protein HETATM small compounds, ions, solvent, ... CONNECT optional; bonds between atoms ANISOU optional; anisotropic B-factors



Body								
ATOM	1	Ν	LYS	A	1	-3.371 -10.156 -8.720 1.00 15.86	N	
ATOM	2	CA	LYS	А	1	-2.483 -10.521 -9.827 1.00 16.67	С	
ATOM	3	С	LYS	А	1	-2.518 -12.032 -10.037 1.00 19.82	С	
ATOM	4	0	LYS	А	1	-2.506 -12.813 -9.073 1.00 15.15	0	
ATOM	5	СВ	LYS	А	1	-1.053 -10.037 -9.545 1.00 17.67	С	
ATOM	6	CG	LYS	А	1	0.017 -10.569 -10.498 1.00 18.43	С	
ATOM	7	CD	LYS	А	1	1.392 -10.047 -10.097 1.00 22.48	С	
ATOM	8	CE	LYS	А	1	2.517 -10.653 -10.939 1.00 22.32	С	
ATOM	9	NZ	LYS	А	1	2.338 -10.354 -12.381 1.00 30.88	Ν	
ATOM	10	Ν	VAL	А	2	-2.600 -12.449 -11.295 1.00 19.96	Ν	
ATOM	11	CA	VAL	А	2	-2.494 -13.860 -11.611 1.00 17.16	С	
ATOM	12	С	VAL	А	2	-1.067 -14.107 -12.085 1.00 21.51	С	
ATOM	13	0	VAL	А	2	-0.624 -13.516 -13.071 1.00 24.51	0	
ATOM	14	CB	VAL	А	2	-3.514 -14.293 -12.665 1.00 16.56	С	
ATOM	15	CG1	VAL	А	2	-3.329 -15.760 -13.007 1.00 18.51	С	
ATOM	16	CG2	VAL	А	2	-4.938 -14.050 -12.163 1.00 17.76	С	



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

Launch PyMOL



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

Launch PyMOL

2 Load both models:

fetch 5hmv fetch 4gcf



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

Launch PyMOL

2 Load both models:

fetch 5hmv fetch 4gcf

3 Align both models: cealign 5hmv, 4gcf



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

Launch PyMOL

2 Load both models:

fetch 5hmv fetch 4gcf

3 Align both models: cealign 5hmv, 4gcf

4 Visualize as cartoon:

hide everything show cartoon



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

Launch PyMOL

2 Load both models:

fetch 5hmv fetch 4gcf

3 Align both models: cealign 5hmv, 4gcf

4 Visualize as cartoon:

hide everything show cartoon

5 Show side chains:

set cartoon_side_chain_helper, 1
show sticks

Excercise - Spot the difference







Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

- Only subtile differences in side chain conformations
- Overall fold virtually identical



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

- Only subtile differences in side chain conformations
- Overall fold virtually identical

Both models are equally good!



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

- Only subtile differences in side chain conformations
- Overall fold virtually identical

Really?



Compare the given models of chicken egg lysozyme. Imagine you want to predict ligand interactions and need precise information on e.g. side chain conformations. Which model suits your needs best?

- Only subtile differences in side chain conformations
- Overall fold virtually identical

Resolution is the key!





Figure: Resolution of PDB entries (March 2018)

https://www.rcsb.org/stats/distribution_resolution





Figure: Resolution of PDB entries (March 2018)

https://www.rcsb.org/stats/distribution_resolution





Figure: Resolution of PDB entries (March 2018)

https://www.rcsb.org/stats/distribution_resolution



- Higher resolution: More observations (data-to-parameter ratio)
- More observations: More accurate atomic positions
- Sharper and cleaner ED maps
- Easier interpretation and more reliable model building

Electron density





Figure: Electron density map of PDB 5HMV (0.98 Å, 1.5 σ)
Electron density





Figure: Electron density map of PDB 4WO6 (2.0 Å, 1.5 σ)

Electron density





Figure: Electron density map of PDB 4GFC (3.5 Å, 1.5 σ)

Electron density



- Density around model
- Missing/excess model











We will complete a lysozyme structure that is missing a patch of 14 aa. Try to complete the structure. Do you recognize the secondary structure element these residues form?

Copy files to desktop: map.ccp4 and seed.pdb





- Copy files to desktop: map.ccp4 and seed.pdb
- 2 Launch Coot





- Copy files to desktop: map.ccp4 and seed.pdb
- 2 Launch Coot
- 3 Load model:
 - File: Open Coordinates (seed.pdb)





- Copy files to desktop: map.ccp4 and seed.pdb
- 2 Launch Coot
- 3 Load model:
 - File: Open Coordinates (seed.pdb)
- 4 Load map:
 - File: Open Map (map.ccp4)

Quality of X-Ray data



Quality indicators of structural models

- Resolution
- Completeness/Redundancy
- Data merging statistics (R_{merge/meas/pim})
- Fit of model and data (R/R_{free}, RSCC)
- B-factors (atomic displacement parameters)
- Correct chemistry/geometry

Bond/torsion angles (e.g. Ramachandran plot) Bond lengths (Binding) geometry of ligands Clashes

Quality of X-Ray data



Quality indicators of structural models

- Resolution
- Completeness/Redundancy
- Data merging statistics (R_{merge/meas/pim})
- Fit of model and data (R/R_{free}, RSCC)
- B-factors (atomic displacement parameters)
- Correct chemistry/geometry

Bond/torsion angles (e.g. Ramachandran plot) Bond lengths (Binding) geometry of ligands Clashes

Most of these can be found in Table 1!





Read the paper!

Important aspects of the structure in question may be explained there...

Understanding Table 1



034701-3 S. W. M. Tanley and J. R. Helliwell

Struct. Dyn. 1, 034701 (2014)

TABLE I. X-ray crystallographic data and protein model refinement statistics.

	Cisplatin	Cisplatin	Cisplatin	Carboplatin
Unit cell parameters (Å)/(deg)	a = 26.99	a = 26.77	A = 27.34	a = 26.96
	b=31.81	b=31.38	B = 32.13	b=31.79
	c = 34.07	c=33.86	C = 34.29	c = 34.05
	$\alpha = 89.08$	$\alpha = 88.90$	A = 88.04	$\alpha = 88.76$
	$\beta = 72.00$	$\beta = 72.31$	B = 71.17	$\beta = 71.99$
	$\gamma = 67.81$	$\gamma = 68.46$	$\gamma = 68.35$	$\gamma = 68.33$
PDB id's	4mwk ^a	4mwm ^a	4mwn ^a	4oxe ^a
Data collection temperature (K)	150	200	295	200
Crystal size (mm)	0.6	0.3	0.25	0.2
Total absorbed X-ray dose (MGy) ^b	0.31	0.37	0.48	0.31
Crystal growth time	5 weeks	8 days	6 weeks	11 days
Observed reflections	202732	118846	112029	156662
Unique reflections	51605	37288	19160	35817
Resolution (Å)	29.28-0.98 (1.02-0.98)	32.09-1.12 (1.15-1.12)	32.31–1.42 (1.51–1.42)	32.21–1.13 (1.16–1.13)
Completeness (%)	90.6 (51.7)	95.1 (84.6)	99.2 (98.4)	94.1 (77.0)
Rmerge (%)	0.045 (0.209)	0.087 (0.179)	0.145 (0.554)	0.086 (0.310)

Understanding Table 1



Mean I/sig(I)	15.6 (3.2) ^c	5.9 (2.1)	6.8 (1.1) ^d	7.2 (2.0)
Redundancy	3.5 (0.6)	2.2 (0.9)	5.7 (2.0)	2.9 (1.1)
Cruickshank diffraction precision index (Å) for coordinate error	0.022	0.037	0.084	0.046
Number of protein atoms	1007	998	992	998
Average B factor (\AA^2) for protein atoms	7.2	11.4	14.7	9.4
Number of water molecules	142	94	41	98
Average B factor (Å ²) for water molecules	16.4	19.7	22.6	30.2
Number of cisplatin/carboplatin atoms	15	11	4	13
Average B factor $(Å^2)$ for cisplatin and carboplatin atoms	33.5	38.8	27.1	36.3
Number of other bound atoms	85	72	28	54
Average B factor (Å ²) for other bound atoms	18.9	35.4	29.3	32.0
R factor/R free (%)	11.7/14.5	14.7/18.7	20.8/23.5	17.7/22.1
R factor all	11.9	14.9	21.0	17.9
root mean square deviation bond lengths (Å)/ Angles (deg)	0.039/2.907	0.023/2.323	0.021/2.247	0.031/2.592
Ramachandran favoured	97.5	97.5	96.0	95.8
Ramachandran allowed	2.5	2.5	3.2	4.2
Ramachandran disallowed	0	0	0.8 ^e	0

Protein backbone angles





Figure: Protein backbone angles¹

¹© (i) Jane S. Richardson, adapted from *The Anatomy and Taxonomy of Protein Structure*

Protein backbone angles





Figure: Protein backbone angles¹ Figure: Ramachandran plot of PDB 5HMV ¹©(•) Jane S. Richardson, adapted from *The Anatomy and Taxonomy of Protein Structure*

Bond lengths



Table: Common bond lengths found in protein structures

Peptide bond	Bond length (Å)	Single bond	Bond length (Å)
$C_{\alpha} - C$	1.525 ± 0.026	C – C	1.540 ± 0.027
C – N	1.336 ± 0.023	C – N	1.489 ± 0.030
$N - C_{\alpha}$	1.459 ± 0.020	C – O	1.420 ± 0.020
C = O	1.229 ± 0.019	C – S	1.807 ± 0.026
$C_{lpha} - C_{eta}$	1.530 ± 0.020	S – S	2.033 ± 0.016

Bernhard Rupp (2010) Biomolecular Crystallography p. 629

Secondary structure elements





Figure: α -helix with backbone (black) and sidechains (white).

Jane S. Richardson (2007) The Anatomy and Taxonomy of Protein Structure

Secondary structure elements





Figure: Parallel β -sheet with backbone (black) and sidechains (white).

Jane S. Richardson (2007) The Anatomy and Taxonomy of Protein Structure

Secondary structure elements





Figure: Antiparallel β -sheet in top (a) and side view (b).

(c) Jane S. Richardson, adapted from The Anatomy and Taxonomy of Protein Structure





Definition

B-factors describe the uncertainty in atomic positions.

- For rigid parts usually below 20 Å²
- Dependent on the resolution (data-to-parameter ratio) different models are used:
 - isotropic model (low resolution)
 - isotropic + TLS (low to medium resolution)
 - anisotropic (high resolution)
- Effect in practice hard to distinguish from occupancy





- Termini and loops naturally tend to have higher B-factors
- May hint to functional protein movements
- Atoms not supported by enough density will have extremely high B-factors

A word of warning

Don't take the exact position of a certain feature (e.g. side chain) for granted! Always check the corresponding B-factors!







Figure: PDB 5HMV with B-factors represented by color and thickness.







Figure: Example of isotropic (left) and anisotropic (right) B-factors.

https://www.phenix-online.org/documentation/dictionary.html

B-factors





Figure: Dynamics of pyruvate phosphate dikinase (PPDK).

Minges et al. (2017) doi:10.1038/srep45389

B-factors





Figure: B-factors are related to protein dynamics (PDB 5JVN).

Minges et al. (2017) doi:10.1038/srep45389





Definition

Fraction of unit cells in which a specific atom occurs at a given position.

- Usually 1.0 (100 %) for protein backbone and most side chains (see alternate conformations)
- Lower values quite common for ligands
- In practice its effects are hard to tell apart from B-factor influence

Alternate conformations



- Usually 1.0 (100 %) for protein backbone and most side chains (see alternate conformations)
- Lower values quite common for ligands

Ligand geometry





Figure: Ligand bound to PDB 1FQH (retracted, contoured at 0.8 σ).

Ligand geometry



- Ligand geometry often neglected
- Unusual or chemically unlikely conformations
- Especially for "uncommon" ligands
- Use ligand validation tools in Coot!

Ligand geometry





Figure: Ligand bound to PDB 1FQH (retracted, contoured at 0.8 σ).

Contacts & Clashes





Figure: Contact visualization in Coot (PDB 5JVN).

Minges et al. (2017) doi:10.1038/srep45389

Contacts & Clashes





Figure: Visualization of severe clashes in Coot.

https://www.phenix-online.org/documentation/dictionary.html



- Validation reports for each structure
- Annual update of reports
- Comprehensive summary of quality indicators



C News X Hysoxyme) Save search ▲Download Per page: 10 ∨	🕯 View basket (0)
Entries Macromolecules Compounds Protein families	Outlite (deed)
Simury Restlemented of Amak. Hellweid IR Society Problement of Amak. Hellweid IR Society Problement Problement of Amak. Hellweid IR Society Problement Problement of Amak. Interveting composition: Problement of Amak. Society Problement Problement of Amak. Society Problement Problement of Amak. Society Problement Probl	X-ray diffraction 0.84 resolution Relaxed: B My 2016 PH model(dats
4ps0 Crystal structure of an inhibitor of vertilatrate lysozyme (PA3802) from Pseudomonas aeroginosa PAO(at 1.25 A resolution Soint Crefer of Structural Generics (C24) Rource organism: Pseudomonas aeroginata PAO(Assembly composition; protein why structure Image: Add to basket 4. Download files	X-ray diffraction 1-234 records Rodel generation Pit model/data

https://www.pdbe.org



C Ruce X lysozyme		🛱 View basket (0)
Entries Macromolecules Compounds Protein families		(
Simure Resettement of drawk. Nettwell 3R Second synchronic to drawk. Second synchronic to draw by the transmission of the synchronic transmission of the synchronic transmission. Second synchronic to draw by the transmission of the synchronic transmission. Financial Componentiation of the synchronic transmission. Second synchronic transmission. Second synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. Second Synchronic transmission. </th <th>X-ray diffraction 0.98A resolution Released: 18 Hay 2 Model geometry Pit model/data</th> <th></th>	X-ray diffraction 0.98A resolution Released: 18 Hay 2 Model geometry Pit model/data	
4965 Crystal structure of at inhibitor of unitabilitier (suprame (PA300)) from Perudomonas annuginosa PAOI at 12.5 resolution Joint Conter for Structure Germinal (SCG) To te publication Source equipains: Perudomonas areograpains Assembly compositions protein only structure é Add to basket & Download files	X-ray diffraction 1.25Å resolution Released: 16 Apr 2 Model geometry Fit model/data	

https://www.pdbe.org



B POB Shmw structure summ: X +					
(←) → C [*] @ 0			回 ··· ③ ☆ ○	2, Suchen	🛓 🗈 🖬 🕈 🗖 O 🗉
He Tellinement OI 4mmkr. Source organismic Galaxis galitu Primary publicationic Di Comment on Structural dynamics of stipfatto binding to Hatidine I (9770) (2014).	n a protein" [Struct. Dyn. 1,	Model geometry Fit model/data		Show overview Citations Structure analysis Function and Biology Liborate and Biology	
1646 (97 10) 8046 (97 10) - 3 (37) (16) (2014) 94602: 27226979 ℃		L		Experiments and Validation View Downloads X Close	2
Function and Biology	Ligands and Env	ironments		PDB He PDB header Archive mmCDF File	
Reaction catalysed: Hydrolysis of (1-04)-beta-linkapss between H-acetylkurnatic acid and H-acetyl-0-glucosasine residens in a poptidelysen and between H-acetyl- 0-glucosasine residens in a chitedriches in chitedricher.	4 bound ligands: H ₀ C 0 N H ₀ C 0	-0 Pt *2	C1-	Updated swyctir file PDR (fie (pz) PDRMs_(ATOM lines) PDRMs_(roz datem) Structure Rattore	try
Biochemical function: 0 trydrolane activity, acting on glycoxyl bonds 15 Biological process: 0 defense response to bacterium 15 Celhidra component: 0 stockoam 15	3 x DHS 7 x No modified residues	Validation	2 x CL	EDS map EDS difference map Assembly composition XML Assembly 1 (mmCIF; gz) Assembly 1 (mmCIF; gz)]
begatates demantic e dynamic hypotiana, rankiy 22 (f) e dynamic hypotiana, rankiy 22, tyranyme (f) e dynamic c) e dynamic c) e dynamic c) e tyranyme-tike domain sportamity (f)	Motric Sfree Claincore Ransobandran outliers Sidechain outliers RSRZ outliers	Percentile Ra	nits Value 0.164	PASTA (Entry) SITE XM, Re with residue-level measurings Summary report (IPD) Potential apport (IPD) Percential apt((IPD) Percential apt((IPD) Validation data (IPM)	
Structure analysis Details	in In	u uomieoiotena di soe ruoren uomieoiotena soe ruoruse d'un	ie webrie		
Assembly composition: monomeric (preferred) Entry contents: 1 distinct polypeptide molecule	X-ray source: DF Spacegroup: P1	UKER AXS MICROSTAR			
Macromolecule:	Unit cell: a:	26.998Å b: 31.80 89.08° β: 72°	rÅ c: 34.072Å γ: 67.81°		
Chain: A Difference details >	R-values: R 0.	R seek 125 0.124	R free 0.147		

https://www.pdbe.org
Validation provided by the PDB



BPDB Shimvistructure summix × +						
(←) → ୯ @ 0	https://www.ebiacuk/pdbe/entry/pdb/thmv			🖸 🛆 :	achen	🛓 🗈 🖬 🕈 🗖 O 🗉
	The resements of armsk. Storace equalities (long plute Phases postbacketies OU252 (2014)). OU252 (2014)). Marcel 2014 (2014). Marcel 2014 (2014).	s a protéin" (Struct, Dyn.),	Model geometry Rt model/data		Shinty overview Calaions Structure analysis Function and Biology Ugands and Biology Ugands and Biology Ugands and Biology Ugands and Validation View Lowercloads X Close Top Rs	
	Function and Biology	Ligands and Envi	ironments		PDB header Archive mmCDF File	
	Reaction catalysed: Nydrolysis of (1-x4)-beta-linkapss between N-acetylmuranic acid and M-acetyl-D-glucosamine residues in a poptidoglycan and between N-acetyl- D-glucosamine residues in chitodextrins	4 bound ligands: $H_{9}C$ = 0 σ^{*}	-0 Pt +2	C1 ⁺	Vodelas Innur III POB Re (st) POBRe, (aTOM Innes) POBRe, (atom Innes) Structure Ratore	·
	Biochemical function: • hydrolase activity, acting on glycosyl bonds 🗹	3 x DMS 7 x No modified residues	NO3 7 x PT	2 x CL	EDS map Bit EDS difference map Bit Assembly composition XML	<u>600</u>
	Biological process: o defense response to bacterium Ef Cellular component: o cytoplasm Ef	Experiments and	Validation	🗟 Details	Assembly 1 (mmCIF; gz) Assembly 1 (storn only; mmCIF) RATA (fairs)	
	Sequence domains: o Chycoside hydrolase, family 22. [5] o Chycoside hydrolase, family 22, hysozyme [5] o Lynoxyme [5] o Chycoside hydrolase, family 22, conserved site [5] o Lynoxyme-like domain superfamily [5]	Notric Pfree Clashicore Banaoibandum estilers Sidechain untiers RSRZ estilers	Percentile Ranks	Value 0.164	Simmary report (PDF) Parentine Jak (PMG) Percentine Jak (PMG) Velideton data (PML)	
	Structure analysis	in In	u nemicolativo al Sor roota n neticolativo Sor rootus d'Ande sodo	94		
	Assembly composition: monomeric (preferred) Entry contents: 1 distinct polypeptide molecule	X-ray source: DR Spacegroup: P1	UKER AXS MICROSTAR			
	Macromolecule: E tysozyme C	Unit cell: a: 0: R-values: R	26.998Å b: 31.807Å 89.08° β: 72° R _{vente}	c: 34.072Å y: 67.81° R iree		
	Chain: A	0.1	125 0.124	0.147		

https://www.pdbe.org



Page 2

wwPDB X-ray Structure Validation Summary Report

5HMV

1 Overall quality at a glance i

The following experimental techniques were used to determine the structure: $X\text{-}RAY \ DIFFRACTION$

The reported resolution of this entry is 0.98 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.





The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain	
			.%	
1	A	128	93%	6% •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	DMS	A	201	-	-	-	Х
3	NO3	A	204	-	-	-	Х

Continued on next page...





Page 7

wwPDB X-ray Structure Validation Summary Report

5HMV

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Lysozyme C





Page 7

wwPDB X-ray Structure Validation Summary Report

5HMV

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Lysozyme C







Provides automated re-refinement pipeline of existing PDB entries





- Provides automated re-refinement pipeline of existing PDB entries
- Older entries may take advantage of new software developments





- Provides automated re-refinement pipeline of existing PDB entries
- Older entries may take advantage of new software developments
- Often leads to significant improvement of model quality and/or completeness

PDB-REDO



88 PDB Shmv structure summ × +					
(← → ♂ @	Intersection and the second s Second second sec		目 … 日 ☆	Q, Suchen	🛓 🗈 🖬 🕈 🖬 O 🗉
€ → C @	Interna Aducati photomorphotomy For retainments (of 0 470000 For retainments) (of 0 470000 For retainments) The second se	a protein* [Struct, Dyn.],	Noted geometry Court References on the second secon	Control of Source and Source	* K 1 * * 10 =
	Function and Biology	Ligands and Envir	ronments		
	Reaction catalysed:	4 bound ligands:		PDB_REDO	
	Hydrolysis of (1->4)-beta-linkapes between N-acetylburanci acid and N-acetyl-0-glucosamine residues in a perticlylycan and between N-acetyl- D-glucosamine residues in chitodextrins		-0 Pt *2 C1-	The sliders below show the change in model quality between original PDB entry and the PDB_REDO entry Nodel Geometry	
	Biochemical function: hydrolase activity, acting on glycosyl bonds 	3 x DMS 7 x b No modified residues	103 <u>7 x PT</u> <u>2 x CL</u>	Fit model/data	
	Biological process: o defense response to bacterium II Cellular component: o cytoplasm II	Experiments and	Validation Detai	ils	
	Sequence donahe: c cliposità lipoticala, family 22. d' c cliposità hydrollas, family 22. jysazyme d' t yrazyma C d' c cliposità hydrollasa, family 22. conserved alta d' c cliposità hydrollasa, family 22. conserved alta d' e lystaryme-tike donain superfamily d'	Motric Sitre Clashnore Ransobanden eithers Sidechain eithers RSR2 eithers	Percentile Ranits Value	• •	
	Structure analysis	l rec	mile slater in al 5 og ensem mile slater in 25 og ensem mile slater in 5 og ensem eftimle medetin		
	Assembly composition: monomeric (preferred)	X-ray source: DRL	JKER AKS MICROSTAR		
	Entry contents: 1 distinct polypeptide molecule	Spacegroup: P1			
	Macromolecule: Lysozyme C	Unit cell: a: 3 o: 6	26.998Å b: 31.807Å c: 34.072Å 99.08° β: 72° γ: 67.81°		
	Chain: A Diffecule details >	R-values: R 0.1:	R work R free 25 0.124 0.147		

https://www.pdbe.org



db redo.eu/db/Shmv × +							
→ C [*]	//pdb-redo.eu/db/5hmv		E	··· 🖸 🕁	Q, suchan	IN 🖬 🗢	+
PDBe	RCSB PDB	3D bionotes	Proteopedia				Г
Validation matrice from PDB-	REDO						ŀ
valuation metrics from FDD	PDB	PDB-REDO					Ľ
Crystallographic refinement							
R	0.1455	0.1232					
R-free	0.1639	0.1438					
Bond length RMS Z-score	0.931	0.800					
Bond angle RMS Z-score	1.013	0.967					
Model quality (raw scores percen	tiles)						
Ramachandran plot appearance	64	66					
Rotamer normality	73	68					
Coarse packing	N/A	N/A					
Fine packing	82	84					
Bump severity	93	93					
Hydrogen bond satisfaction	47	50					
WHAT_CHECK	Report	Report					
Model quality compared to re	solution neighbours						
R-free	Ramachandran plot	Rotamer quality					
2000 2000 2000 2000 2000 2000 2000 200							

https://pdb-redo.eu/db/5hmv





Assessment of model quality

Choose a model from the PDB (you may choose freely or use one of the examples below). Use the tools discussed in this course to explore the model and evaluate its quality and possible limitations.





Assessment of model quality

Choose a model from the PDB (you may choose freely or use one of the examples below). Use the tools discussed in this course to explore the model and evaluate its quality and possible limitations.

Example IDs: 4BHX, 5D1F, 1KBL, 5JVN, 2XKG





Assessment of model quality

Choose a model from the PDB (you may choose freely or use one of the examples below). Use the tools discussed in this course to explore the model and evaluate its quality and possible limitations.

Example IDs:

4BHX, 5D1F, 1KBL, 5JVN, 2XKG

- Ligands bound?
- Fit to electron density
- Geometry
- Flexibility

Concluding remarks



Use your common sense when working with structural modelsKeep in mind that all structures in the PDB are merely models



Use your common sense when working with structural modelsKeep in mind that all structures in the PDB are merely models

Thank you for your attention!

Concluding remarks





Maki Noro: http://www.popsci.com/blognetwork/tags/crystallography