wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 07:37 pm BST

PDB ID	:	1HC1
Title	:	CRYSTAL STRUCTURE OF HEXAMERIC HAEMOCYANIN FROM PANULIRUS INTERRUPTUS REFINED AT 3.2 ANGSTROMS RESOLUTION
Authors	:	Volbeda, A.; Hol, W.G.J.
Deposited on	:	1991-05-15
Resolution	:	3.20 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

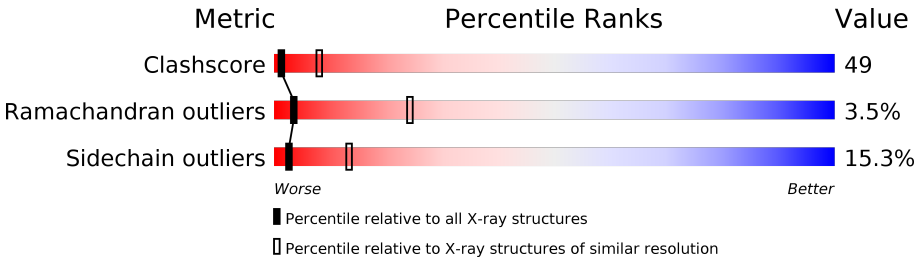
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)

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 respectively. 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	657	8% 29% 46% 14% •
1	B	657	7% 32% 42% 16% •
1	C	657	35% 49% 11% • •
1	D	657	31% 50% 13% • •
1	E	657	34% 47% 13% • •
1	F	657	32% 50% 13% • •

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 32166 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called ARTHROPODAN HEMOCYANIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	634	Total	C	N	O	S	0	0	0
			5173	3283	892	977	21			
1	B	634	Total	C	N	O	S	0	0	0
			5173	3283	892	977	21			
1	C	634	Total	C	N	O	S	0	0	0
			5173	3283	892	977	21			
1	D	634	Total	C	N	O	S	0	0	0
			5173	3283	892	977	21			
1	E	634	Total	C	N	O	S	0	0	0
			5173	3283	892	977	21			
1	F	634	Total	C	N	O	S	0	0	0
			5173	3283	892	977	21			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	32	ASP	GLU	CONFLICT	UNP P04254
A	163	PRO	GLN	CONFLICT	UNP P04254
A	458	ASN	LYS	CONFLICT	UNP P04254
A	514	SER	LYS	CONFLICT	UNP P04254
B	32	ASP	GLU	CONFLICT	UNP P04254
B	163	PRO	GLN	CONFLICT	UNP P04254
B	458	ASN	LYS	CONFLICT	UNP P04254
B	514	SER	LYS	CONFLICT	UNP P04254
C	32	ASP	GLU	CONFLICT	UNP P04254
C	163	PRO	GLN	CONFLICT	UNP P04254
C	458	ASN	LYS	CONFLICT	UNP P04254
C	514	SER	LYS	CONFLICT	UNP P04254
D	32	ASP	GLU	CONFLICT	UNP P04254
D	163	PRO	GLN	CONFLICT	UNP P04254
D	458	ASN	LYS	CONFLICT	UNP P04254
D	514	SER	LYS	CONFLICT	UNP P04254
E	32	ASP	GLU	CONFLICT	UNP P04254

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	163	PRO	GLN	CONFLICT	UNP P04254
E	458	ASN	LYS	CONFLICT	UNP P04254
E	514	SER	LYS	CONFLICT	UNP P04254
F	32	ASP	GLU	CONFLICT	UNP P04254
F	163	PRO	GLN	CONFLICT	UNP P04254
F	458	ASN	LYS	CONFLICT	UNP P04254
F	514	SER	LYS	CONFLICT	UNP P04254

- Molecule 2 is COPPER (II) ION (three-letter code: CU) (formula: Cu).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	2	Total Cu 2 2	0	0
2	E	2	Total Cu 2 2	0	0
2	B	2	Total Cu 2 2	0	0
2	C	2	Total Cu 2 2	0	0
2	A	2	Total Cu 2 2	0	0
2	F	2	Total Cu 2 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	186	Total O 186 186	0	0
3	B	186	Total O 186 186	0	0
3	C	186	Total O 186 186	0	0
3	D	186	Total O 186 186	0	0
3	E	186	Total O 186 186	0	0
3	F	186	Total O 186 186	0	0

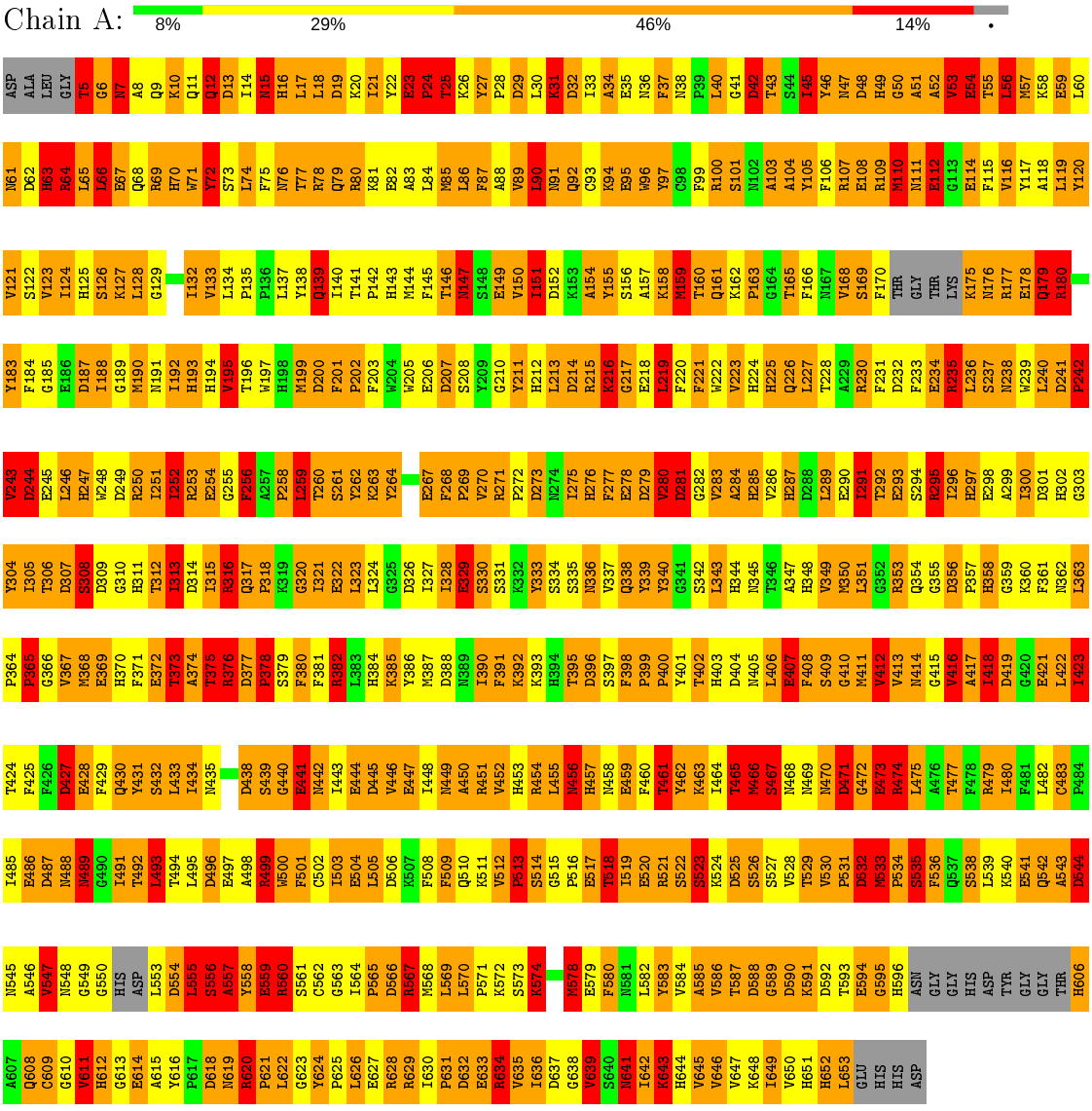
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

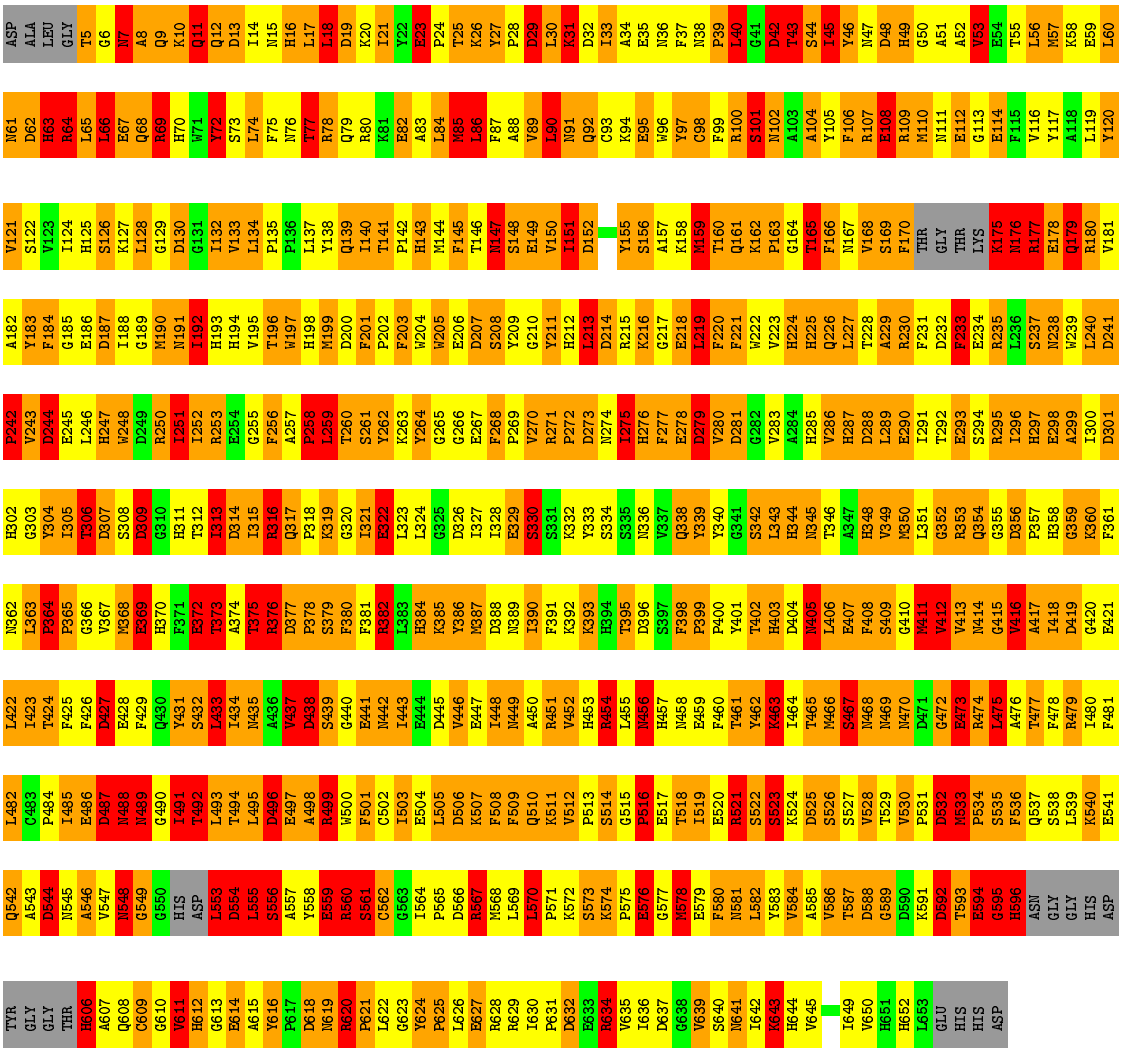
- Molecule 1: ARTHROPODAN HEMOCYANIN

Chain A:



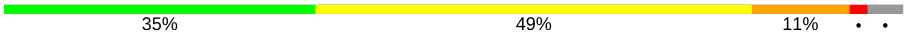
- Molecule 1: ARTHROPODAN HEMOCYANIN

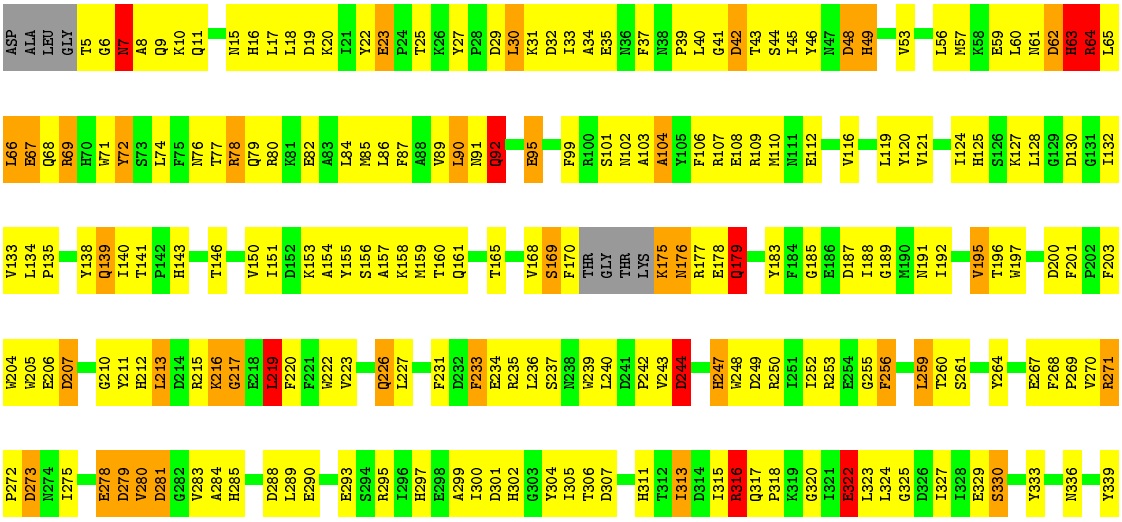
Chain B:

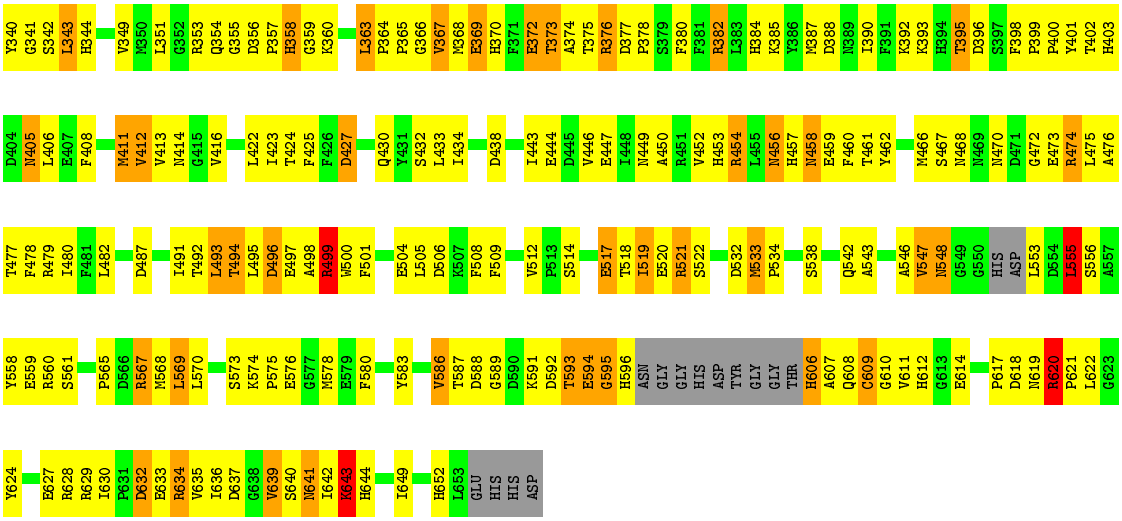


• Molecule 1: ARTHROPODAN HEMOCYANIN

Chain C:

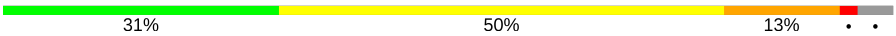


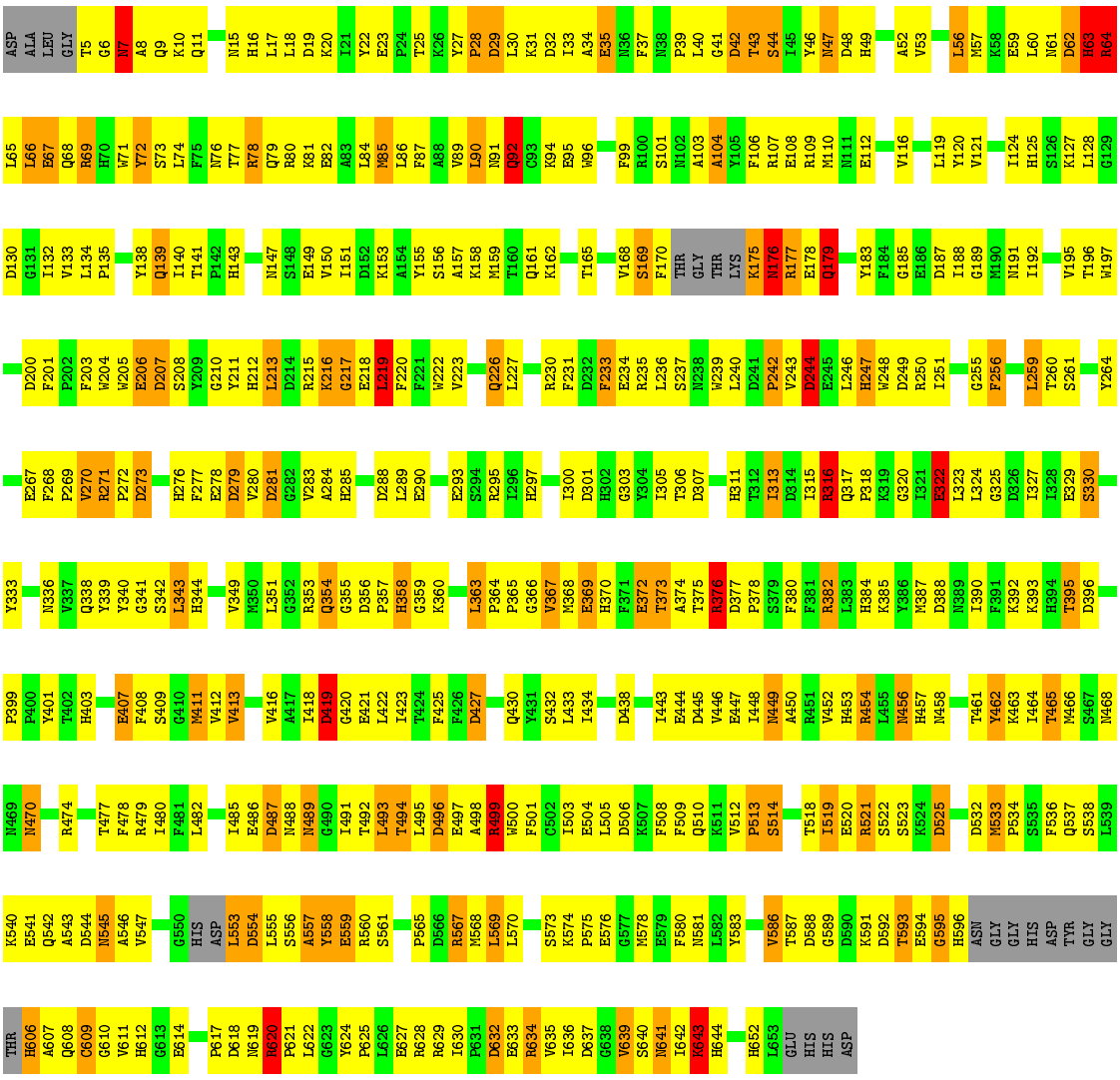




- Molecule 1: ARTHROPODAN HEMOCYANIN

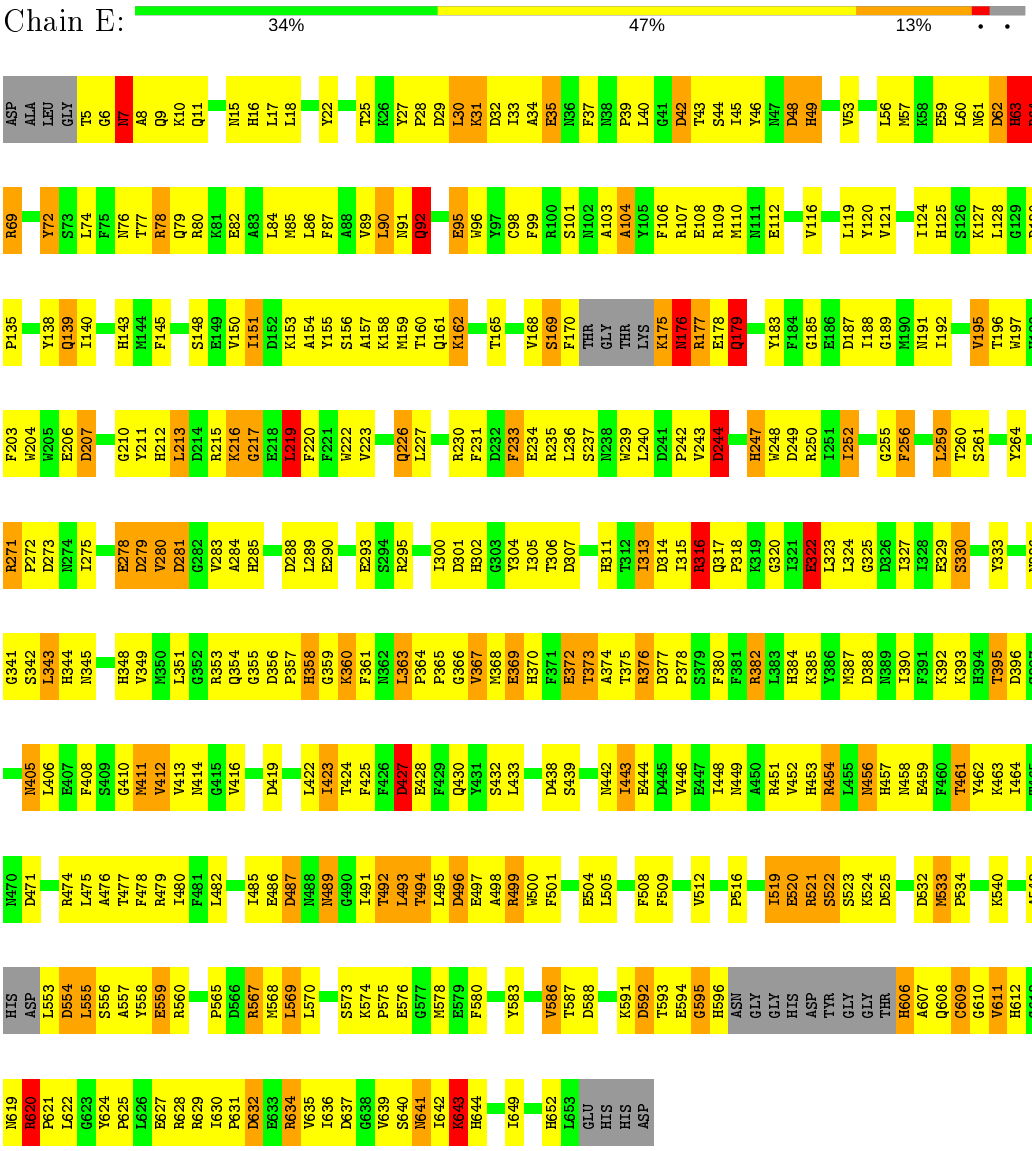
Chain D:





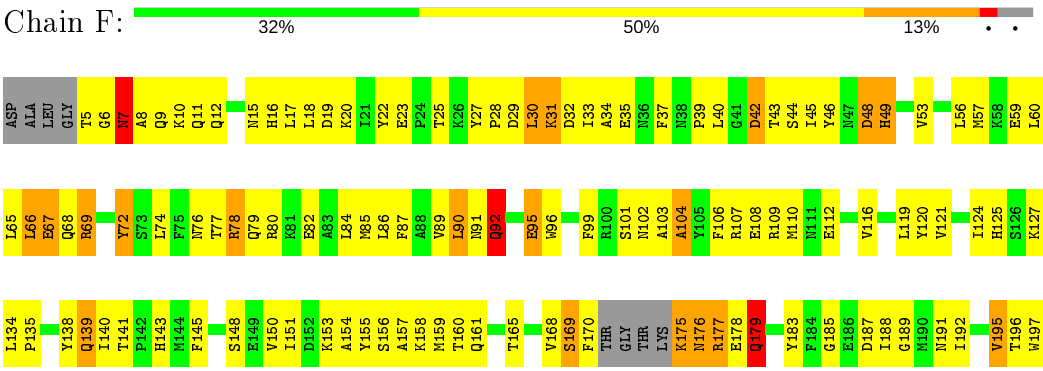
- Molecule 1: ARTHROPODAN HEMOCYANIN

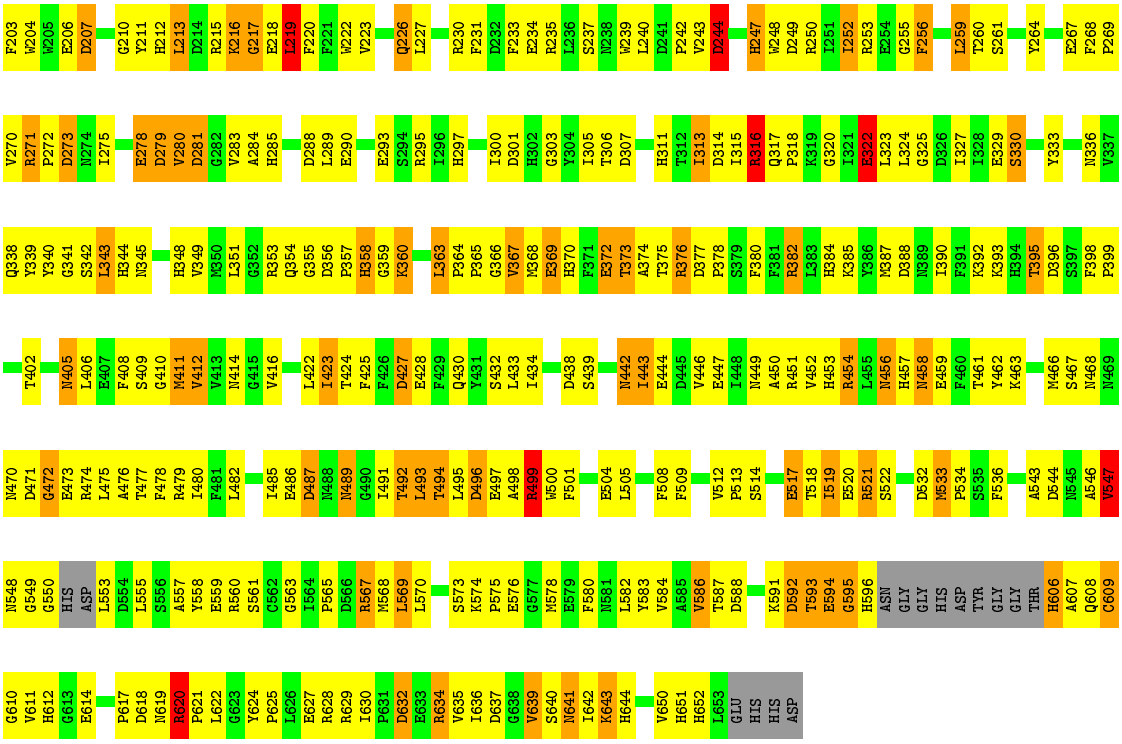
Chain E:



- Molecule 1: ARTHROPODAN HEMOCYANIN

Chain F:




F203 W204 W205 E206 D207 G210 Y211 H212 L213 D214 R215 K216 G217 E218 L219 F220 F221 W222 V223 Q226 L227 R230 F231 D232 F233 E234 R235 L236 S237 N238 W239 L240 D241 P242 V243 D244 H247 W248 D249 R250 I251 I252 R253 E254 G255 F256 L259 T260 S261 Y264 E267 F268 P269
V270 R271 P272 D273 N274 I275 E278 D279 V280 D281 G282 V283 A284 H285 D288 L289 E290 E293 S294 R295 I296 H297 I300 D301 H302 G303 Y304 I305 T306 D307 H311 T312 I313 D314 I315 R316 Q317 P318 K319 G320 I321 E322 L323 L324 G325 D326 I327 I328 E329 S330 Y333 N336 V337
Q338 Y339 Y340 G341 S342 L343 H344 N345 H348 V349 M350 L351 G352 R353 Q354 G355 D356 P357 H358 G359 K360 L363 P364 P365 G366 V367 M368 E369 H370 F371 E372 T373 A374 T375 R376 D377 P378 S379 F380 F381 R382 L383 H384 K385 Y386 M387 D388 N389 I390 F391 K392 K393 H394 T395 D396 S397 F398 P399
T402 N405 L406 E407 F408 S409 G410 M411 V412 V413 N414 G415 V416 L422 I423 T424 F425 F426 D427 E428 F429 Q430 Y431 S432 L433 I434 D438 S439 N442 I443 E444 D445 V446 E447 I448 N449 A450 R451 V452 H453 R454 L455 N456 H457 N458 E459 F460 T461 Y462 K463 M466 S467 N468 N469
N470 D471 G472 E473 R474 L475 A476 T477 F478 R479 I480 F481 L482 I485 E486 D487 N488 N489 G490 I491 T492 L493 T494 L495 D496 E497 A498 R499 W500 F501 E504 L505 F508 F509 V512 P513 S514 E517 T518 I519 E520 R521 S522 D532 M533 P534 S535 F536 A543 D544 N545 A546 V547
N548 G549 G550 HIS ASP L553 D554 L555 S556 A557 Y558 E559 R560 S561 C562 G563 I564 P565 D566 R567 M568 L569 L570 S573 K574 P575 E576 G577 M578 E579 F580 N581 L582 Y583 V584 A585 V586 T587 D588 K591 D592 T593 E594 G595 H596 ASN GLY GLY HIS ASP TYR GLY GLY THR H606 A607 Q608 C609
G610 V611 H612 G613 E614 P617 D618 N619 R620 P621 L622 G623 Y624 P625 L626 E627 R628 R629 I630 P631 D632 E633 R634 V635 I636 D637 G638 V639 S640 N641 I642 K643 H644 V650 H651 H652 L653 GLU HIS HIS ASP

4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	119.80Å 193.10Å 122.20Å 90.00° 118.10° 90.00°	Depositor
Resolution (Å)	8.00 – 3.20	Depositor
% Data completeness (in resolution range)	(Not available) (8.00-3.20)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ	Depositor
R, R_{free}	0.201 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	32166	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CU

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	A	1.41	11/5316 (0.2%)	4.13	1053/7205 (14.6%)
1	B	1.40	14/5316 (0.3%)	3.74	1016/7205 (14.1%)
1	C	0.75	0/5316	1.64	87/7205 (1.2%)
1	D	0.77	1/5316 (0.0%)	1.66	94/7205 (1.3%)
1	E	0.76	2/5316 (0.0%)	1.64	87/7205 (1.2%)
1	F	0.75	0/5316	1.65	94/7205 (1.3%)
All	All	1.02	28/31896 (0.1%)	2.64	2431/43230 (5.6%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	2
All	All	0	3

The worst 5 of 28 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	550	GLY	C-O	7.84	1.36	1.23
1	B	208	SER	CB-OG	7.79	1.52	1.42
1	A	441	GLU	CB-CG	7.42	1.66	1.52
1	A	526	SER	CB-OG	6.86	1.51	1.42
1	B	267	GLU	CD-OE2	-6.23	1.18	1.25

The worst 5 of 2431 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	64	ARG	NE-CZ-NH2	89.27	164.93	120.30
1	A	207	ASP	CB-CG-OD1	43.98	157.88	118.30
1	B	271	ARG	NE-CZ-NH1	43.24	141.92	120.30
1	A	273	ASP	CB-CG-OD1	40.24	154.52	118.30
1	B	273	ASP	CB-CG-OD1	39.23	153.61	118.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	295	ARG	Sidechain
1	B	177	ARG	Sidechain
1	B	521	ARG	Sidechain

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5173	0	4880	687	0
1	B	5173	0	4883	651	1
1	C	5173	0	4888	416	3
1	D	5173	0	4888	472	1
1	E	5173	0	4888	428	0
1	F	5173	0	4888	416	1
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
2	E	2	0	0	0	0
2	F	2	0	0	0	0
3	A	186	0	0	39	0
3	B	186	0	0	15	0
3	C	186	0	0	11	0
3	D	186	0	0	11	0
3	E	186	0	0	10	0
3	F	186	0	0	9	0
All	All	32166	0	29315	2986	3

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 49.

The worst 5 of 2986 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:161:GLN:OE1	1:B:443:ILE:HD13	1.28	1.29
1:A:422:LEU:CD2	1:A:570:LEU:HD21	1.66	1.23
1:A:316:ARG:HD3	3:A:829:HOH:O	1.41	1.19
1:A:165:THR:CG2	1:A:449:ASN:HB2	1.73	1.17
1:B:456:ASN:HD22	1:B:457:HIS:N	1.42	1.17

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:309:ASP:OD1	1:C:49:HIS:CD2[2_647]	1.85	0.35
1:C:594:GLU:OE1	1:F:471:ASP:CB[2_657]	2.13	0.07
1:C:474:ARG:NH2	1:D:41:GLY:O[2_656]	2.17	0.03

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	626/657 (95%)	501 (80%)	97 (16%)	28 (4%)	2	18
1	B	626/657 (95%)	506 (81%)	89 (14%)	31 (5%)	2	16
1	C	626/657 (95%)	517 (83%)	95 (15%)	14 (2%)	6	35
1	D	626/657 (95%)	516 (82%)	91 (14%)	19 (3%)	4	28
1	E	626/657 (95%)	512 (82%)	95 (15%)	19 (3%)	4	28
1	F	626/657 (95%)	503 (80%)	102 (16%)	21 (3%)	3	24
All	All	3756/3942 (95%)	3055 (81%)	569 (15%)	132 (4%)	3	24

5 of 132 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	42	ASP
1	A	147	ASN
1	A	176	ASN
1	A	441	GLU
1	A	471	ASP

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	564/580 (97%)	438 (78%)	126 (22%)	1 4
1	B	564/580 (97%)	450 (80%)	114 (20%)	1 6
1	C	564/580 (97%)	494 (88%)	70 (12%)	4 21
1	D	564/580 (97%)	489 (87%)	75 (13%)	4 18
1	E	564/580 (97%)	498 (88%)	66 (12%)	5 23
1	F	564/580 (97%)	498 (88%)	66 (12%)	5 23
All	All	3384/3480 (97%)	2867 (85%)	517 (15%)	2 13

5 of 517 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	596	HIS
1	C	461	THR
1	F	313	ILE
1	C	23	GLU
1	C	279	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 158 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	405	ASN
1	D	147	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	338	GLN
1	C	435	ASN
1	D	11	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 12 ligands modelled in this entry, 12 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.