

SGC

Case Studies in Molecular Replacement

Melanie Vollmar

Structural Genomics Consortium

University of Oxford



SGC Toronto



SGC Oxford



Karolinska
Institutet

SGC Stockholm

- Information about target protein
- Tools for model editing
- Examples

- Related PDB entries known from literature
 - Know as much about your target as possible
- Data base search based on sequence and/or structural homology/alignment
 - <http://toolkit.tuebingen.mpg.de/hhpred>
 - http://ekhidna.biocenter.helsinki.fi/dali_server
 - <http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl>

- Coot
 - Superpositioning of multiple models and eliminate non-aligning parts to create a core
- Sequence alignment
 - Sequence alignment between sequence of unknown structure and sequence of search model
 - <http://www.ebi.ac.uk/Tools/msa/clustalw2>
 - <http://probcons.stanford.edu/>
 - <http://mafft.cbrc.jp/alignment/software/>
 - <http://www.ebi.ac.uk/Tools/msa/tcoffee/>
- Chainsaw
 - Creating merged model based on target sequence and search model details
 - Creating of poly-Ala models
 - Within CCP4



- Intermediate resolution
 - Muscle specific human β -enolase (PDB-ID: 2XSX)
- High resolution
 - PH domain of human AKT3 kinase (PDB-ID: 2X18)
- Low homology
 - PH domain of human Actin-binding protein anillin ANLN (PDB-ID: 2Y7B)
- Low resolution
 - Rotating transmembrane ring of F_0F_1 ATPsynthase from spinach chloroplasts (PDB-ID: 2W5J)
- NMR model
 - Thioredoxin domain of human TXNL2 (PDB-ID: 2WZ9)
- Twinning
 - Macro domain of human core histone H2A (PDB-ID: 2XD7)



- Muscle specific human β -enolase
- One of three isoforms of enolase
- Location: cytoplasm
- Function: involved in glycolysis
- Enzymatic reaction:
$$\text{2-phospho-D-glycerate} = \text{phosphoenolpyruvate} + \text{H}_2\text{O}$$
- Co-factor: Mg^{2+}
- Biologically active form: dimer
- Number of residues: 434
- Molecular weight: 46.9 kDa
- UniProt-ID: P13929



- Processing results (XDS, Scala)

	Overall	Inner Shell	Outer Shell
Low resolution limit	29.49	29.49	1.79
High resolution limit	1.70	5.38	1.70
Rmerge	0.093	0.036	0.716
Rmerge in top intensity bin	0.036	-	-
Rmeas (within I+/I-)	0.099	0.039	0.759
Rmeas (all I+ & I-)	0.099	0.039	0.759
Rpim (within I+/I-)	0.033	0.013	0.250
Rpim (all I+ & I-)	0.033	0.013	0.250
Fractional partial bias	0.000	0.000	0.000
Total number of observations	996168	32548	143502
Total number of unique	114102	3820	16176
Mean ((I)/sd(I))	15.7	39.1	3.1
Completeness	96.4	95.0	94.6
Multiplicity	8.7	8.5	8.9
Average unit cell	95.88	105.53	106.09
	90.0	90.0	90.0
Space group	P 2 2 2		
Maximum resolution	1.70 Å		

Matthews - Cell Content Analysis Initial parameters from /work/ENO3A/5-phase-d001-x0

Job title ENO3A_P222_XDS_proc.mtz

Calculate Matthews coefficient for protein only

☒ Read crystal parameters from MTZ file

MTZ file O3A-phase-d0C ENO3A_P222_XDS_proc.mtz

Space group P 2 2 2

Cell a 95.8780 b 105.5310 c 106.0910 alpha 90.0000 beta 90.0000 gamma 90.0000

☒ High resolution limit 1.700

Use molecular weight estimated from number of residues

Number of residues 450

Solvent content analysis

Cell volume: 1073439.500

Molecular weight estimated from number of residues: 450

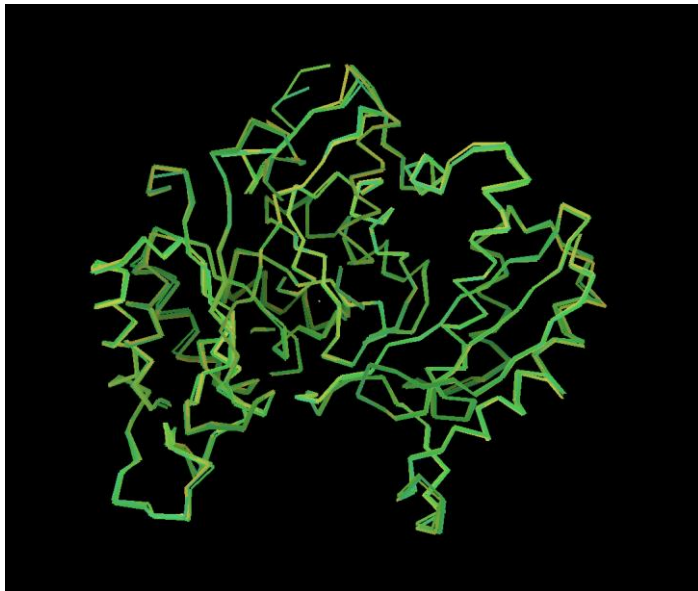
Nmol/asym	Matthews Coeff	%solvent	P(1.70)	P(tot)
1	5.96	76.61	0.86	0.81
2	2.65	53.62	0.89	0.95
3	1.77	30.43	0.10	0.04

Reset Run Now Close



Model preparation

- Multiple homologues structures
e.g. 1TE6 (82%), 2AKM (82%), 2AKZ (82%), 2PSN (83%), 3B97 (83%)
- Reduce model PDB files to one chain (e.g. chain A); delete water molecules and ligands
- Use Coot for superpositioning
- Sequence alignment between model and target
- Run Chainsaw with edited model and sequence alignment; reduces side chains to last common atom between model and target sequence



Chainsaw - create MR search model Initial parameters from /work/ENO3A/5-phase-d001-x0C

Job title 1TE6_A_SSM.pdb for model creation

Create search model using Chainsaw

Prune non-conserved residues to gamma atom

PDB in O3A-phase-d0C 1TE6_A_SSM.pdb Browse View

Input sequence alignment file in ALN/Clustalw format

Alignment in O3A-phase-d0C clustalw2-20100713-x002-1TE6_A.aln Browse View

PDB out O3A-phase-d0C 1TE6_A_SSM_chainsaw1.pdb Browse View

Run Save or Restore Close



Phasing

- Using Phaser
- Using search models as prepared above
- Using data as resulting from Scala

Phaser result

SpaceGroup of Solution: P 21 21 21

SINGLE solution

Solution written to PDB file:

/work/ENO3A/5-phase-d001-x002-mv/ENO3A-phase-d001_7.1.pdb

Solution written to MTZ file:

/work/ENO3A/5-phase-d001-x002-mv/ENO3A-phase-d001_7.1.mtz

Solution log-likelihood gain: 3623.52

Solution annotation (history):

RFZ=18.6 TFZ=28.0 PAK=0 LLG=739

Molecule 1

RFZ=15.7 TFZ=34.5 PAK=1 LLG=2364

Molecule 2

LLG=3624

Maximum Likelihood Molecular Replacement Initial parameters from /work/ENO3A/5-phase-

Job title ENO3A_P222_XDS_proc.mtz find 2 copies

Mode for molecular replacement automated search

Define data

MTZ in O3A-phase-d0C ENO3A_P222_XDS_proc.mtz

F F d001 SIGF SIGF_d001

Resolution range 106.091 A to 2.5 A

Space group read from mtz file 'P 2 2 2'

Run Phaser with all choices of alternative space group

Number of top solutions to output as PDB and MTZ files 100: MTZ files not output for all modes

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name ensemble1 Define ensemble via pdb file(s)

PDB #1 O3A-phase-d0C 1TE6_A-SSM_chainsaw1.pdb

Similarity of PDB #1 to the target structure sequence identity 0.82

PDB #2 O3A-phase-d0C 2AKM_A-SSM_chainsaw1.pdb

Similarity of PDB #2 to the target structure sequence identity 0.82

PDB #3 O3A-phase-d0C 2AKZ_A-SSM_chainsaw1.pdb

Similarity of PDB #3 to the target structure sequence identity 0.82

PDB #4 O3A-phase-d0C 2PSN_A-SSM_chainsaw1.pdb

Similarity of PDB #4 to the target structure sequence identity 0.83

PDB #5 O3A-phase-d0C 3B97_A-SSM_chainsaw1.pdb

Similarity of PDB #5 to the target structure sequence identity 0.83

Edit list Add superimposed PDB file to the ensemble

Edit list Add ensemble

Search details

Allow search with alternative ensembles (models) for a single component of the ASU off

Perform search using ensemble1 Number of copies to search for 2

Edit list Add another search

Final selection criterion for ROTATION search peaks percentage of top peak

Final selection criterion for TRANSLATION search peaks percentage of top peak

Packing criterion best packing solutions Max # clashes 10 Clash distance 3.0

Composition of the asymmetric unit

Total scattering determined by components in asymmetric unit

Component #1 protein sequence file Number in asymmetric unit 2

SEQ file O3A-phase-d0C ENO3A-x002.seq

Edit list Define another component

Run Save or Restore Close



Change space group

- Using Reindex from CCP4
- Input: Scala MTZ file

* Cell Dimensions : (obsolete - refer to dataset cell dimensions above)

95.8780 105.5310 106.0910 90.0000 90.0000 90.0000

* Resolution Range :

0.00009 0.34602 (106.091 - 1.700 A)

* Sort Order :

1 2 3 0 0

* Space group = 'P 2 2 2' (number 16)

MTZDMP before changing

* Cell Dimensions : (obsolete - refer to dataset cell dimensions above)

95.8780 105.5310 106.0910 90.0000 90.0000 90.0000

* Resolution Range :

0.00009 0.34602 (106.091 - 1.700 A)

* Sort Order :

0 0 0 0 0

* Space group = 'P 21 21 21' (number 19)

MTZDMP after changing



- Intermediate resolution
 - Muscle specific human β -enolase (PDB-ID: 2XSX)
- High resolution
 - PH domain of human AKT3 kinase (PDB-ID: 2X18)
- Low homology
 - PH domain of human Actin-binding protein anillin ANLN (PDB-ID: 2Y7B)
- Low resolution
 - Rotating transmembrane ring of F_0F_1 ATPsynthase from spinach chloroplasts (PDB-ID: 2W5J)
- NMR model
 - Thioredoxin domain of human TXNL2 (PDB-ID: 2WZ9)
- Twinning
 - Macro domain of human core histone H2A (PDB-ID: 2XD7)



- PH domain of human AKT3 kinase
- AKT3 = RAC-gamma serine/threonine-protein kinase
- Location: cytoplasm and membrane associated
- Function: regulating cell survival
- Enzymatic reaction (full length):
ATP + a protein = ADP + a phosphoprotein
- Co-factor (PH domain): -
- Biologically active form (PH domain): -
- Number of residues (PH domain): 103
- Molecular weight (PH domain): 12.3 kDa
- UniProt-ID: Q9Y243



- Processing results (HKL2000, Scalepack)

	Overall	Inner Shell	Outer Shell
Low resolution limit	50.00	50.00	1.64
High resolution limit	1.58	3.40	1.58
Chi**2	1.287	1.387	0.798
Rmerge	0.050	0.032	0.250
Total number of observations	126450	12875	11431
Total number of unique			
Mean ($\langle I \rangle / \text{sd}(I)$)	24.3	37.0	3.4
Completeness	95.8	97.5	86.5
Multiplicity	2.6	2.6	2.6
Average unit cell	62.46	62.39	71.40
	111.48	102.73	94.37
Space group	P 1		
Maximum resolution	1.58 Å		

Cell volume: 249998.031

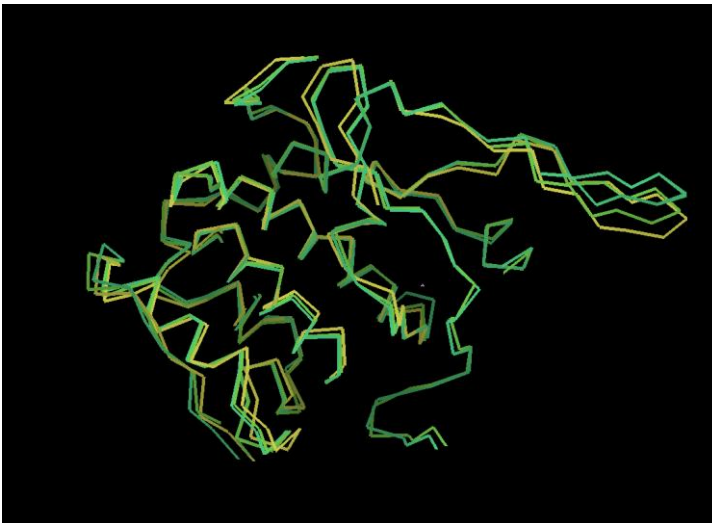
For estimated molecular weight 12760.

Nmol/asym Matthews Coeff %solvent P(1.43)
P(tot)

1	19.59	93.73	0.00	0.00
2	9.80	87.45	0.00	0.00
3	6.53	81.18	0.00	0.00
4	4.90	74.90	0.00	0.00
5	3.92	68.63	0.00	0.02
6	3.27	62.36	0.01	0.07
7	2.80	56.08	0.07	0.18
8	2.45	49.81	0.23	0.31
9	2.18	43.53	0.38	0.29
10	1.96	37.26	0.25	0.11
11	1.78	30.99	0.04	0.01
12	1.63	24.71	0.00	0.00
13	1.51	18.44	0.00	0.00
14	1.40	12.16	0.00	0.00
15	1.31	5.89	0.00	0.00

Model preparation

- Multiple homologues structures
e.g. 1MRY (84%), 1O6K (83%), 1GZ0 (86%), 3CQW (85%), 1UNP (82%)
- Reduce model PDB files to one chain (e.g. chain A); delete water molecules and ligands
- Use Coot for superpositioning
- Sequence alignment between model and target
- Run Chainsaw with edited model and sequence alignment; reduces side chains to last common atom between model and target sequence



Core of superpositioned search models of
full length homologues



PH domain as in 1UNP



Phasing

- Using Phaser
- One search model was picked randomly
- Data was further truncated and Rfree flag was added

Phaser result

SpaceGroup of Solution: P 1

There were 7 solutions

Pdb and Mtz files have been output for 1 of these solutions

Solution #1 written to PDB file: /work/AKT3A/8-phase-mv/AKT3_phase_12.1.pdb

Solution #1 written to MTZ file: /work/AKT3A/8-phase-mv/AKT3_phase_12.1.mtz

Solution #1 log-likelihood gain: 302.771

Solution #1 annotation (history):

RFZ=10.2 TFZ=100.0 PAK=0 LLG=89

Molecule 1

RFZ=9.0 TFZ=5.8 PAK=0 LLG=158

Molecule 2

LLG=303

Maximum Likelihood Molecular Replacement Initial parameters from /work/AKT3A/8-phase-1

Job title: phasing with 1UNP.pdb after integration with HKL2000

Mode for molecular replacement: automated search

Define data

MTZ in: AKT3A_phase, AKT3_HKL2000_1_truncate1.mtz

F: , F: , SIGF: , SIGF:

Resolution range: 64.017 A to 2.5 A

Space group read from mtz file: P1

Run Phaser with: the mtz space group

Number of top solutions to output as PDB and MTZ files: NB: MTZ files not output for all modes

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name: ensemble1, Define ensemble via: pdb file(s)

PDB #1: 1UNP-WAT.pdb

Similarity of PDB #1 to the target structure: sequence identity, 0.82

Edit list, Add superimposed PDB file to the ensemble

Edit list, Add ensemble

Search details

Allow search with alternative ensembles (models) for a single component of the ASU: off

Perform search using: ensemble1, Number of copies to search for: 2

Edit list, Add another search

Final selection criterion for ROTATION search peaks: percentage of top peak

Final selection criterion for TRANSLATION search peaks: percentage of top peak

Packing criterion: best packing solutions, Max # clashes: 10, Clash distance: 3.0

Composition of the asymmetric unit

Total scattering determined by: components in asymmetric unit

Component #1: protein, sequence file, Number in asymmetric unit: 8

SEQ file: AKT3A-x012-d001.seq

Edit list, Define another component

Define search sets (any already known rotations/translations)

Additional parameters

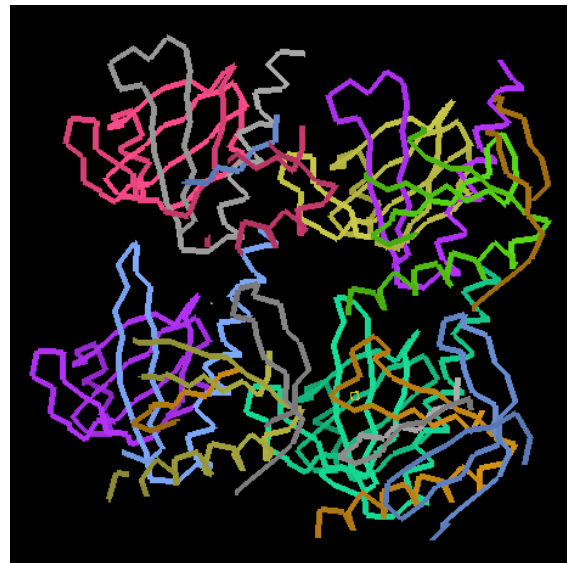
Run, Save or Restore, Close



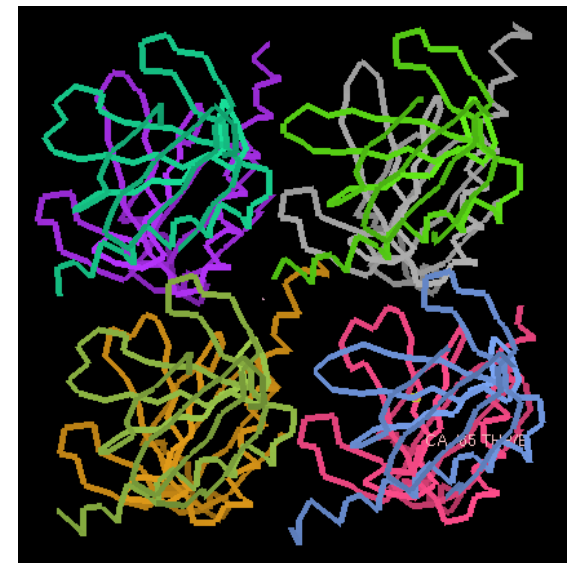
Phasing results and autobuilding



Initial phases



1st round of autobuild
with Buccaneer



2nd round of autobuild
with Buccaneer



- Intermediate resolution
 - Muscle specific human β -enolase (PDB-ID: 2XSX)
- High resolution
 - PH domain of human AKT3 kinase (PDB-ID: 2X18)
- Low homology
 - PH domain of human Actin-binding protein anillin ANLN (PDB-ID: 2Y7B)
- Low resolution
 - Rotating transmembrane ring of F_0F_1 ATPsynthase from spinach chloroplasts (PDB-ID: 2W5J)
- NMR model
 - Thioredoxin domain of human TXNL2 (PDB-ID: 2WZ9)
- Twinning
 - Macro domain of human core histone H2A (PDB-ID: 2XD7)



- PH domain of human Actin-binding protein anillin ANLN
- Actin-binding protein anillin
- Location: nucleus and cytoplasm
- Function: Required for cytokinesis; interacts with cytoskeleton
- Enzymatic reaction:
 -
- Co-factor: -
- Biologically active form: -
- Number of residues (PH domain): 125
- Molecular weight (PH domain): 14.8 kDa
- UniProt-ID: Q9NQW6



- Processing results (XDS, Scala)

	Overall	Inner Shell	Outer Shell
Low resolution limit	18.94	18.94	2.0
High resolution limit	1.90	6.01	1.9
Rmerge	0.136	0.037	0.962
Rmerge in top intensity bin	0.037	-	-
Rmeas (within I+/I-)	0.153	0.043	1.077
Rmeas (all I+ & I-)	0.153	0.043	1.077
Rpim (within I+/I-)	0.069	0.021	0.478
Rpim (all I+ & I-)	0.069	0.021	0.478
Fractional partial bias	0.000	0.000	0.000
Total number of observations	50794	1595	7380
Total number of unique	10888	391	1550
Mean ((I)/sd(I))	9.2	24.0	2.0
Completeness	99.6	96.3	99.8
Multiplicity	4.7	4.1	4.8
Average unit cell	34.30	59.42	64.72
	90.0	90.0	90.0
Space group	P 21 21 21		
Maximum resolution	1.90 Å		

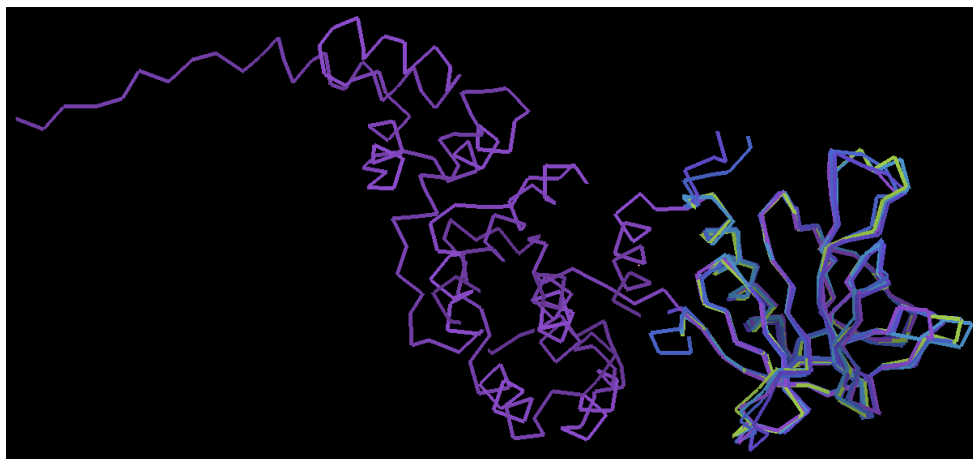
Cell volume: 132210.984

For estimated molecular weight 16313.
 Nmol/asym Matthews Coeff %solvent P(1.94)
 P(tot)

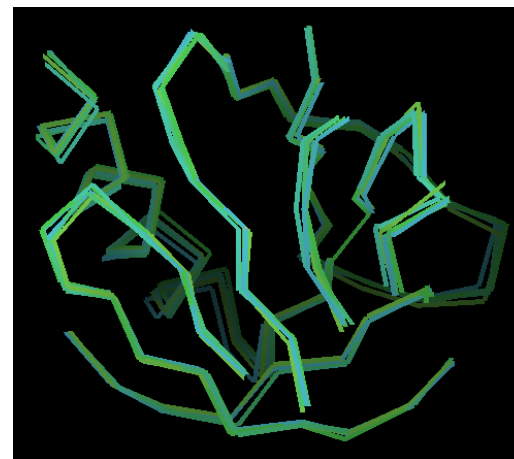
1	2.03	39.33	1.00	1.00
---	------	-------	------	------

Model preparation

- Single homologues structure in multiple conformations
e.g. 1U29 (20%), 1FHW (20%), 1U27 (20%), 1U2B (18%), 2R0D (15%)
- Reduce model PDB files to one chain (e.g. chain A); delete water molecules and ligands
- Use Coot to eliminate flexible protein parts
- Sequence alignment between model and target
- Run Chainsaw with edited model and sequence alignment; reduces side chains to last common atom between model and target sequence



Superpositioned search models



Core of superpositioned search models



Phasing

- Using Phaser
- Using search models as prepared above
- Using data as resulting from Scala

Phaser result

SpaceGroup of Solution: P 21 21 21

SINGLE solution

Solution written to PDB file:

/work/ANLNA/9-phase-d001-x062-mv/ANLNA-phase-x062_8.1.pdb

Solution written to MTZ file:

/work/ANLNA/9-phase-d001-x062-mv/ANLNA-phase-x062_8.1.mtz

Solution log-likelihood gain: 21.6602

Solution annotation (history):

RFZ=3.5 TFZ=6.2 PAK=0 LLG=22

Molecule 1

LLG=22

Maximum Likelihood Molecular Replacement Initial parameters from /work/ANLNA/9-phase-

Job title phase ANLNA_d002_x062_XDS.mtz based on 1FHW 1U27 1U29 1U2B 2R0D

Mode for molecular replacement automated search

Define data

MTZ in LNA-phase-x0C ANLNA-d002-x062_XDS.mtz

F 4001 SIGF d001

Resolution range 43.780 A to 2.5 A

Space group read from mtz file 'P 21 21 21'

Phaser with all choices of alternative space group

Number of top solutions to output as PDB and MTZ files NB: MTZ files not output for all modes

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name ensemble1 Define ensemble via pdb file(s)

PDB #1 LNA-phase-x0C 1FHW_A-SSM-frag_chainsaw1.pdb

Similarity of PDB #1 to the target structure sequence identity 0.2

PDB #2 LNA-phase-x0C 1U27_A-SSM-frag_chainsaw1.pdb

Similarity of PDB #2 to the target structure sequence identity 0.2

PDB #3 LNA-phase-x0C 1U29_A-SSM-frag_chainsaw1.pdb

Similarity of PDB #3 to the target structure sequence identity 0.2

PDB #4 LNA-phase-x0C 1U2B_A-SSM-frag_chainsaw1.pdb

Similarity of PDB #4 to the target structure sequence identity 0.18

PDB #5 LNA-phase-x0C 2R0D_A-SSM-frag_chainsaw1.pdb

Similarity of PDB #5 to the target structure sequence identity 0.15

Edit list Add superimposed PDB file to the ensemble

Edit list Add ensemble

Search details

Allow search with alternative ensembles (models) for a single component of the ASU off

Perform search using ensemble1 Number of copies to search for 1

Edit list Add another search

Final selection criterion for ROTATION search peaks percentage of top peak

Final selection criterion for TRANSLATION search peaks percentage of top peak

Packing criterion best packing solutions Max # clashes 10 Clash distance 3.0

Composition of the asymmetric unit

Total scattering determined by components in asymmetric unit

Component #1 protein sequence file Number in asymmetric unit 1

SEQ file LNA-phase-x0C ANLNA-x062.seq

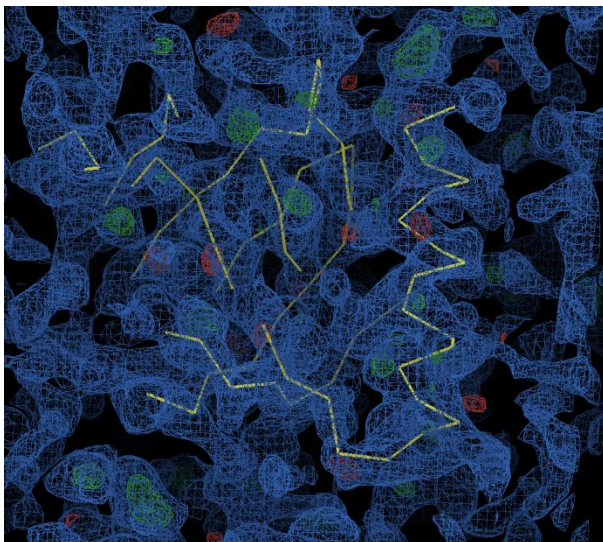
Edit list Define another component

Define search sets (any already known rotations/translations)

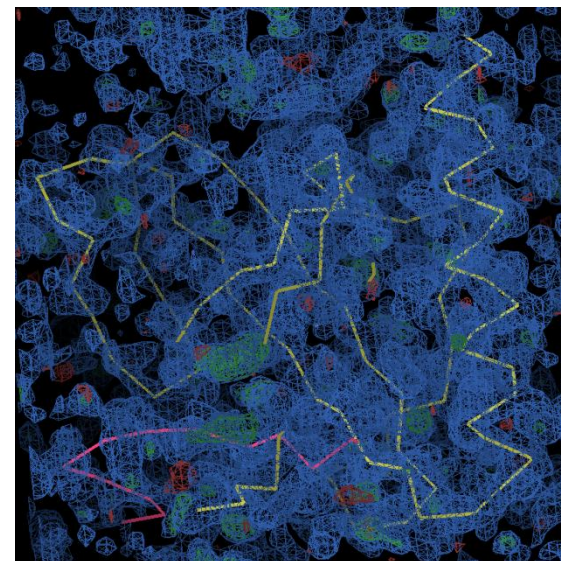
Additional parameters

Run Save or Restore Close

Results



After phasing



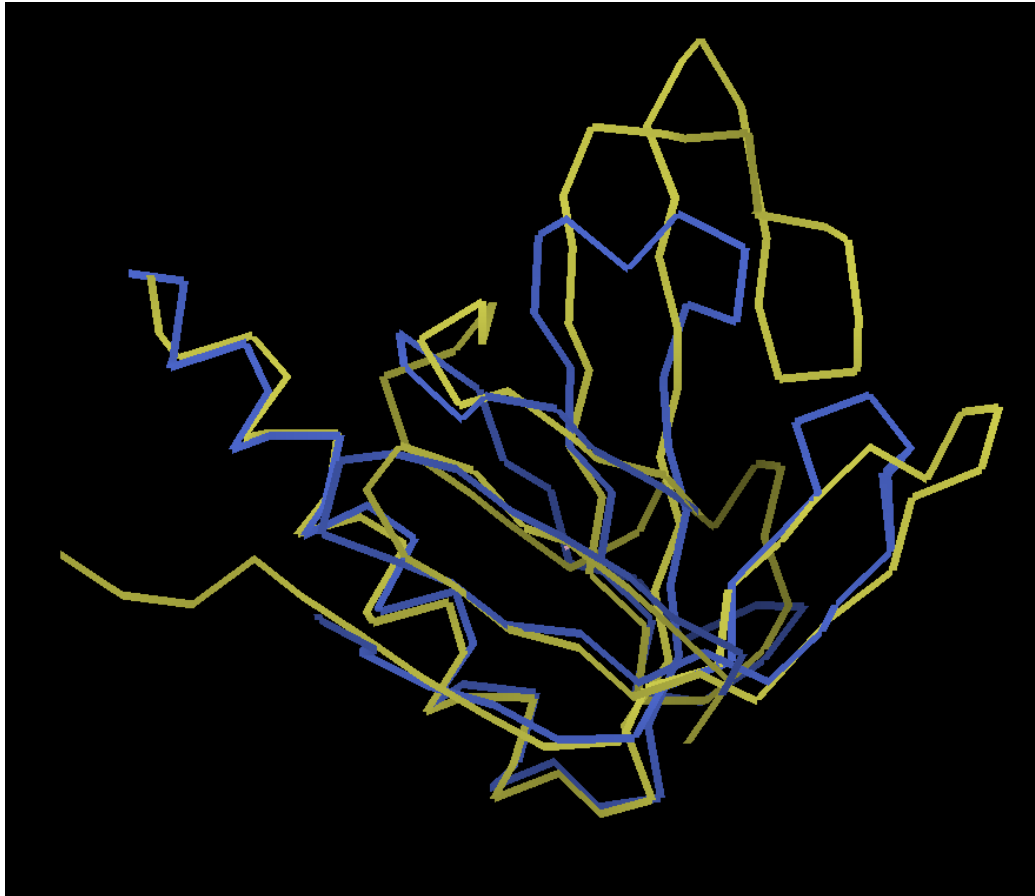
Some rounds of autobuilding and manual editing

	Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL
rmsCHIRAL										
\$\$										
0	0.5281	0.5462	0.344	51639.	2595.4	0.0310	1.084	1.921	1.070	0.111
1	0.4966	0.5533	0.358	51086.	2577.5	0.0345	1.563	1.613	0.803	0.119
2	0.4804	0.5389	0.397	50883.	2571.6	0.0236	1.037	1.470	0.678	0.096
3	0.4740	0.5203	0.421	50496.	2553.0	0.0131	0.595	1.331	0.608	0.096
4	0.4649	0.5133	0.448	50599.	2561.3	0.0125	0.561	1.402	0.642	0.098
5	0.4586	0.5102	0.470	50424.	2555.4	0.0129	0.578	1.495	0.681	0.102
6	0.4537	0.5081	0.484	50358.	2554.7	0.0136	0.617	1.580	0.732	0.104
7	0.4508	0.5043	0.489	50145.	2545.4	0.0146	0.657	1.645	0.770	0.109
8	0.4479	0.5024	0.489	50093.	2544.0	0.0139	0.624	1.697	0.775	0.111
9	0.4462	0.4999	0.493	50051.	2542.9	0.0138	0.616	1.713	0.754	0.111
10	0.4441	0.5010	0.496	50169.	2549.7	0.0138	0.622	1.761	0.765	0.112

	Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL
rmsCHIRAL										
\$\$										
0	0.3663	0.3868	0.739	33811.	2638.0	0.0273	1.694	2.393	1.223	0.121
1	0.3158	0.3571	0.757	31368.	2595.3	0.0251	1.094	2.058	0.985	0.129
2	0.2949	0.3439	0.761	29929.	2567.2	0.0225	0.975	2.010	0.954	0.132
3	0.2876	0.3407	0.761	29487.	2557.6	0.0226	0.962	2.019	0.960	0.131
4	0.2853	0.3420	0.760	29323.	2554.6	0.0224	0.952	2.020	0.961	0.132
5	0.2844	0.3450	0.759	29255.	2553.4	0.0225	0.948	2.021	0.961	0.133
6	0.2842	0.3469	0.758	29222.	2553.4	0.0225	0.946	2.025	0.962	0.133
7	0.2840	0.3483	0.758	29200.	2553.8	0.0226	0.950	2.030	0.964	0.133
8	0.2837	0.3481	0.758	29188.	2554.1	0.0227	0.950	2.038	0.967	0.134
9	0.2833	0.3484	0.758	29174.	2554.4	0.0228	0.958	2.044	0.970	0.134
10	0.2830	0.3481	0.758	29174.	2554.8	0.0229	0.965	2.050	0.973	0.134

Comparison of PH domains from AKT3 and ANLN

- Although both identified as PH domains differences are big enough to show low homology for MR
- Yellow → ANLN
- Blue → AKT3





- Intermediate resolution
 - Muscle specific human β -enolase (PDB-ID: 2XSX)
- High resolution
 - PH domain of human AKT3 kinase (PDB-ID: 2X18)
- Low homology
 - PH domain of human Actin-binding protein anillin ANLN (PDB-ID: 2Y7B)
- Low resolution
 - Rotating transmembrane ring of F_0F_1 ATPsynthase from spinach chloroplasts (PDB-ID: 2W5J)
- NMR model
 - Thioredoxin domain of human TXNL2 (PDB-ID: 2WZ9)
- Twinning
 - Macro domain of human core histone H2A (PDB-ID: 2XD7)



- Rotating transmembrane ring of chloroplast F_0F_1 ATP synthase from spinach chloroplasts
- F-type ATPase subunit c
- Location: chloroplast thylacoid membrane
- Function: transport of protons across the membrane and thereby fuelling ATP production
- Enzymatic reaction:
 -
- Co-factor: -
- Biologically active form: 14mer
- Number of residues (monomer): 81
- Molecular weight (14mer): 112 kDa
- UniProt-ID: P69447



• Processing results (Mosflm, Scala)

	Overall	Inner Shell	Outer Shell
Low resolution limit	28.28	28.28	4.0
High resolution limit	3.80	12.02	3.8
Rmerge	0.085	0.065	0.201
Rmeas (within I+/-)	0.081	0.039	0.248
Rmeas (all I+ & I-)	0.081	0.039	0.248
Total number of observations	132804		
Total number of unique	11213		
Mean ($\langle I \rangle / \text{sd}(I)$)	10.4	18.8	4.2
Completeness	92.2	71.6	92.2
Multiplicity	2.6	2.4	2.5
Average unit cell	128.60	89.99	124.89
	90.0	104.69	90.0
Space group	C2		
Maximum resolution	3.80 Å		

Cell volume: 1398161.375

For estimated molecular weight 9000.

Nmol/asyM Matthews Coeff %solvent P(3.80) P(tot)

1	38.84	96.83	0.00	0.00
2	19.42	93.67	0.00	0.00
3	12.95	90.50	0.00	0.00
4	9.71	87.34	0.00	0.00
5	7.77	84.17	0.00	0.00
6	6.47	81.01	0.00	0.00
7	5.55	77.84	0.00	0.00
8	4.85	74.68	0.00	0.00
9	4.32	71.51	0.01	0.00
10	3.88	68.35	0.01	0.01
11	3.53	65.18	0.03	0.02
12	3.24	62.02	0.05	0.04
13	2.99	58.85	0.08	0.06
14	2.77	55.69	0.11	0.09
15	2.59	52.52	0.14	0.13
16	2.43	49.36	0.16	0.16
17	2.28	46.19	0.15	0.17
18	2.16	43.03	0.12	0.14
19	2.04	39.86	0.07	0.10
20	1.94	36.70	0.03	0.05
21	1.85	33.53	0.01	0.02
22	1.77	30.37	0.00	0.00
23	1.69	27.20	0.00	0.00
24	1.62	24.04	0.00	0.00
25	1.55	20.87	0.00	0.00
26	1.49	17.71	0.00	0.00
27	1.44	14.54	0.00	0.00
28	1.39	11.38	0.00	0.00
29	1.34	8.21	0.00	0.00
30	1.29	5.05	0.00	0.00
31	1.25	1.88	0.00	0.00

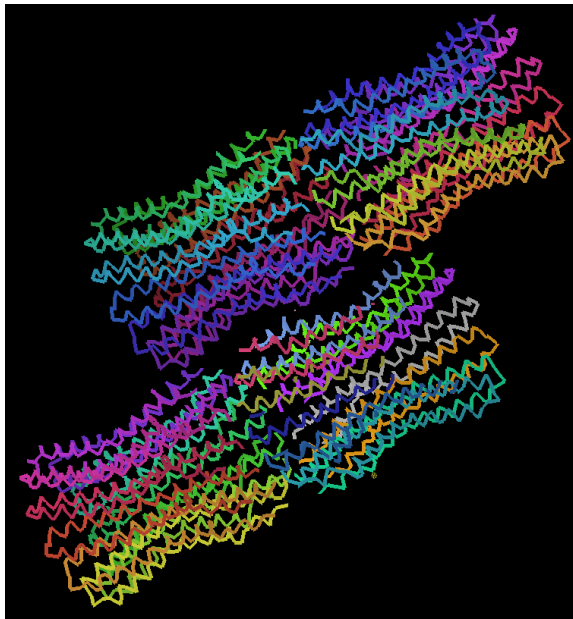
Note:
There is no
preferred
Matthews
coefficient

But:
Number of
14
monomers
known from
biochemical
experiments
and electron
microscopy
results

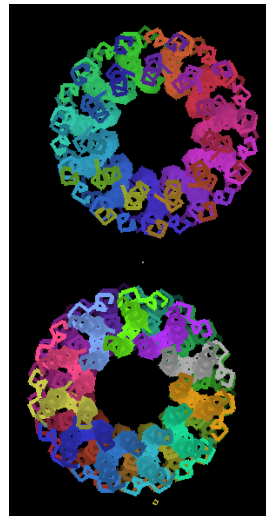


Model preparation

- Multiple homologues structures
e.g. 1YCE (56%)
- Reduce model PDB files to one chain (e.g. chain A); delete water molecules and ligands
- Sequence alignment between model and target
- Run Chainsaw with edited model and sequence alignment; reduces side chains to last common atom between model and target sequence



Closest homologue

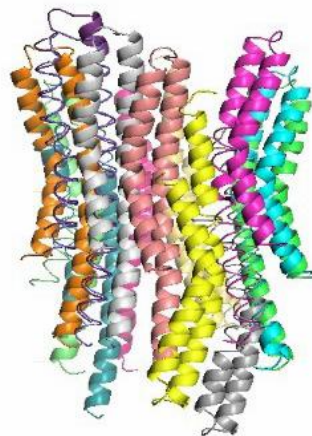
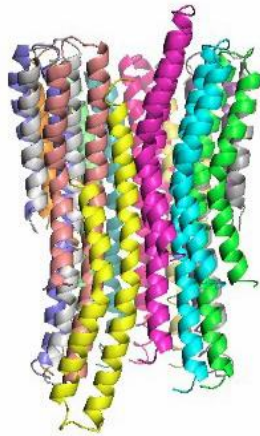


Isolated single chain



Phasing

- Using Phaser and single chain as search model
- Searching for 14 chains

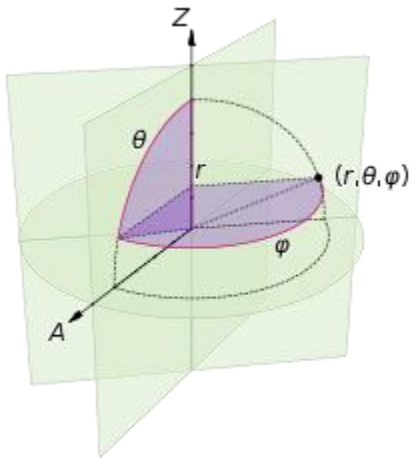




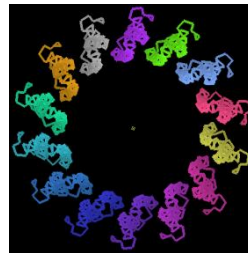
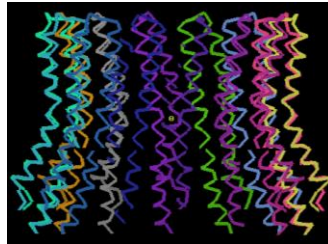
Phasing

- Molrep and scripting

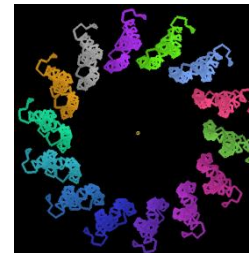
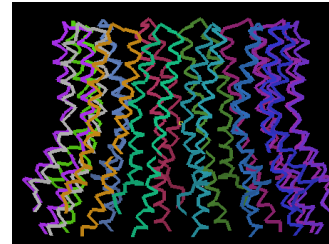
Step 1



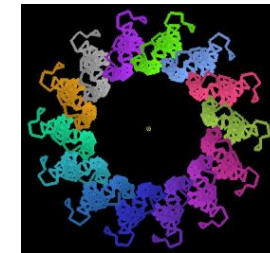
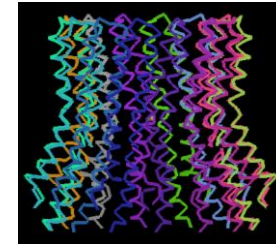
Step 2



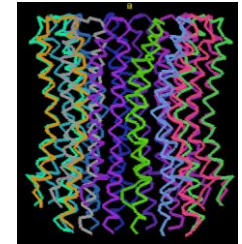
d = 62 Å



d = 59 Å



d = 43 Å



d = 43 Å

- Monomer at origin
- Rotate by angle φ
- Rotate by angle θ
- Rotate around r

- Monomer in new orientation
- Copy monomer
- Combine to 14mer ring
- Vary radius/diameter of ring



Phasing

- Molrep and scripting

Step 3

Run Molrep from command line with default settings

Contrast = 2.63 solution

After stick correction:

Move closer to origin

l_sym_operator : 1

new position(frac): 0.179 0.000 0.214

Sol_

S_Nmon RF TF theta phi chi tx ty tz

TFcnt wRfac Scor

S__1 2 1 180.00 0.00 0.43 0.179 0.000

0.214 4.35 0.404 0.869

--- convert "molrep.crd" to "molrep.pdb" ---

Contrast values for solutions:

3.0 → definitely solution

2.0 – 3.0 → solution

1.5 – 2.0 → maybe solution

Step 4

Rigid body refinement followed by restrained refinement of solution with Refmac5 with default settings and using 14fold NCS

		decrease		increase										
Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL	rmsCHIRAL				
\$\$														
0	0.4932	0.4620	0.432	71684.	3688.5	0.0213	0.963	1.758	0.891	0.094				
1	0.4362	0.4480	0.461	70870.	3688.3	0.0304	1.504	1.787	0.967	0.106				
2	0.4133	0.4404	0.482	70589.	3680.2	0.0152	0.723	1.074	0.531	0.053				
3	0.4006	0.4364	0.495	70388.	3675.5	0.0094	0.444	0.912	0.415	0.042				
4	0.3919	0.4320	0.509	70216.	3668.9	0.0082	0.379	0.916	0.416	0.042				
5	0.3856	0.4275	0.517	70060.	3661.5	0.0076	0.343	0.922	0.420	0.043				
6	0.3749	0.4257	0.525	69889.	3662.6	0.0091	0.402	1.030	0.465	0.048				
7	0.3648	0.4231	0.529	69784.	3667.9	0.0104	0.466	1.151	0.517	0.055				
8	0.3610	0.4213	0.535	69657.	3665.7	0.0101	0.451	1.162	0.522	0.055				
9	0.3570	0.4238	0.535	69590.	3666.5	0.0103	0.460	1.185	0.532	0.058				
10	0.3485	0.4244	0.539	69491.	3672.1	0.0125	0.552	1.346	0.604	0.064				
11	0.3459	0.4261	0.540	69516.	3676.8	0.0122	0.542	1.353	0.607	0.064				
12	0.3440	0.4251	0.540	69483.	3676.9	0.0122	0.539	1.367	0.611	0.064				
13	0.3424	0.4251	0.540	69491.	3679.1	0.0121	0.534	1.367	0.611	0.064				
14	0.3413	0.4228	0.542	69472.	3678.6	0.0121	0.535	1.373	0.613	0.064				
15	0.3402	0.4230	0.541	69501.	3682.2	0.0122	0.537	1.373	0.611	0.064				
16	0.3394	0.4198	0.546	69462.	3679.9	0.0122	0.538	1.378	0.614	0.064				



- Intermediate resolution
 - Muscle specific human β -enolase (PDB-ID: 2XSX)
- High resolution
 - PH domain of human AKT3 kinase (PDB-ID: 2X18)
- Low homology
 - PH domain of human Actin-binding protein anillin ANLN (PDB-ID: 2Y7B)
- Low resolution
 - Rotating transmembrane ring of F_0F_1 ATPsynthase from spinach chloroplasts (PDB-ID: 2W5J)
- NMR model
 - Thioredoxin domain of human TXNL2 (PDB-ID: 2WZ9)
- Twinning
 - Macro domain of human core histone H2A (PDB-ID: 2XD7)



- Thioredoxin domain of human TXNL2
- Glutaredoxin-3, PKC-interacting cousin of thioredoxin
- Location: cytoplasm
- Function: critical negative regulator of cardiac hypertrophy and a positive inotropic regulator
- Enzymatic reaction (thioredoxin domain):
does not possess any thioredoxin activity; no conserved catalytic motif
- Co-factor (full length): 2Fe-2S clusters
- Biologically active form (full length): Monomer and homodimer; the homodimer is probably linked by 2 2Fe-2S clusters
- Number of residues (thioredoxin domain): 116
- Molecular weight (thioredoxin domain): 12.6 kDa
- UniProt-ID: O76003



- Processing results (Mosflm, Scala)

	Overall	Inner Shell	Outer Shell
Low resolution limit	34.47	34.47	1.63
High resolution limit	1.55	4.90	1.55
Rmerge	0.091	0.023	0.581
Rmerge in top intensity bin	0.038	-	-
Rmeas (within I+/I-)	0.106	0.028	0.673
Rmeas (all I+ & I-)	0.106	0.028	0.673
Rpim (within I+/I-)	0.054	0.015	0.337
Rpim (all I+ & I-)	0.054	0.015	0.337
Fractional partial bias	-0.006	-0.034	0.049
Total number of observations	73276	2442	10641
Total number of unique	19534	703	2808
Mean (I /sd(I))	9.5	23.4	2.2
Completeness	99.8	99.5	99.6
Multiplicity	3.8	3.5	3.8
Average unit cell	42.63	52.06	58.60
	90.0	90.0	90.0
Space group	P 2 2 2		
Maximum resolution	1.55 Å		

Cell volume: 130052.023

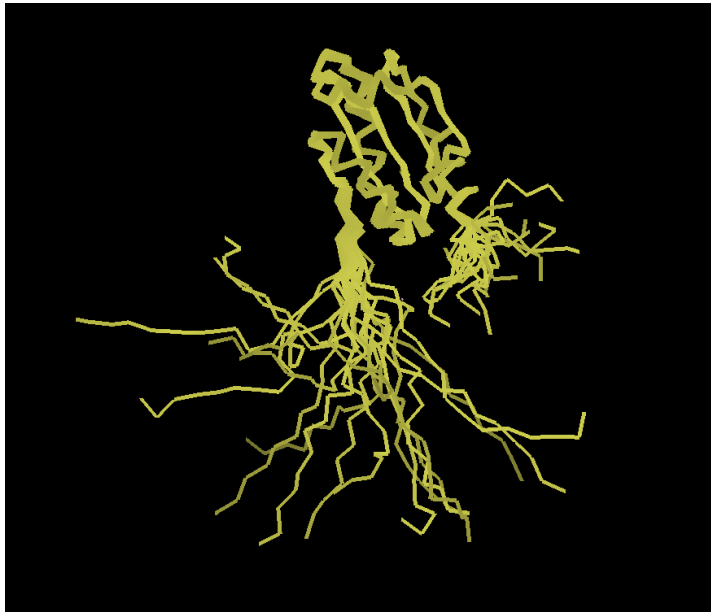
For estimated molecular weight 14683.

Nmol/asym Matthews Coeff %solvent P(1.50)
P(tot)

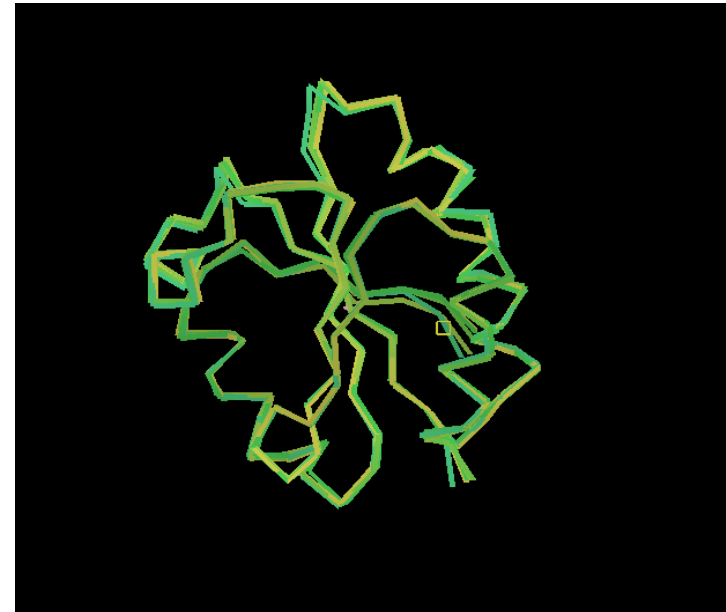
1	2.21	44.49	1.00	1.00
---	------	-------	------	------

Model preparation

- Single homologues structure in multiple conformations
e.g. 1DIY (90%)
- Reduce model PDB files to one chain (e.g. chain A); delete water molecules and ligands
- Use Coot to eliminate flexible protein parts
- Sequence alignment between model and target
- Run Chainsaw with edited model and sequence alignment; reduces side chains to last common atom between model and target sequence



Superpositioned search models



Core of superpositioned search models

Using an NMR model PDB-ID 2WZ9



Phasing

- Using Phaser
- Using search models as prepared above
- Using data as resulting from Scala

Phaser result

SpaceGroup of Solution: P 2 2 2

SINGLE solution

Solution written to PDB file:

/work/TXNL2A/4-phaser-d001-

mv/TXNL2A_phaser_19.1.pdb

Solution written to MTZ file:

/work/TXNL2A/4-phaser-d001-

mv/TXNL2A_phaser_19.1.mtz

Solution log-likelihood gain: -93.6145

Solution annotation (history):

RFZ=6.3 TFZ=3.8 PAK=2 LLG=-96

Molecule 1

LLG=-94

Maximum Likelihood Molecular Replacement Initial parameters from /work/TXNL2A/4-phase

Job title d001_x019 default 5 search ensembles cutter

Mode for molecular replacement automated search

Define data

MTZ in XNL2A_phaser d001-x019-mv.mtz

F d001 E d001 SIGF SIGF_d001

Resolution range 58.600 A to 2.5 A

Space group read from mtz file P222

Run Phaser with the mtz space group

Number of top solutions to output as PDB and MTZ files No. MTZ files not output for all modes

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name ensemble1 Define ensemble via pdb file(s)

PDB #1 XNL2A_phaser 2diy_model_16_cutter.pdb

Similarity of PDB #1 to the target structure sequence identity 0.9

Ensemble # 2

Ensemble Name ensemble2 Define ensemble via pdb file(s)

PDB #1 XNL2A_phaser 2diy_model_17_cutter.pdb

Similarity of PDB #1 to the target structure sequence identity 0.9

Ensemble # 3

Ensemble Name ensemble3 Define ensemble via pdb file(s)

PDB #1 XNL2A_phaser 2diy_model_18_cutter.pdb

Similarity of PDB #1 to the target structure sequence identity 0.9

Ensemble # 4

Ensemble Name ensemble4 Define ensemble via pdb file(s)

PDB #1 XNL2A_phaser 2diy_model_19_cutter.pdb

Similarity of PDB #1 to the target structure sequence identity 0.9

Ensemble # 5

Ensemble Name ensemble5 Define ensemble via pdb file(s)

PDB #1 XNL2A_phaser 2diy_model_20_cutter.pdb

Similarity of PDB #1 to the target structure sequence identity 0.9

Search details

Allow search with alternative ensembles (models) for a single component of the ASU on

Run Save or Restore Close

Using an NMR model PDB-ID 2WZ9



SGC

Changing space group

- Use unmerged, unscaled data file to check on space with Pointless and get a new data file with appropriate symmetry
- or
- Use Re-indexing to change space group

Pointless: prepare intensity data for scaling Initial parameters from /work/TXNL2A/3-mosflm_scala-d001-mv/CC

Job title: check space group

☒ Determine Laue group ☐ Match index to reference ☐ Choose a previous solution ☐ Just combine input files

Input reflection file type: MTZ file

Project name: TXNL2A crystal name: x019 dataset name: d001

MTZ #1 TXNL2A_scala d001-x019_2_002.mtz

Write output reflections in the best space/pointgroup

Output MTZ TXNL2A_scala d001-x019_2_002_pointless1.mtz

Always set primitive orthorhombic groups in cell length order (a-b-c) & allow monoclinic I2 setting of C2

Excluded Data ☐

Lattice Symmetry Determination ☐

Criteria For Accepting Partials ☐

Additional Options ☐

Run Save or Restore Close

* Cell Dimensions : (obsolete - refer to dataset cell dimensions above)

42.6300 52.0600 58.6000 90.0000 90.0000 90.0000

* Resolution Range :

0.00029 0.44443 (58.600 - 1.500 A)

* Sort Order :

1 2 3 0 0

* Space group = 'P222' (number 16)

Before changing

Reindex Reflections Initial parameters from /work/TXNL2A/3-mosflm_scala-d001-mv/CCF

Job title: from P222 to P212121

MTZ in TXNL2A_scala d001-x019_scala_1p55.mtz

MTZ out TXNL2A_scala d001-x019_scala_1p55_P212121.mtz

Reindex Details

Define transformation matrix by: choosing a standard transformation

Apply reindex matrix: None

☒ Change spacegroup to: P212121

☒ Reduce reflections to the asymmetric unit

Advanced reindexing options ☐

Run Save or Restore Close

* Cell Dimensions : (obsolete - refer to dataset cell dimensions above)

42.6300 52.0600 58.6000 90.0000 90.0000 90.0000

* Resolution Range :

0.00029 0.44443 (58.600 - 1.500 A)

* Sort Order :

0 0 0 0 0

* Space group = 'P 21 21 21' (number 19)

After changing

Using an NMR model PDB-ID 2WZ9



SGC

Phasing

- Using Phaser
- Using search models as prepared above
- Using data as resulting from Scala

Phaser result

SpaceGroup of Solution: P 21 21 21

SINGLE solution

Solution written to PDB file:

/work/TXNL2A/4-phaser-d001-
mv/TXNL2A_phaser_21.1.pdb

Solution written to MTZ file:

/work/TXNL2A/4-phaser-d001-
mv/TXNL2A_phaser_21.1.mtz

Solution log-likelihood gain: 67.6119

Solution annotation (history):

RFZ=8.0 TFZ=11.1 PAK=0 LLG=36

Molecule 1

LLG=68



- Intermediate resolution
 - Muscle specific human β -enolase (PDB-ID: 2XSX)
- High resolution
 - PH domain of human AKT3 kinase (PDB-ID: 2X18)
- Low homology
 - PH domain of human Actin-binding protein anillin ANLN (PDB-ID: 2Y7B)
- Low resolution
 - Rotating transmembrane ring of F_0F_1 ATPsynthase from spinach chloroplasts (PDB-ID: 2W5J)
- NMR model
 - Thioredoxin domain of human TXNL2 (PDB-ID: 2WZ9)
- Twinning
 - Macro domain of human core histone H2A (PDB-ID: 2XD7)



- Macro domain of human core histone H2A
- Location: nucleus
- Function: histone formation, DNA repair, DNA replication, chromosomal stability
- Enzymatic reaction:
 -
- Co-factor: -
- Biologically active form: assembled in histone
- Number of residues (macro domain): 187
- Molecular weight (macro domain): 20.2 kDa
- UniProt-ID: Q9P0M6



• Processing results (Mosflm, Scala) (before twinning was detected)

	Overall	Inner Shell	Outer Shell
Low resolution limit	41.85	41.85	2.31
High resolution limit	2.19	6.93	2.19
Rmerge	0.082	0.037	0.371
Rmerge in top intensity bin	0.049	-	-
Rmeas (within I+/I-)	0.093	0.042	0.416
Rmeas (all I+ & I-)	0.093	0.042	0.416
Rpim (within I+/I-)	0.043	0.019	0.185
Rpim (all I+ & I-)	0.043	0.019	0.185
Fractional partial bias	-0.064	-0.069	-0.015
Total number of observations	101359	3048	14575
Total number of unique	22223	696	3180
Mean (I /sd(I))	10.4	19.6	3.9
Completeness	99.0	90.2	99.0
Multiplicity	4.6	4.4	4.6
Average unit cell	57.04	61.58	242.51
	90.0	90.0	90.0
Space group	C 2 2 2		
Maximum resolution	2.19 Å		

Cell volume: 851822.000

For estimated molecular weight 22364.

Nmol/asym Matthews Coeff %solvent P(1.50)
P(tot)

1	4.76	74.18	0.00	0.01
2	2.38	48.36	0.99	0.99
3	1.59	22.55	0.00	0.00

Pointless

No changes in space group were suggested

Phenix.Xtriage

Determining possible twin laws.

0 merohedral twin operators found
0 pseudo-merohedral twin operators found
In total, 0 twin operator were found

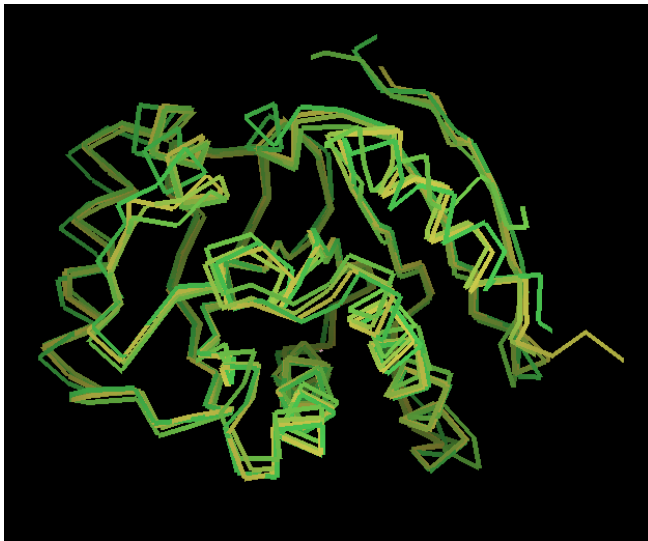
The results of the L test indicate that the intensity statistics are significantly different than is expected from good to reasonable, untwinned data.

As there are no twin laws possible given the crystal symmetry, there could be a number of reasons for the departure of the intensity statistics from normality.

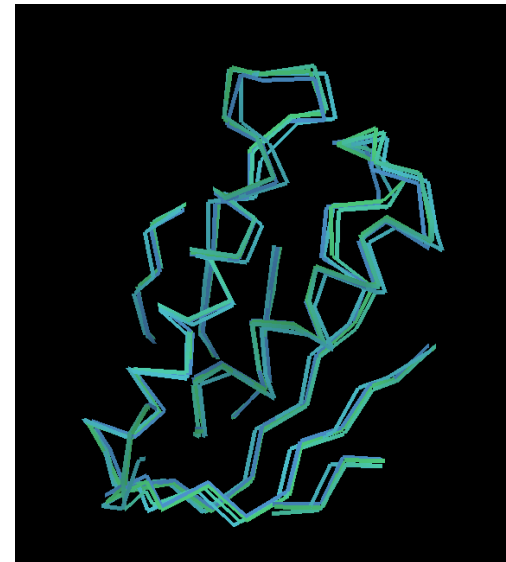
Overmerging pseudo-symmetric or twinned data, intensity to amplitude conversion problems as well as bad data quality might be possible reasons. It could be worthwhile considering reprocessing the data.

Model preparation

- Single homologues structure in multiple conformations
e.g. 1YD9 (64%), 1ZR3 (62%), 1ZR5 (65%), 3IID (62%)
- Reduce model PDB files to one chain (e.g. chain A); delete water molecules and ligands
- Use Coot to eliminate flexible protein parts
- Sequence alignment between model and target
- Run Chainsaw with edited model and sequence alignment; reduces side chains to last common atom between model and target sequence



Superpositioned search models



Core of superpositioned search models



Phasing

- Using Phaser
- Using search models as prepared above
- Using data as resulting from Scala

Phaser result

SpaceGroup of Solution: C 2 2 21

SINGLE solution

Solution written to PDB file: /work/H2AFY2A/6-phase-mv/H2A_phase_2_4.1.pdb

Solution written to MTZ file: /work/H2AFY2A/6-phase-mv/H2A_phase_2_4.1.mtz

Solution log-likelihood gain: 648.297

Solution annotation (history):

RFZ=8.2 TFZ=13.8 PAK=0 LLG=181

Molecule 1

RFZ=8.0 TFZ=23.9 PAK=0 LLG=543

Molecule 2

LLG=648

Maximum Likelihood Molecular Replacement Initial parameters from /work/H2AFY2A/6-phas

Job title quick phase

Mode for molecular replacement automated search

Define data

MTZ in H2A_phase_2 H2AFY2A-d001-x010.mtz

F d001 SIGF SIGF_d001

Resolution range 242.510 A to 2.5 A

Space group read from mtz file C222

Run Phaser with all choices of alternative space group

Number of top solutions to output as PDB and MTZ files 10

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name ensemble1 Define ensemble via pdb file(s)

PDB #1 H2A_phase_2 1YD9_A-coot-0.pdb

Similarity of PDB #1 to the target structure sequence identity 0.64

PDB #2 H2A_phase_2 1ZR3_A-coot-0.pdb

Similarity of PDB #2 to the target structure sequence identity 0.62

PDB #3 H2A_phase_2 1ZR5_A-coot-0.pdb

Similarity of PDB #3 to the target structure sequence identity 0.65

PDB #4 H2A_phase_2 3IID_A-coot-0.pdb

Similarity of PDB #4 to the target structure sequence identity 0.62

Edit list Add superimposed PDB file to the ensemble

Edit list Add ensemble

Search details

Allow search with alternative ensembles (models) for a single component of the ASU off

Perform search using ensemble1 Number of copies to search for 2

Edit list Add another search

Final selection criterion for ROTATION search peaks percentage of top peak

Final selection criterion for TRANSLATION search peaks percentage of top peak

Packing criterion best packing solutions Max # clashes 10 Clash distance 3.0

Composition of the asymmetric unit

Total scattering determined by components in asymmetric unit

Component #1 protein sequence file Number in asymmetric unit 2

SEQ file H2A_phase_2 H2AFY2A-x010.seq

Edit list Define another component

Define search sets (any already known rotations/translations)

Additional parameters

Run Save or Restore Close



Refinement after phasing and changing space group

- Using Refmac5
- All settings are default

Refinement statistics:

Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL	rmsCHIRAL
------	-------	-------	-----	-----	---------	---------	-------	---------	-------	-----------

0	0.4507	0.4454	0.546	127990.	6857.3	0.
1	0.4362	0.4407	0.554	127316.	6838.6	0.
2	0.4299	0.4391	0.559	127020.	6829.0	0.
3	0.4257	0.4348	0.566	126892.	6826.0	0.
4	0.4222	0.4307	0.568	126797.	6825.2	0.
5	0.4199	0.4283	0.571	126731.	6825.2	0.
6	0.4179	0.4257	0.573	126703.	6826.1	0.
7	0.4164	0.4232	0.575	126684.	6826.9	0.
8	0.4151	0.4224	0.577	126661.	6827.0	0.
9	0.4141	0.4215	0.579	126654.	6827.7	0.
10	0.4133	0.4205	0.581	126625.	6826.9	0.

The results of the L-test indicate that the intensity statistics are significantly different than is expected from good to reasonable, untwinned data.

As there are no twin laws possible given the crystal symmetry, there could be a number of reasons for the departure of the intensity statistics from normality.

Overmerging pseudo-symmetric or twinned data, intensity to amplitude conversion problems as well as bad data quality might be possible reasons.

It could be worthwhile considering reprocessing the data.

Note:

Only marginal change in Rfactors and FOM



• Processing results (Mosflm, Scala) (assuming twinning)

	Overall	Inner Shell	Outer Shell
Low resolution limit	41.86	41.86	2.21
High resolution limit	2.10	6.64	2.10
Rmerge	0.070	0.029	0.372
Rmerge in top intensity bin	0.040	-	-
Rmeas (within I+/I-)	0.091	0.039	0.476
Rmeas (all I+ & I-)	0.091	0.039	0.476
Rpim (within I+/I-)	0.057	0.026	0.294
Rpim (all I+ & I-)	0.057	0.026	0.294
Fractional partial bias	-0.041	-0.062	-0.021
Total number of observations	109576	3184	12575
Total number of unique	45536	1454	5542
Mean (I /sd(I))	7.5	14.9	2.3
Completeness	92.7	88.8	77.7
Multiplicity	2.4	2.2	2.3
Average unit cell	57.03	61.64	242.23
	90.0	89.64	90.0
Space group	C 2		
Maximum resolution	2.10 Å		

Cell volume: 851502.000

For estimated molecular weight 22364.

Nmol/asym Matthews Coeff %solvent P(1.50)
P(tot)

1	9.52	87.09	0.00	0.00
2	4.76	74.17	0.00	0.01
3	3.17	61.26	0.12	0.17
4	2.38	48.34	0.70	0.68
5	1.90	35.43	0.17	0.14
6	1.59	22.52	0.00	0.00
7	1.36	9.60	0.00	0.00

Phenix.Xtriage

The following twin laws have been found:

| Type | Axis | R metric (%) | delta (le Page) | delta (Lebedev) | Twin law |

| PM | 2-fold | 0.278 | 0.360 | 0.005 | -h,-k,l |

M: Merohedral twin law

PM: Pseudomerohedral twin law

0 merohedral twin operators found

1 pseudo-merohedral twin operators found

In total, 1 twin operator were found



Phasing

- Using Phaser
- Using search models as prepared above
- Using data as resulting from Scala

Phaser result

SpaceGroup of Solution: C 1 2 1

SINGLE solution

Solution written to PDB file: /work/H2AFY2A/9-phase-C2-mv/H2A_phaseC2_4.1.pdb

Solution written to MTZ file: /work/H2AFY2A/9-phase-C2-mv/H2A_phaseC2_4.1.mtz

Solution log-likelihood gain: 1101.92

Solution annotation (history):

RFZ=8.2 TFZ=4.9 PAK=0 LLG=89

Molecule 1

RFZ=8.2 TFZ=16.9 PAK=0 LLG=321

Molecule 2

RFZ=7.2 TFZ=19.1 PAK=0 LLG=585

Molecule 3

RFZ=7.2 TFZ=23.9 PAK=0 LLG=924

Molecule 4

LLG=1102

Maximum Likelihood Molecular Replacement Initial parameters from /work/H2AFY2A/9-phas

Job title: try phasing in C2

Mode for molecular replacement: automated search

Define data

MTZ in: H2A_phaseC2 H2AFY2A-d001-x010_C2.mtz

F: d001 SIGF: SIGF_d001

Resolution range: 242.225 A to 2.5 A

Space group read from mtz file: C121

Run Phaser with: the mtz space group

Number of top solutions to output as PDB and MTZ files: NB: MTZ files not output for all modes

Define ensembles (each comprising possible model(s) for a protein/nucleic acid component)

Ensemble # 1

Ensemble Name: ensemble1 Define ensemble via: pdb file(s)

PDB #1 H2A_phaseC2: 1YD9_A-coot-0.pdb Similarity of PDB #1 to the target structure: sequence identity 0.64

PDB #2 H2A_phaseC2: 1ZR3_A-coot-0.pdb Similarity of PDB #2 to the target structure: sequence identity 0.62

PDB #3 H2A_phaseC2: 1ZR5_A-coot-0.pdb Similarity of PDB #3 to the target structure: sequence identity 0.65

PDB #4 H2A_phaseC2: 3IID_A-coot-0.pdb Similarity of PDB #4 to the target structure: sequence identity 0.62

Search details

Allow search with alternative ensembles (models) for a single component of the ASU: off

Perform search using: ensemble1 Number of copies to search for: 4

Final selection criterion for ROTATION search peaks: percentage of top peak

Final selection criterion for TRANSLATION search peaks: percentage of top peak

Packing criterion: best packing solutions Max # clashes: 10 Clash distance: 3.0

Composition of the asymmetric unit

Total scattering determined by: components in asymmetric unit

Component #1: protein sequence file: Number in asymmetric unit: 4

SEQ file: H2A_phaseC2 H2AFY2A-x010.seq

Define search sets (any already known rotations/translations)

Additional parameters

Run Save or Restore Close



The following parameters are worth checking before setting up a phasing experiment:

- Check for R_{merge} overall and highest resolution shell
- Desired **meanI/sigI** in **highest resolution shell >2**; alter maximum resolution if necessary
- Desired **completeness >95%**
- Desired **multiplicity >3**
- Check for **twinning**
- Check **space/point group**
- Matthews coefficient → expected number of molecules to search for

The following parameters are worth checking before setting up a phasing experiment:

- **Data base** search to find closest homologues
- Prepare **sequence alignments** between model sequence and target sequence
- Delete all **water molecules** and **ligands**
- Use only **one chain per homologue** as search model
- Reduce molecules to a **core** and **superpose models**
- Based on sequence alignment create search models only having **last common atom in side chains** present or a **polyAla** model



The following programs have been used:

- Mosflm, XDS, HKL2000 → data processing
- Scala, Scalepack, Truncate/Ctruncate, Scalepack2MTZ → merging, scaling data and adding Rfree
- Matthews coefficient, Phenix_xtriage, Pointless, Reindex → check and edit/manipulate data
- Various data bases, Chainsaw, Coot → model preparation
- Phaser, Molrep → phasing
- Parrot, DM → density modification after phasing
- Coot, ARP/wARP, Buccaneer, Refmac, Phenix_refine, Buster_TNT, Tslanl various server and data bases → building and refinement