Structure proof-reading at SGC (Structural Genomics Consortium)

Melanie Vollmar

Structural Genomics Consortium

University of Oxford

SGC Toronto

SGC Oxford

SGC Stockholm

Reason for structure proof-reading

- Importance shown by structures recently retracted from the PDB
- Discuss interesting features in relation to biological relevance

- Second opinion whether refinement is complete

- Keep a general standard for all structures deposited by SGC
- Ensure high quality structures for public domain

You would never submit a paper to a journal without giving it to a colleague/supervisor to read, would you?

Files for proof-reading

- Processing log-file, e.g. from Scala, Truncate, Denzo
- MTZ file from processing (optional unmerged, unscaled data file)
- Phasing log-file, e.g. from Phaser (optional)
- Refinement log-file of last refinement step, e.g. Refmac5
- PDB and MTZ file from last refinement step
- Dictionary (.CIF)
- TLS input file
- Protein sequence
- Any other keyword file used in refinement
- A text file with comments for proof reader, e.g. problematic areas in the structure, notes about processing and phasing

Proof-reading steps

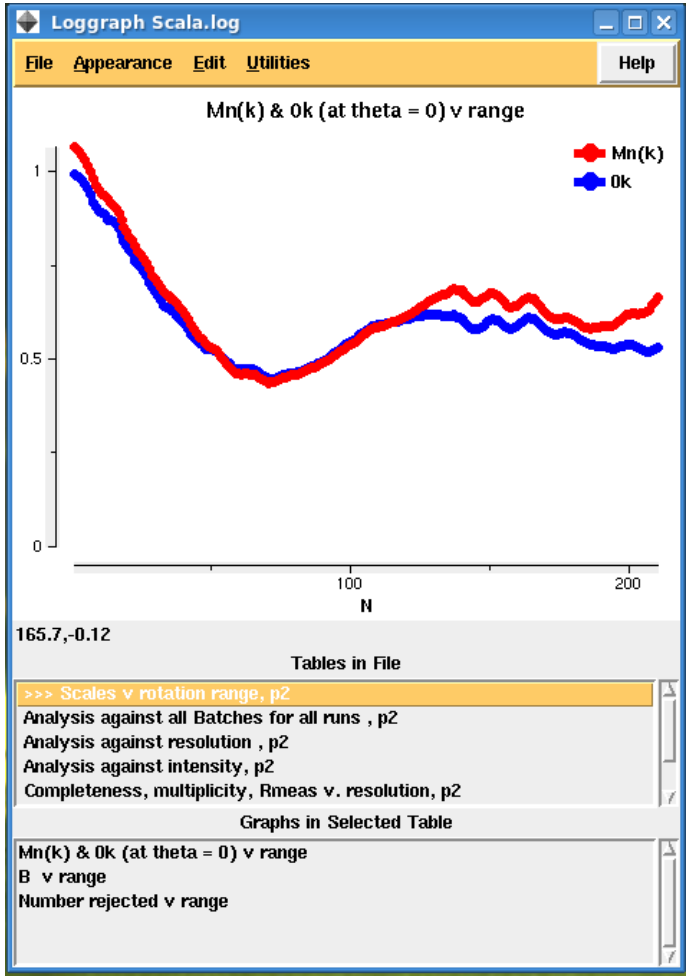
- Proof-reading check list:
  - ftp://hestia.sgc.ox.ac.uk/pub/px
- Checking data processing
- Check space group
- Check for twinning
- Check phasing result (optional → see appendix)
- Check last refinement step
- Check waters
- Run Molprobity or Quality Control Check
- Run Moleman2
- Blast sequence and align sequence
- Check ligand geometry
- Walk structure and map

Checking data processing

Critical values from Scala/Truncate log-file

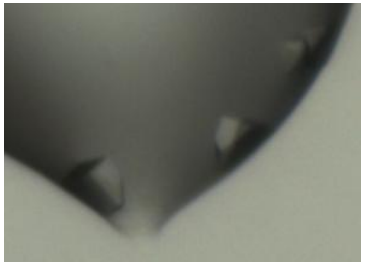
Number of rejections per image

N	Run.Rot	MidPhi	Batch	Bfactor	Mn(k)	0k	Number	NumReject
1	1.1	-49.50	1	-0.694	1.0651	0.9940	1703	0
2	1.2	-48.50	2	-0.688	1.0622	0.9905	2193	0
3	1.3	-47.50	3	-0.677	1.0564	0.9851	2219	0
4	1.4	-46.50	4	-0.668	1.0453	0.9774	2202	0
5	1.5	-45.50	5	-0.656	1.0339	0.9671	2198	0
6	1.6	-44.50	6	-0.641	1.0180	0.9542	2217	1
7	1.7	-43.50	7	-0.629	1.0017	0.9395	2208	0
8	1.8	-42.50	8	-0.614	0.9811	0.9185	2217	0



What to look out for:

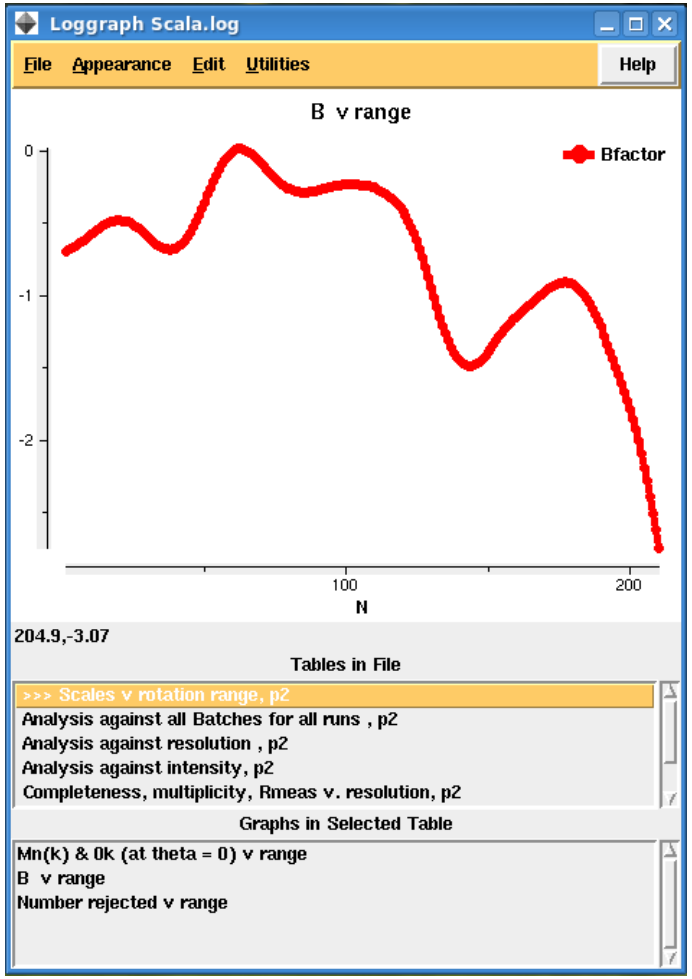
- Low number of rejected reflections per image; a maximum of maybe 5
- Ideally a constant scaling factor of 1; except if crystals have irregular shape



Checking data processing

Critical values from Scala/Truncate log-file

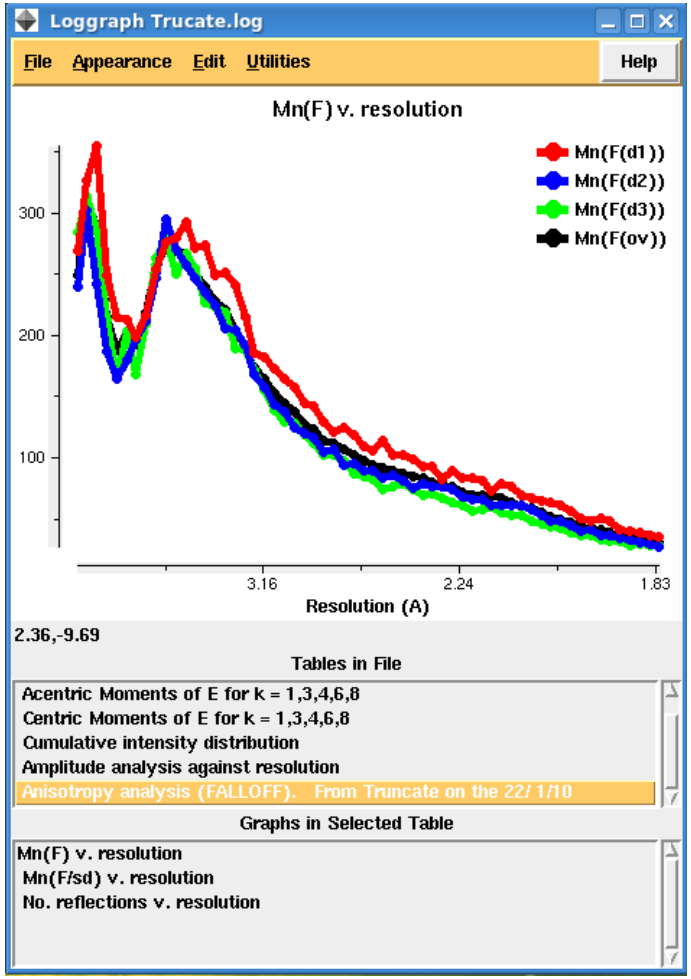
Scaling Bfactor



What to look out for:

- Drop in scaling Bfactor below -10 indicates radiation damage
- Severe anisotropy of data

Anisotropy



Checking data processing

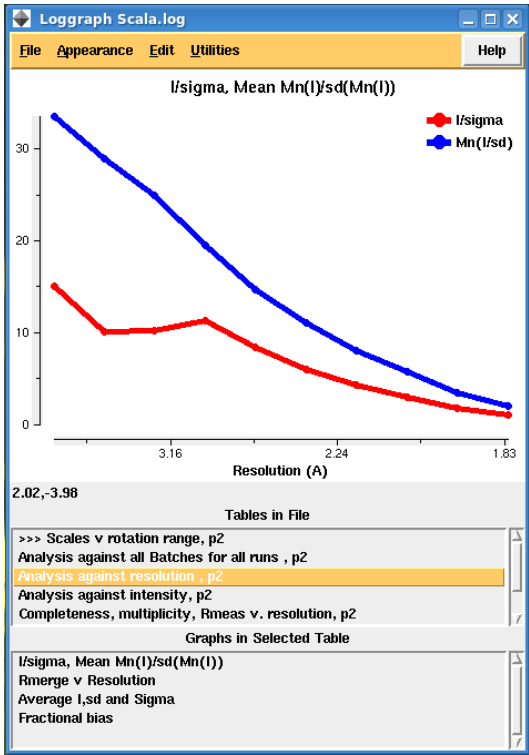
Critical values from Scala/Truncate log-file

Rmerge and Mn (I/sd)

N	1/d^2	Dmin(A)	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
$$																	
1	0.0302	5.76	0.034	0.033	0.034	-	0	2569	171	15.0	169	33.6	15181	3471	282	-0.034	6396
2	0.0604	4.07	0.051	0.045	0.045	-	0	2766	274	10.1	198	29.0	25328	6141	258	-0.030	10449
3	0.0906	3.32	0.053	0.046	0.048	-	0	2226	217	10.2	179	24.9	34283	8065	292	-0.026	14346
4	0.1208	2.88	0.058	0.051	0.050	-	0	1024	91	11.3	99	19.5	41036	9532	296	-0.018	17632
5	0.1509	2.57	0.083	0.071	0.053	-	0	545	65	8.4	69	14.7	47223	10882	294	-0.015	20428
6	0.1811	2.35	0.121	0.104	0.058	-	0	359	60	6.0	61	11.1	51759	11871	295	-0.009	22521
7	0.2113	2.18	0.172	0.145	0.064	-	0	251	58	4.3	62	8.0	56805	12968	295	-0.016	25099
8	0.2415	2.03	0.249	0.211	0.070	-	0	177	59	3.0	63	5.8	60838	13841	295	-0.020	26770
9	0.2717	1.92	0.430	0.366	0.078	-	0	99	57	1.7	62	3.5	64727	14717	297	-0.015	28369
10	0.3019	1.82	0.734	0.615	0.086	-	0	56	56	1.0	62	2.0	64070	15068	213	-0.008	27775
$$																	

">For inline graphs use a Java browser</applet>

	Rmrg	Rfull	Rcum	Ranom	Nanom	Av_I	SIGMA	I/sigma	sd	Mn(I/sd)	Nmeas	Nref	Ncent	FRCBIAS	Nbias
Overall:	0.086	0.073	0.086	-	0	665	109	6.1		86	11.6	461250	106556	2817	-0.024	199785



What to look out for:

- Increase of Rmerge from low to high resolution; increase without sudden jumps
- At a reported value for mean I/sigI of 2 Rmerge reaches a maximum where its error can be trusted
- Usually goes along with a sudden decrease in completeness
- For phasing and refinement a higher resolution may be used on the expense of increasing error and incomplete data; beware and don't report it as resolution of the structure

Checking data processing

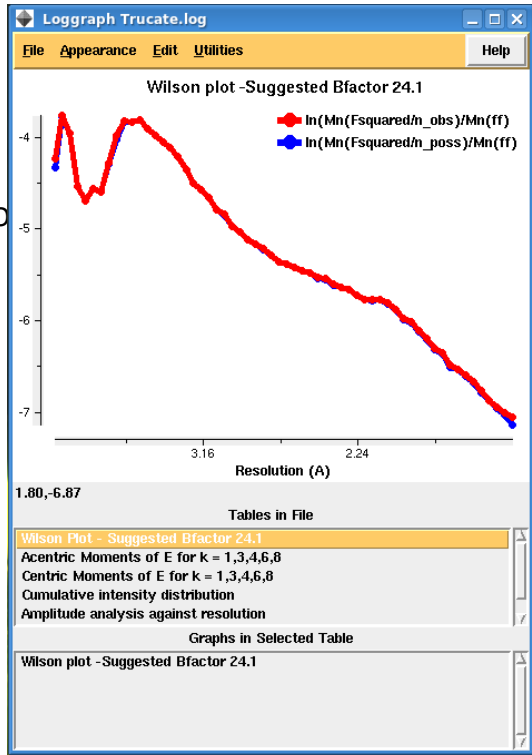
Critical values from Scala/Truncate log-file

Completeness and multiplicity

N	1/resol^2	Dmin	Nmeas	Nref	Ncent	%poss	C%poss	Mlplct	AnoCmp	AnoFrc	AnoMlt	Rmeas	Rmeas0 (Rsym)	Rpim	RpimO	PCV	PCV0
$$	$$																
1	0.030	5.76	15195	3485	286	99.0	99.0	4.4	-	-	-	0.038	0.038	0.034	0.018	0.018	0.045	0.045
2	0.060	4.07	25435	6248	279	98.7	98.8	4.1	-	-	-	0.059	0.059	0.051	0.028	0.028	0.066	0.066
3	0.091	3.32	34328	8110	293	99.8	99.3	4.2	-	-	-	0.061	0.061	0.053	0.028	0.028	0.068	0.068
4	0.121	2.88	41064	9560	296	99.8	99.4	4.3	-	-	-	0.065	0.065	0.058	0.031	0.031	0.075	0.075
5	0.151	2.57	47244	10903	294	99.7	99.5	4.3	-	-	-	0.094	0.094	0.083	0.044	0.044	0.110	0.110
6	0.181	2.35	51784	11896	295	99.6	99.5	4.4	-	-	-	0.137	0.137	0.121	0.064	0.064	0.160	0.160
7	0.211	2.18	56854	13017	295	99.4	99.5	4.4	-	-	-	0.195	0.195	0.172	0.091	0.091	0.227	0.227
8	0.242	2.03	60887	13890	295	99.2	99.5	4.4	-	-	-	0.282	0.282	0.249	0.131	0.131	0.329	0.329
9	0.272	1.92	64787	14777	297	99.0	99.4	4.4	-	-	-	0.487	0.487	0.430	0.227	0.227	0.574	0.574
10	0.302	1.82	64314	15312	252	97.8	99.2	4.2	-	-	-	0.837	0.837	0.734	0.395	0.395	0.988	0.988
Overall			461892	107198	2882	99.2	99.2	4.3	-	-	-	0.098	0.098	0.086	0.046	0.046	0.111	0.111

Nmeas	Nref	Ncent	%poss	C%poss	Mlplct	AnoCmp	AnoFrc	AnoMlt	Rmeas	Rmeas0 (Rsym)	Rpim	RpimO	PCV	PCV0

Wilson Bfactor

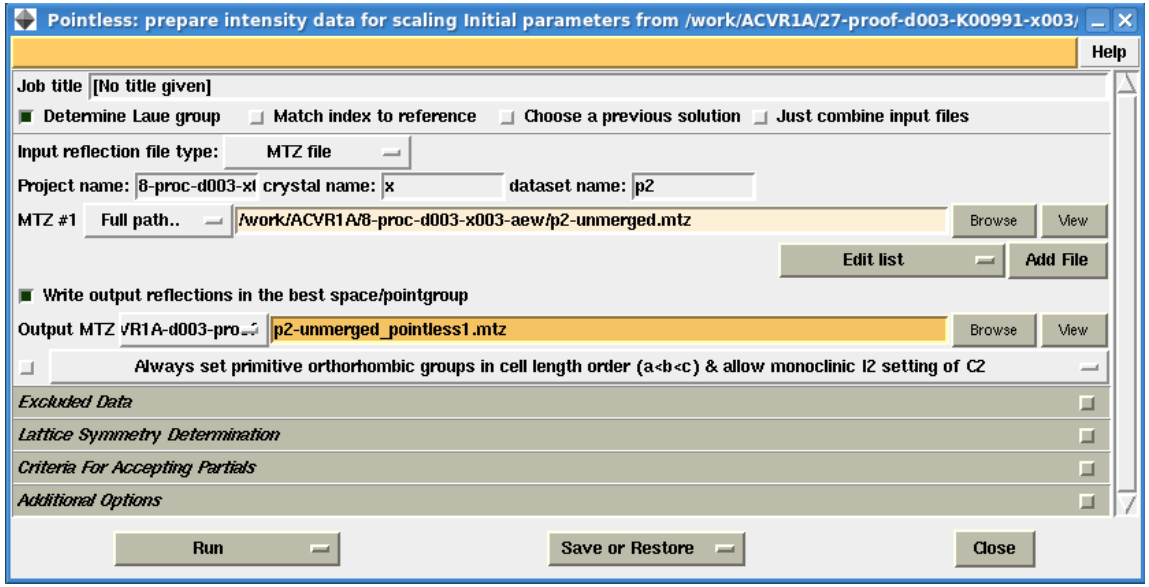


What to look out for:

- Overall completeness of the data > 95%
- Note: in high resolution shell not always guaranteed and may be as low as 60%
- Slope of Wilson Plot is almost a straight line
- Bfactor is sensible for the resolution range; increases with decreasing resolution

Checking space group

Run Pointless on un-merged data file to check for space group




```
* Cell Dimensions : (obsolete - refer to dataset cell dimensions above)
  83.6000   98.7000   83.8500   90.0000  117.4200   90.0000
*  Resolution Range :
   0.00068    0.30189     (     38.461 -      1.820 A )
* Sort Order :
     1     2     3     4     5
* Space group = 'P 1 21 1' (number     4)
```

Cell dimensions and space group as used in refinement

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

83.6000 98.7000 83.8500 90.0000 117.4200 90.0000

* Resolution Range :

0.00068 0.30189 (38.461 - 1.820 A)

* Sort Order :

1 2 3 4 5

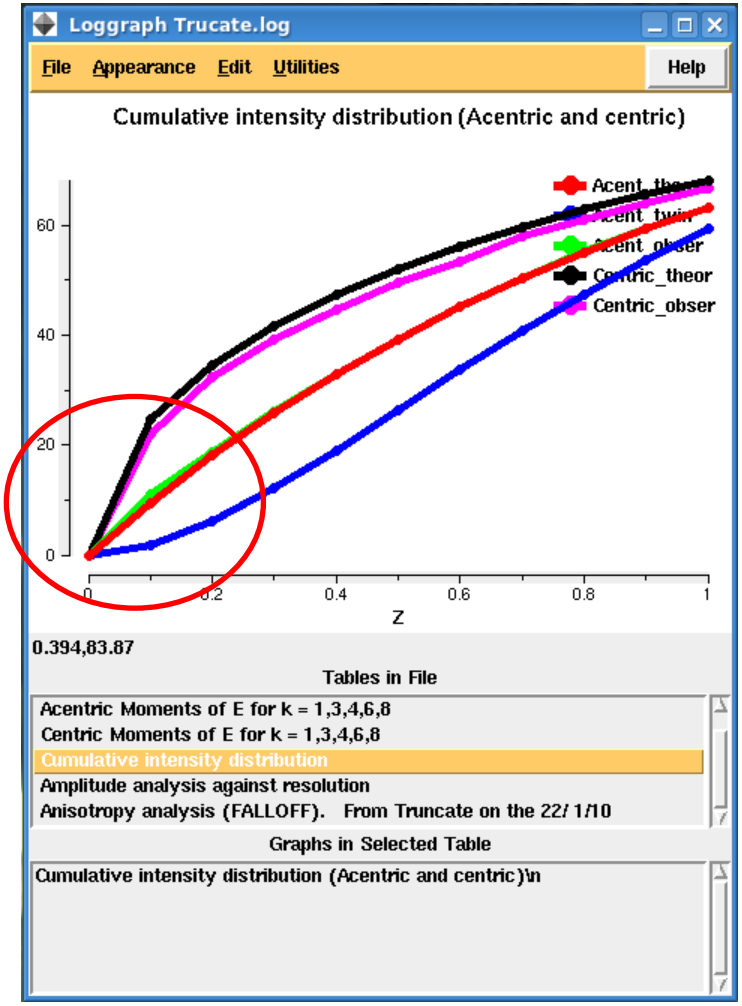
* Space group = 'P 1 21 1' (number 4)

Cell dimensions and space group as found by Pointless from unmerged data

Checking for twinning

- Using Loggraph and Truncate log file
- Or a web server

  http://nihserver.mbi.ucla.edu/Twinning/



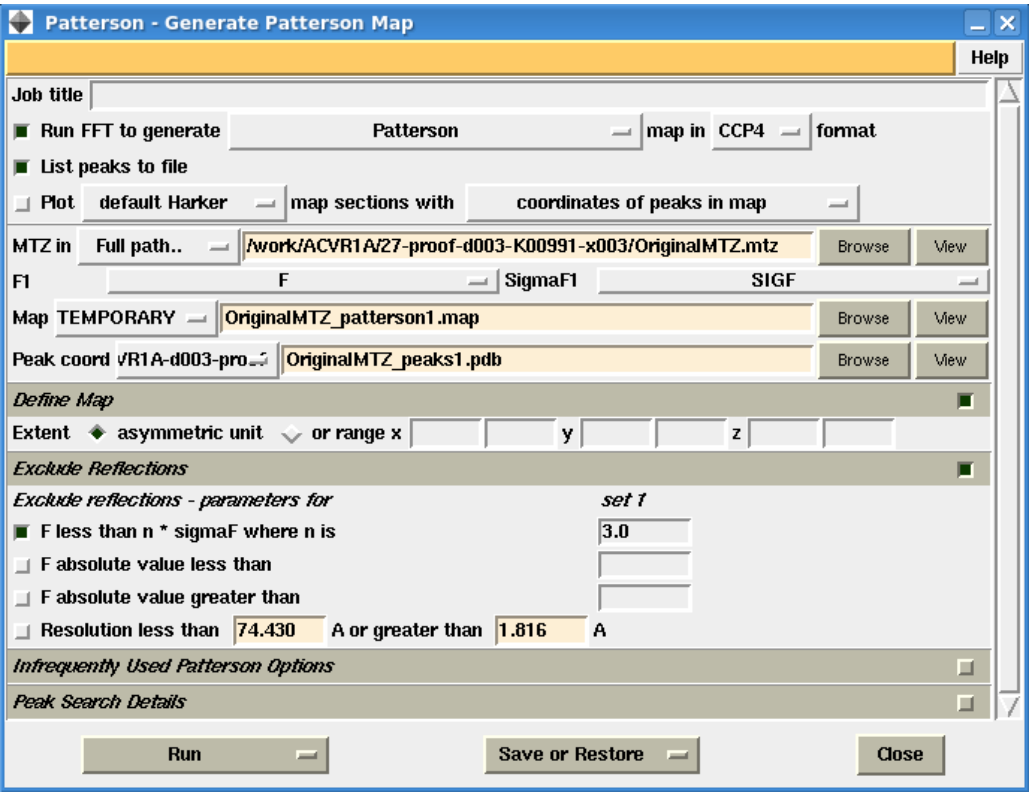
What to look out for:

- Any indication of twinning
- Here: observed data almost identical to acentric theoretically expected data

Checking for pseudo-translation

- Using FFT/Peakmax to generate native Patterson map
- Or a Zanuda web server

http://www.ysbl.york.ac.uk/YSBLPrograms/index.jsp



Order No. Site Height/Rms Grid Fractional coordinates Orthogonal coordinates

Origin

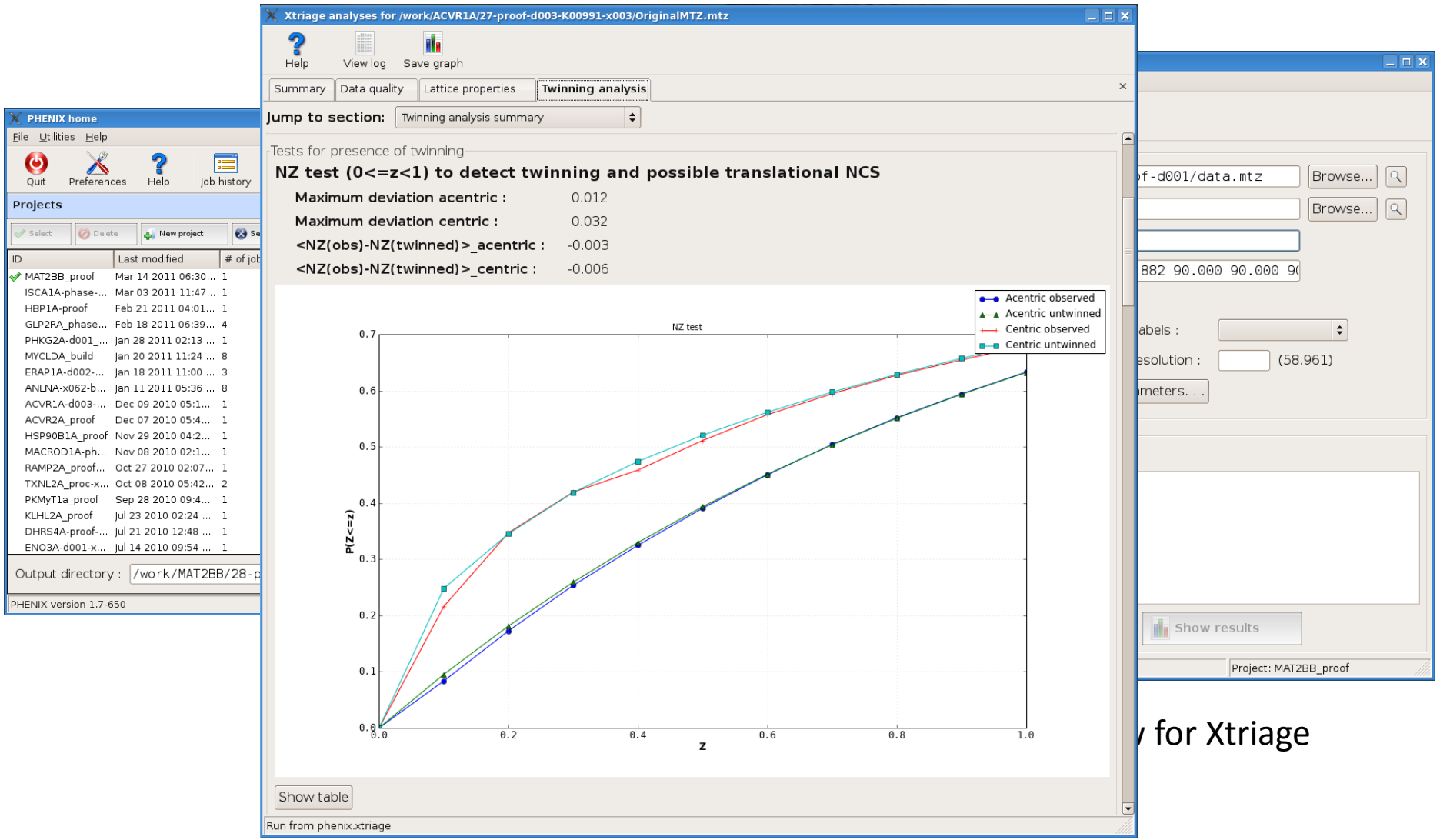
1	1	1	226.28	0	0	0	0.0000	0.0000	0.0000	0.00	0.00	0.00
2	8	7	7.97	4	5	137	0.0289	0.0307	0.9784	-35.36	3.03	72.82
3	3	3	7.75	3	0	9	0.0188	0.0000	0.0630	-0.86	0.00	4.69
4	11	9	7.67	1	6	135	0.0045	0.0367	0.9652	-36.90	3.62	71.84
5	10	9	7.52	1	6	6	0.0054	0.0364	0.0416	-1.16	3.59	3.10
6	6	5	7.16	5	2	136	0.0342	0.0113	0.9726	-34.70	1.12	72.39
7	7	6	7.15	4	5	7	0.0260	0.0309	0.0485	0.30	3.05	3.61
8	5	4	6.79	0	2	132	0.0000	0.0109	0.9435	-36.43	1.07	70.23
9	12	10	5.58	3	7	1	0.0203	0.0446	0.0068	1.43	4.40	0.51
10	9	8	5.54	5	6	0	0.0348	0.0371	0.0000	2.91	3.66	0.00
11	2	2	4.19	8	0	6	0.0581	0.0000	0.0450	3.12	0.00	3.35
12	13	0	3.35	6	80	26	0.0425	0.5000	0.1823	-3.48	49.35	13.57

What to look out for:

- Any indication for pseudo-translation
- Any peak as big as 15% of the origin shows for example a NCS on a crystallographic axes

Checking for twinning

Using Phenix.Xtriage to check for twinning



for Xtriage

Checking for twinning

- Using Phenix.Xtriage to check for twinning

What to look out for:

- Any indication of twinning
- Other tests run in Phenix.Xtriage are to find pseudo-translation and anisotropy

```
##----------------------------------------------##
##               Twinning Analyses              ##
##----------------------------------------------##

Using data between 10.00 to 2.59 Angstrom.

Determining possible twin laws.

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
============================================================
The largest off-origin peak in the Patterson function is 5.41% of the
height of the origin peak. No significant pseudotranslation is detected.

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 twin laws possible given the crystal symmetry, twinning could
be the reason for the departure of the intensity statistics from normality.
It might be worthwhile carrying out refinement with a twin specific target function.

Note that the symmetry of the intensities suggest that the assumed space group  is too low. As
twinning is however suspected, it is not immediatly clear if this is the case. Careful reprocessing
and (twin)refinement for all cases might resolve this question.
```

Checking last refinement step: log file

- Using a text editor to look for critical values

```
Data line--- make check NONE
 Data line--- make hydrogen ALL hout NO peptide NO cispeptide YES
              ssbridge YES symmetry YES sugar YES connectivity YES link YES
 Data line--- refi type REST resi MLKF meth CGMAT bref ISOT
 Data line--- refi tlsc 10
 Data line--- TLSD WATERS EXCLUDE
 Data line--- ncyc 10
 Data line--- scal type SIMP LSSC ANISO EXPE
 Data line--- solvent YES
 Data line--- weight MATRIX 0.31
 Data line--- monitor MEDIUM torsion 10.0 distance 10.0 angle 10.0 plane 10.0
              chiral 10.0 bfactor 10.0 bsphere 10.0 rbond 10.0 ncsr 10.0
 Data line--- labin FP=F SIGFP=SIGF FREE=FreeR_flag
 Data line--- labout FC=FC FWT=FWT PHIC=PHIC PHWT=PHWT DELFWT=DELFWT
              PHDELWT=PHDELWT FOM=FOM
 Data line--- temp 0.5 3.0 5.0 8.0 11.0
 Data line--- angle 1.1
 Data line--- chir 1.65
 Data line--- PNAME 8-proc-d003-x003-aew
 Data line--- DNAME p2
 Data line--- RSIZE 80
 Data line--- END
```

Usage of hydrogens

Refinement of Us, e.g. isotropic

Overall weighting of the data; tightness depends on resolution

Weighting of temperature factors; allow for side chain flexibility

What to look out for:

- How is the data weighted?

  Depends on data quality and should be chosen so that refinement is stable
- Are hydrogens present in refinement?

  To improve electrostatics within the structure hydrogens should always be used even for low resolution data
- Is the refinement of Us isotropic or anisotropic?

  Low resolution data not accurate enough so should be refined isotropically; for high resolution anisotropy can improve the refinement
- How are the Bfactors weighted and are the Bfactors complete?

  Weighting should be sensible for a stable refinement; only complete Bfactors give information about flexibility but also about badly build areas in a structure

Checking last refinement step: log file

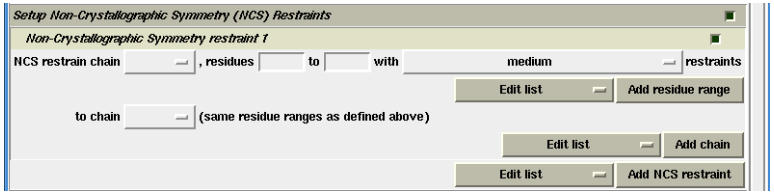
- Using a text editor to look for critical values

  in case of multiple copies of the same molecule within the unit cell; group definition handled differently in different refinement programs

What to look out for:

- Is NCS present and how are the groups defined and weighted?

  Improper NCS definition may be a reason for a stuck refinement; can be released for high resolution data; additionally check NCS definition in COOT by superpositioning



Data line--- NCSR NCHAins 3 CHAIns B A C NSPANS 1 696 710 1
Data line--- NCSR NCHAins 3 CHAIns B A C NSPANS 1 726 735 1
Data line--- NCSR NCHAins 3 CHAIns B A C NSPANS 1 591 609 4
Data line--- NCSR NCHAins 3 CHAIns B A C NSPANS 1 800 808 4
Data line--- NCSR NCHAins 3 CHAIns B A C NSPANS 1 915 933 4
Data line--- NCSR NCHAins 2 CHAIns B C NSPANS 1 761 788 6
Data line--- NCSR NCHAins 2 CHAIns B C NSPANS 1 809 815 6
Data line--- NCSR NCHAins 2 CHAIns B C NSPANS 1 634 855 6

Number of chains on which the NCS restraint will be applied, e.g. 3

Identifier for chains in the NCS restraint, e.g. B A C; 1st in the list is master chain

Weighting of the NCS restraint, e.g. 4 which corresponds to "medium" weighting if picking from GUI

Number of residue spans in the NCS restraint, e.g. 1; residue span, e.g. from residue 634 to 855

Checking last refinement step: log file

- Using a text editor to look for critical values

  To account for local and global flexibility; treated differently in different refinement programs

```
 Data line--- make check NONE
 Data line--- make    hydrogen ALL    hout NO    peptide NO    cispeptide
YES
                ssbridge YES    symmetry YES    sugar YES    connectivity YES
link YES
 Data line--- refi    type REST    resi MLKF    meth CGMAT    bref ISOT
 Data line--- refi tlsc 10
 Data line--- TLSD WATERS EXCLUDE
 Data line--- ncyc 10
 Data line--- scal    type SIMP    LSSC    ANISO    EXPE
 Data line--- solvent YES
 Data line--- weight    MATRIX 0.31
 Data line--- monitor MEDIUM    torsion 10.0    distance 10.0    angle 10.0
plane 10.0
                chiral 10.0    bfactor 10.0    bsphere 10.0    rbond 10.0    ncsr
10.0
 Data line--- labin  FP=F SIGFP=SIGF    FREE=FreeR_flag
 Data line--- labout  FC=FC FWT=FWT PHIC=PHIC PHWT=PHWT
DELFWT=DELFWT
                PHDELWT=PHDELWT FOM=FOM
 Data line--- temp 0.5 3.0 5.0 8.0 11.0
 Data line--- angle 1.1
 Data line--- chir 1.65
 Data line--- PNAME 8-proc-d003-x003-aew
 Data line--- DNAME p2
 Data line--- RSIZE 80
 Data line--- END
```

10 cycles of TLS refinement and waters are excluded

What to look out for:

- Has TLS been used and how are the groups defined?

  Helps often to improve refinement for low resolution data but is not always effective at high resolution

- Have water molecules be excluded in TLS refinement?

  whether to include waters in refinement or not is still under discussion

TLS input defined in an input file

Range of TLS group, e.g. chain A residues 202 to 286 all atoms

```
TLS
RANGE  'A 202.' 'A 286.' ALL
TLS
RANGE  'A 287.' 'A 499.' ALL
TLS
RANGE  'B 199.' 'B 283.' ALL
TLS
RANGE  'B 284.' 'B 375.' ALL
TLS
RANGE  'B 376.' 'B 498.' ALL
```

Checking last refinement step: log file

- Using a text editor to look for critical values

Ncyc	Rfact	Rfree	FOM	-LL	-LLfree	rmsBOND	zBOND	rmsANGL	zANGL	rmsCHIRAL
$$										
$$										
0	0.2102	0.2508	0.831	539584.	29214.7	0.0	0.0	0.0	0.0	0.0
1	0.1992	0.2426	0.844	536522.	29126.0	0.0	0.0	0.0	0.0	0.0
2	0.1954	0.2400	0.848	535100.	29084.8	0.0	0.0	0.0	0.0	0.0
3	0.1938	0.2388	0.850	534404.	29065.7	0.0	0.0	0.0	0.0	0.0
4	0.1928	0.2382	0.851	533965.	29055.2	0.0	0.0	0.0	0.0	0.0
5	0.1922	0.2378	0.852	533670.	29048.1	0.0	0.0	0.0	0.0	0.0
6	0.1917	0.2374	0.852	533422.	29041.3	0.0	0.0	0.0	0.0	0.0
7	0.1913	0.2372	0.853	533220.	29034.9	0.0	0.0	0.0	0.0	0.0
8	0.1910	0.2370	0.853	533039.	29028.4	0.0	0.0	0.0	0.0	0.0
9	0.1908	0.2368	0.854	532890.	29022.8	0.0	0.0	0.0	0.0	0.0
10	0.1907	0.2367	0.854	532763.	29018.2	0.0241	1.323	1.847	0.888	0.290
11	0.1801	0.2310	0.863	527605.	28871.5	0.0187	0.905	1.657	0.769	0.244
12	0.1742	0.2275	0.869	524582.	28779.9	0.0173	0.772	1.600	0.722	0.216
13	0.1705	0.2248	0.873	522668.	28714.7	0.0169	0.730	1.597	0.708	0.195
14	0.1682	0.2230	0.875	521344.	28669.3	0.0166	0.710	1.611	0.706	0.176
15	0.1666	0.2217	0.878	520468.	28636.9	0.0164	0.696	1.627	0.707	0.160
16	0.1654	0.2207	0.879	519788.	28609.5	0.0162	0.686	1.639	0.708	0.145
17	0.1646	0.2200	0.880	519313.	28589.8	0.0160	0.678	1.645	0.708	0.133
18	0.1639	0.2194	0.881	518947.	28575.3	0.0160	0.673	1.649	0.708	0.124
19	0.1635	0.2189	0.882	518680.	28563.9	0.0159	0.670	1.649	0.708	0.118
20	0.1631	0.2185	0.882	518453.	28553.4	0.0158	0.667	1.648	0.707	0.112
$$										

$TEXT:Result: $$ Final results $$

	Initial	Final
R factor	0.2102	0.1631
R free	0.2508	0.2185
Rms BondLength	0.0241	0.0158
Rms BondAngle	1.8468	1.6479
Rms ChirVolume	0.2898	0.1124

What to look out for:

- Do Rfactor and Rfree of last refinement step make sense?

  A gap of about 5% between the two is often desired

- How big is FOM?

  The higher the better the model fits the data

- How big is RMSD bond?

  Value smaller than 0.020 is desired (in general dependent on resolution)

Checking last refinement step

- Using a text editor to look for critical values

What to look out for:

- Is there a ligand present and is the dictionary for it given?

  Ligands which are not in the standard dictionary need to be described by their own dictionary file

- Is a MAKE_U_POSITIVE problem present and is the refinement stable?

  U tensor for temperature factors is not correctly defined in refinement; this needs to be avoided

Contents of a CIF file; gives restraints and dictionary information for ligands

```
data_comp_DRG
#
loop_
_chem_comp_atom.comp_id
_chem_comp_atom.atom_id
_chem_comp_atom.type_symbol
_chem_comp_atom.type_energy
_chem_comp_atom.partial_charge
 DRG      CAP   C   CH2     0.049
 DRG      1HAP  H   HCH2    0.000
 DRG      2HAP  H   HCH2    0.000
 DRG      CAN   C   CH2     0.049
 DRG      1HAN  H   HCH2    0.000
 DRG      2HAN  H   HCH2    0.000
 DRG      NAU   N   NT2     0.316
 DRG      HAA   H   HNT2    0.021
 DRG      HAB   H   HNT2    0.021
 DRG      CAO   C   CH2     0.049
 DRG      1HAO  H   HCH2    0.000
 DRG      2HAO  H   HCH2    0.000
```

Overall weighting term for data

Weighting term for data and Bfactor is not properly chosen

```
CGMAT cycle number =     7
weigth matrix   0.8000001
function value    195881.0
-------------------------------------------------------------------------------
        Restraint type          N restraints  Rms Delta  Av(Sigma)
Bond distances: refined atoms        975     0.015     0.022
Bond distances: others               653     0.002     0.020
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Overall          : scale =   0.826, B =   0.066
Partial structure   1: scale =   0.424, B =  30.042
Overall anisotropic scale factors
  B11 = -0.14 B22 =  0.62 B33 = -0.48 B12 =  0.00 B13 =  0.00 B23 =  0.00
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
Overall R factor              =    0.1364
Free R factor                 =    0.1655
Overall figure of merit       =    0.9201
-------------------------------------------------------------------------------
 Trying gamma equal   2.3013467E-03
 Gamma decreased to    0.000000
```

determines

Geometric RMSDs

Problem in MAKE_U_POSITIVE 2.5330290E-02 2.5330294E-02

Checking water molecules

- Make sure refinement excluded water molecules
- Run actit.csh
  - Script is based on ACT from CCP4 package and can be found here:

  ftp://hestia.sgc.ox.ac.uk/pub/px
- Input

  actit.csh <pdbin>

  where

  <pdbin> is the PDB file you want to check the waters in
- Output

```
khan:/work/ACVR1A/27-proof-d003-K00991-x003/proof-build-mv 127: actit.csh final_tlsanl1.pdb
   0.0 -  5.0      0
   5.0 - 10.0      0
  10.0 - 15.0     10 **
  15.0 - 20.0     32 ********
  20.0 - 25.0     84 ********************
  25.0 - 30.0    117 ****************************
  30.0 - 35.0    163 ***************************************
  35.0 - 40.0    209 **************************************************
  40.0 - 45.0    206 *************************************************
  45.0 - 50.0    137 *********************************
  50.0 - 55.0     49 ************
  55.0 - 60.0      2
  60.0 - 65.0      0
  65.0 - 70.0      0
  70.0 - 75.0      0
  75.0 - 80.0      0
  80.0 - 85.0      0
  85.0 - 90.0      0
  90.0 - 95.0      0
  95.0 - 100.0     0
  .gt. 100.0       0
```

Gaussian distribution of water molecules

What to look out for:

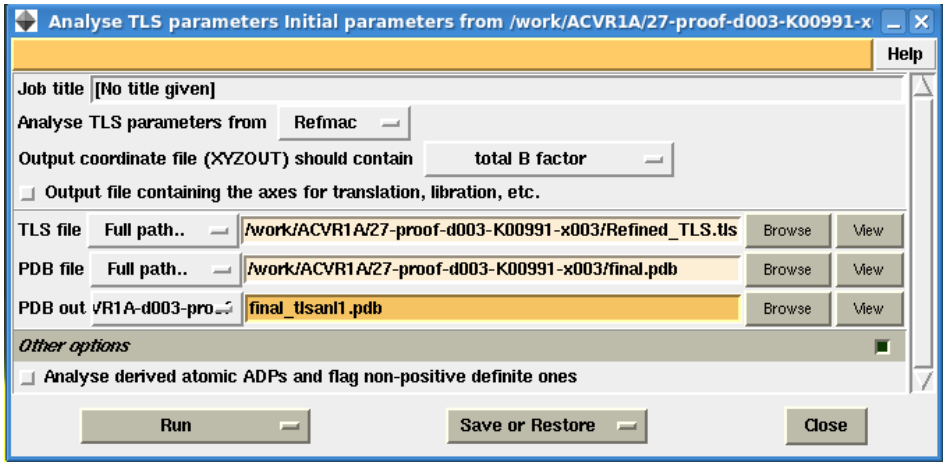
- Suspicious water molecules with high temperature factors

List of water molecules with Bfactors higher than 50

```
O     HOH  50W   50.03
O     HOH ***W   50.08
O     HOH 677W   50.11
O     HOH 423W   50.19
O     HOH 779W   50.22
O     HOH 662W   50.23
O     HOH 953W   50.23
O     HOH 399W   50.28
O     HOH 660W   50.29
O     HOH 778W   50.32
O     HOH ***W   50.37
O     HOH 534W   50.40
O     HOH 499W   50.43
O     HOH 424W   50.65
O     HOH ***W   50.68
O     HOH 312W   50.71
O     HOH 287W   50.73
O     HOH 625W   50.75
O     HOH 425W   50.86
O     HOH 981W   50.86
O     HOH 497W   50.92
O     HOH 759W   50.94
O     HOH ***W   51.06
O     HOH 850W   51.13
O     HOH 866W   51.23
O     HOH ***W   51.24
O     HOH  18W   51.30
O     HOH 856W   51.37
O     HOH 294W   51.49
O     HOH 966W   51.54
O     HOH 980W   51.54
O     HOH 714W   51.59
O     HOH 879W   51.71
O     HOH 863W   51.78
O     HOH 913W   51.81
O     HOH 838W   52.01
O     HOH 334W   52.19
O     HOH 828W   52.19
O     HOH 709W   52.21
O     HOH 252W   52.41
O     HOH ***W   52.45
O     HOH 878W   52.60
O     HOH ***W   52.69
O     HOH 771W   52.76
O     HOH 487W   52.81
O     HOH 698W   53.60
O     HOH 776W   53.92
O     HOH 702W   53.97
O     HOH 359W   55.00
O     HOH 704W   55.04
O     HOH 590W   57.19
```

Checking Bfactors and close atoms

- TLSanl
  - Within CCP4 as stand-alone
  - Get complete Bfactor of the model
- Input
  - TLS output file from refinement
  - PDB file of final model
- Output
  - PDB file with complete Bfactors




```
CRYST1  83.600   98.700   83.850  90.00 117.42  90.00 P 1 21 1
SCALE1     0.011962  0.000000  0.006206        0.00000
SCALE2     0.000000  0.010132  0.000000        0.00000
SCALE3     0.000000  0.000000  0.013435        0.00000
ATOM      1  N   ARG A 202      11.597  -4.441  -9.468  1.00 53.89           N
ATOM      2  CA  ARG A 202      12.056  -3.528  -8.374  1.00 59.23           C
ATOM      4  CB  ARG A 202      12.087  -2.077  -8.873  1.00 55.23           C
ATOM     13  C   ARG A 202      11.164  -3.634  -7.123  1.00 59.20           C
ATOM     14  O   ARG A 202       9.995  -3.215  -7.155  1.00 58.95           O
ATOM     18  N   THR A 203      11.719  -4.206  -6.042  1.00 57.09           N
ATOM     19  CA  THR A 203      11.054  -4.288  -4.717  1.00 54.92           C
ATOM     21  CB  THR A 203      10.343  -5.649  -4.477  1.00 57.16           C
ATOM     23  OG1 THR A 203      11.312  -6.645  -4.125  1.00 63.26           O
ATOM     25  CG2 THR A 203       9.556  -6.101  -5.719  1.00 68.24           C
ATOM     29  C   THR A 203      12.061  -4.080  -3.572  1.00 45.50           C
```

Bfactors before running TLSanl

```
CRYST1  83.600   98.700   83.850  90.00 117.42  90.00 P 1 21 1
SCALE1     0.011962  0.000000  0.006206        0.00000
SCALE2     0.000000  0.010132  0.000000        0.00000
SCALE3     0.000000  0.000000  0.013435        0.00000
ATOM      1  N   ARG A 202      11.597  -4.441  -9.468  1.00 65.78           N
ANISOU    1  N   ARG A 202     8044   8737   8213    977    -50    146       N
ATOM      2  CA  ARG A 202      12.056  -3.528  -8.374  1.00 70.65           C
ANISOU    2  CA  ARG A 202     8678   9258   8909    915     13    198       C
ATOM      4  CB  ARG A 202      12.087  -2.077  -8.873  1.00 67.23           C
ANISOU    4  CB  ARG A 202     8253   8817   8473    952     54    308       C
ATOM     13  C   ARG A 202      11.164  -3.634  -7.123  1.00 70.10           C
ANISOU   13  C   ARG A 202     8594   9107   8932    860      5    145       C
ATOM     14  O   ARG A 202       9.995  -3.215  -7.155  1.00 70.11           O
ANISOU   14  O   ARG A 202     8577   9098   8963    883    -18    143       O
ATOM     18  N   THR A 203      11.719  -4.206  -6.042  1.00 67.32           N
ANISOU   18  N   THR A 203     8250   8705   8623    792     22    102       N
```

Bfactors after running TLSanl

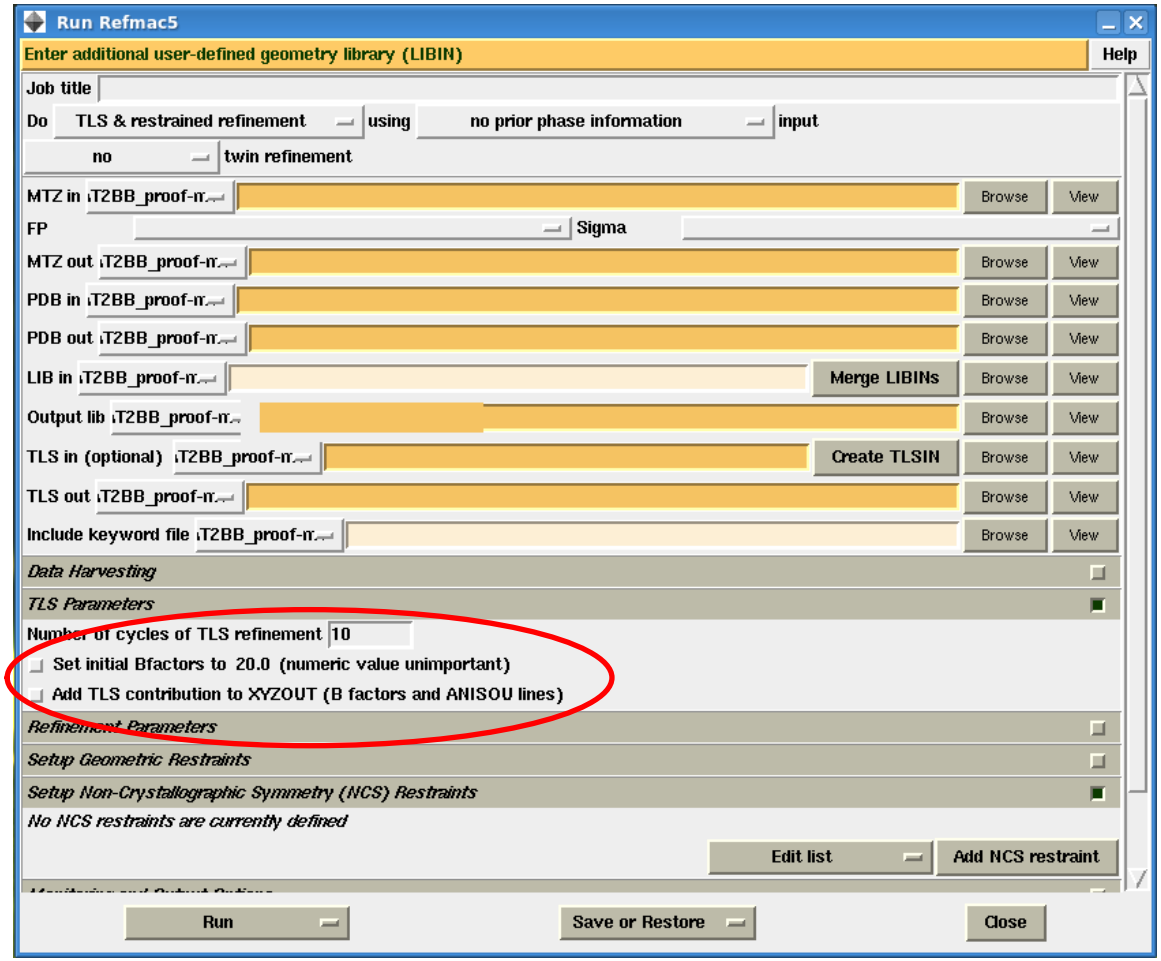
Checking Bfactors and close atoms

- TLSanl
  - Within CCP4 as part of Refmac5
  - Is then run when using TLS in refinement
  - Get complete Bfactor of the model

TLS group definition:

- By hand
- Using TLS MD server

http://skuld.bmsc.washington.edu/~tlsmd/



What to look out for:

- High Bfactors give information about flexible/unordered/badly build areas in the structure

Checking Bfactors and close atoms

- Moleman2
  - Download and more information can be found here: http://xray.bmc.uu.se/usf/xutil.html
  - Get started by loading the PDB file of interest
    → MOLEMAN2 > re <pdbin.pdb>
  - Check Bfactors for PDB file
    → MOLEMAN2 > bf st

See manual for more possibilities

Chain name	Atoms	Bave	Bsdv	Bmin	Bmax	Brms	Bharm.av
A<->	2422	26.68	12.72	8.26	105.63	29.55	22.84
B<->	2408	36.54	18.45	11.51	108.90	40.93	29.65
C<->	2452	25.09	13.46	8.40	102.33	28.47	20.32
D<->	2460	24.99	14.60	8.48	105.86	28.94	20.06
W<->	1009	36.47	8.91	12.20	57.19	37.54	33.79
E<->	124	22.80	12.47	9.81	69.77	25.99	19.05
I<->	68	35.00	7.85	13.30	50.48	35.86	32.90
J<->	143	42.71	8.10	19.20	63.66	43.47	40.85
F<->	13	33.10	11.69	9.77	57.20	35.10	27.47

Protein chains

Water chain

Other ligands

High Bfactor indicate badly ordered areas in protein chains

What to look out for:

- High Bfactors give information about flexible/unordered/badly build areas in the structure
- High temperature factors for ligands → ligand is likely not to fit the density, e.g. salt ion where there should be water

Checking Bfactors and close atoms

- Moleman2
  - Download and more information can be found here: http://xray.bmc.uu.se/usf/xutil.html
  - Check duplicated atoms and errors in occupancy for PDB file

    → MOLEMAN2 > pd sa

See manual for more possibilities

Coordinates for CA of Ser in two conformations

Occupancy

Bfactor

```
ERROR   - ALTERNATE LOCATIONS IN IDENTICAL POSITIONS
the following two atoms have alternate locations
but their distance is only  0.02 A, suggesting
that they are in identical positions in space and
should be merged into a single location:
ATOM   577  CA ASER A 239      19.651  -7.171  19.092  0.50 30.68           C
ATOM   578  CA BSER A 239      19.655  -7.168  19.073  0.50 30.70           C

WARNING - OCCUPANCIES DO NOT SUM TO 1.00
for the following atom, the occupancies of the   1
alternate locations add up to  0.70 instead of
1.00 (this can be okay if you are sure that the
atom has partial occupancy, or if it lies in a
special position, such as on a twofold axis):
ATOM 23216  CGC FLC F   1      -3.969 -14.778  27.908  0.70 20.62           C
```

Incomplete occupancy

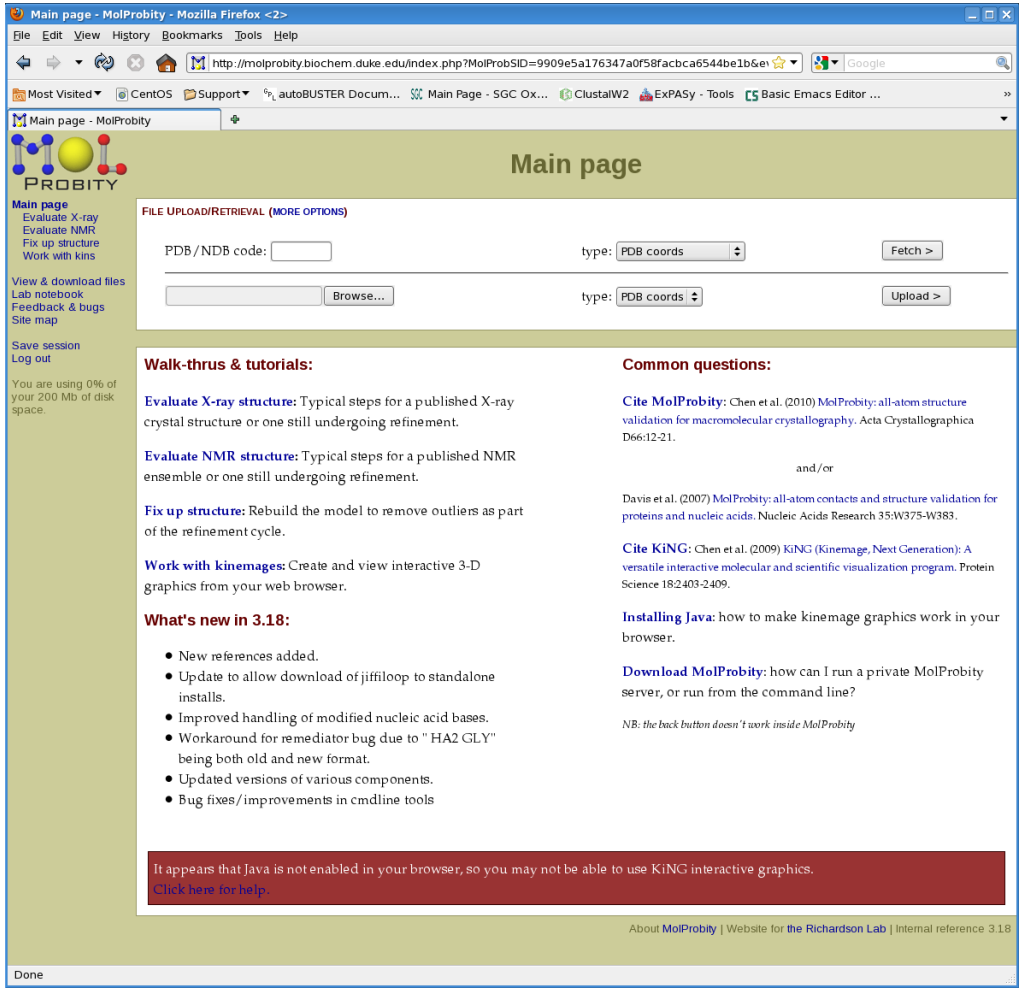
What to look out for:

- Duplicated atoms and which can maybe be merged
- Occupancy has to add up to 1
- Big differences in Bfactors between atoms of alternative conformations shows that occupancy needs to be assigned in different ratio

Checking geometry and more

- Molprobity
  - The web server can be found here:

    http://molprobity.biochem.duke.edu/
  - Works for PDB files refined with most common refinement programs



What to look out for:

- Ramachandran outlier
- Outlier in bond length and bond angle
- Bad rotamers
- Outliers in Cβ atom chirality
- Amount of serious clashes between residues but also between residues and ligand/water

Main page and input

Checking geometry and more

- Molprobity
  - The web server can be found here:

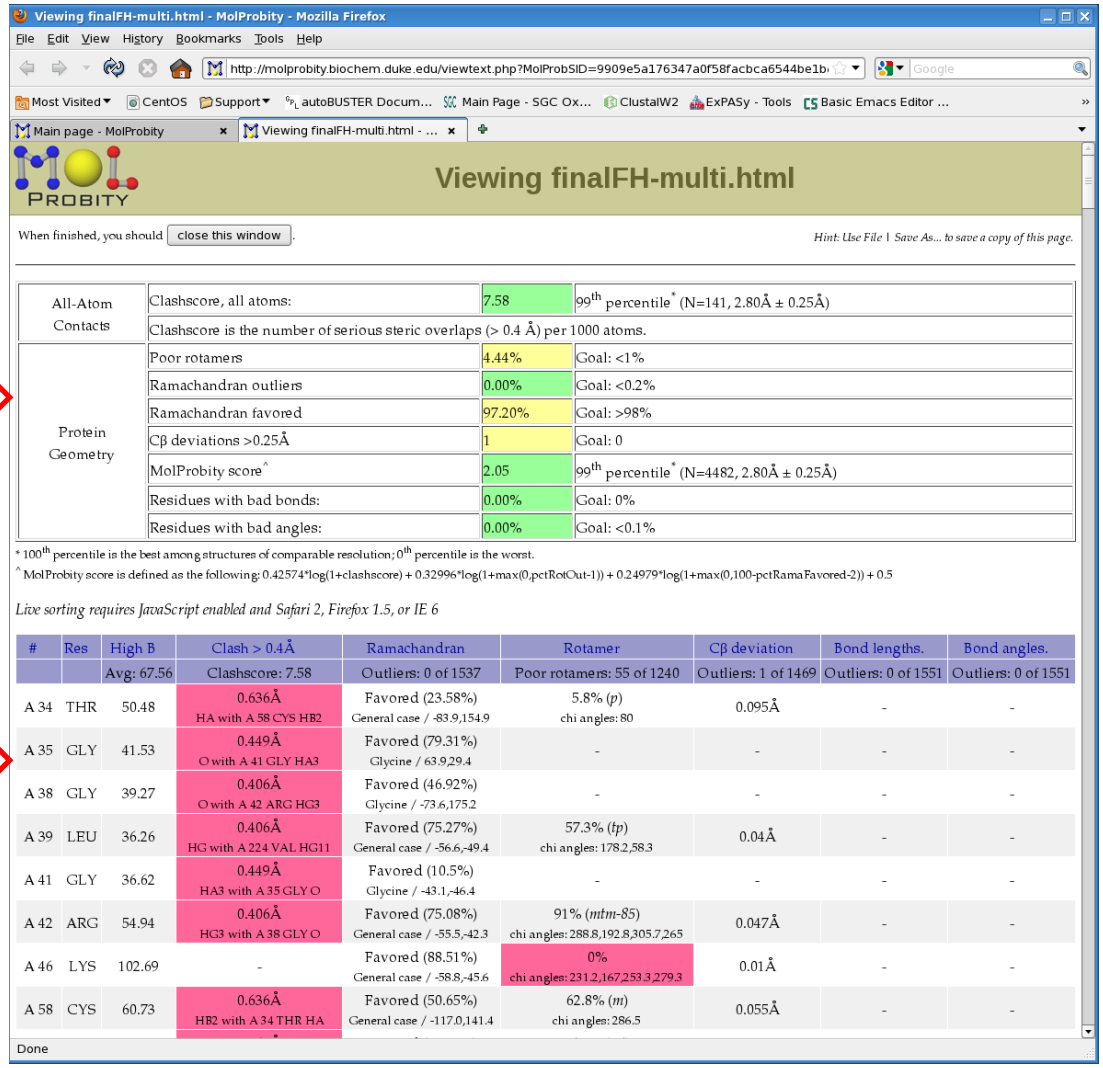
http://molprobity.biochem.duke.edu/

HTML version of geometry analysis

Overall statistics;
Green: accepted
Yellow: Ok but can be improved
Red: Failed

Detailed statistics given for each residue

Click on column label to list from worst to best



Viewing finalFH-multi.html - MolProbity - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://molprobity.biochem.duke.edu/viewtext.php?MolProbSID=9909e5a176347a0f58facbca6544be1b

Most Visited CentOS Support autoBUSTER Docum... Main Page - SGC Ox... ClustalW2 ExPASy - Tools Basic Emacs Editor ...

Main page - MolProbity | Viewing finalFH-multi.html - ...

PROBITY

Viewing finalFH-multi.html

When finished, you should close this window.

Hint: Use File | Save As... to save a copy of this page.

All-Atom Contacts	Clashscore, all atoms:	7.58	99th percentile* (N=141, 2.80Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	4.44%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	97.20%	Goal: >98%
	Cβ deviations >0.25Å	1	Goal: 0
	MolProbity score^	2.05	99th percentile* (N=4482, 2.80Å ± 0.25Å)
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: <0.1%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

^ MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5

Live sorting requires JavaScript enabled and Safari 2, Firefox 1.5, or IE 6

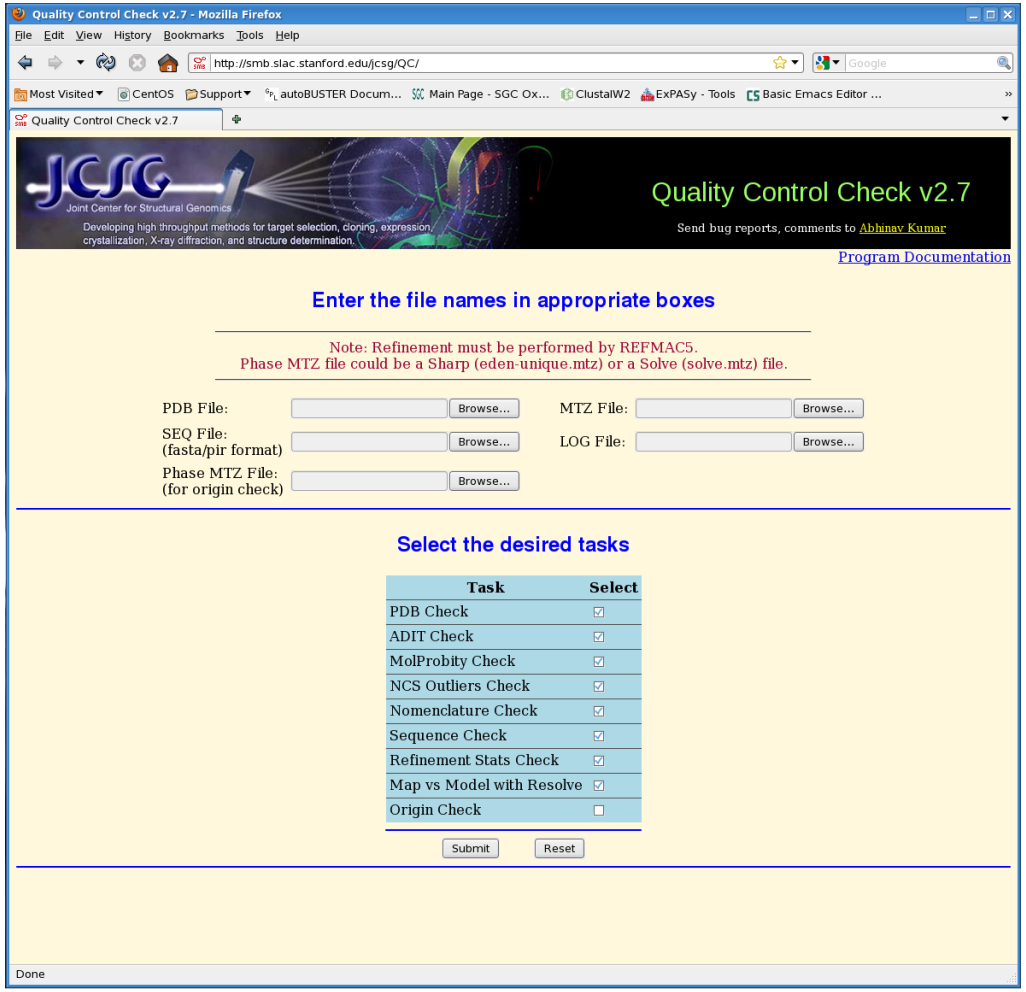
#	Res	High B	Clash > 0.4Å	Ramachandran	Rotamer	Cβ deviation	Bond lengths.	Bond angles.
		Avg: 67.56	Clashscore: 7.58	Outliers: 0 of 1537	Poor rotamers: 55 of 1240	Outliers: 1 of 1469	Outliers: 0 of 1551	Outliers: 0 of 1551
A 34	THR	50.48	0.636Å HA with A 58 CYS HB2	Favored (23.58%) General case / -83.9,154.9	5.8% (p) chi angles: 80	0.095Å	-	-
A 35	GLY	41.53	0.449Å O with A 41 GLY HA3	Favored (79.31%) Glycine / 63.9,29.4	-	-	-	-
A 38	GLY	39.27	0.406Å O with A 42 ARG HG3	Favored (46.92%) Glycine / -73.6,175.2	-	-	-	-
A 39	LEU	36.26	0.406Å HG with A 224 VAL HG11	Favored (75.27%) General case / -56.6,-49.4	57.3% (tp) chi angles: 178.2,58.3	0.04Å	-	-
A 41	GLY	36.62	0.449Å HA3 with A 35 GLY O	Favored (10.5%) Glycine / -43.1,-46.4	-	-	-	-
A 42	ARG	54.94	0.406Å HG3 with A 38 GLY O	Favored (75.08%) General case / -55.5,-42.3	91% (mtm-85) chi angles: 288.8,192.8,305.7,265	0.047Å	-	-
A 46	LYS	102.69	-	Favored (88.51%) General case / -58.8,-45.6	0% chi angles: 231.2,167,253.3,279.3	0.01Å	-	-
A 58	CYS	60.73	0.636Å HB2 with A 34 THR HA	Favored (50.65%) General case / -117.0,141.4	62.8% (m) chi angles: 286.5	0.055Å	-	-

Done

Checking geometry and more

- Quality Control Check from JCSG
  - The web server can be found here:

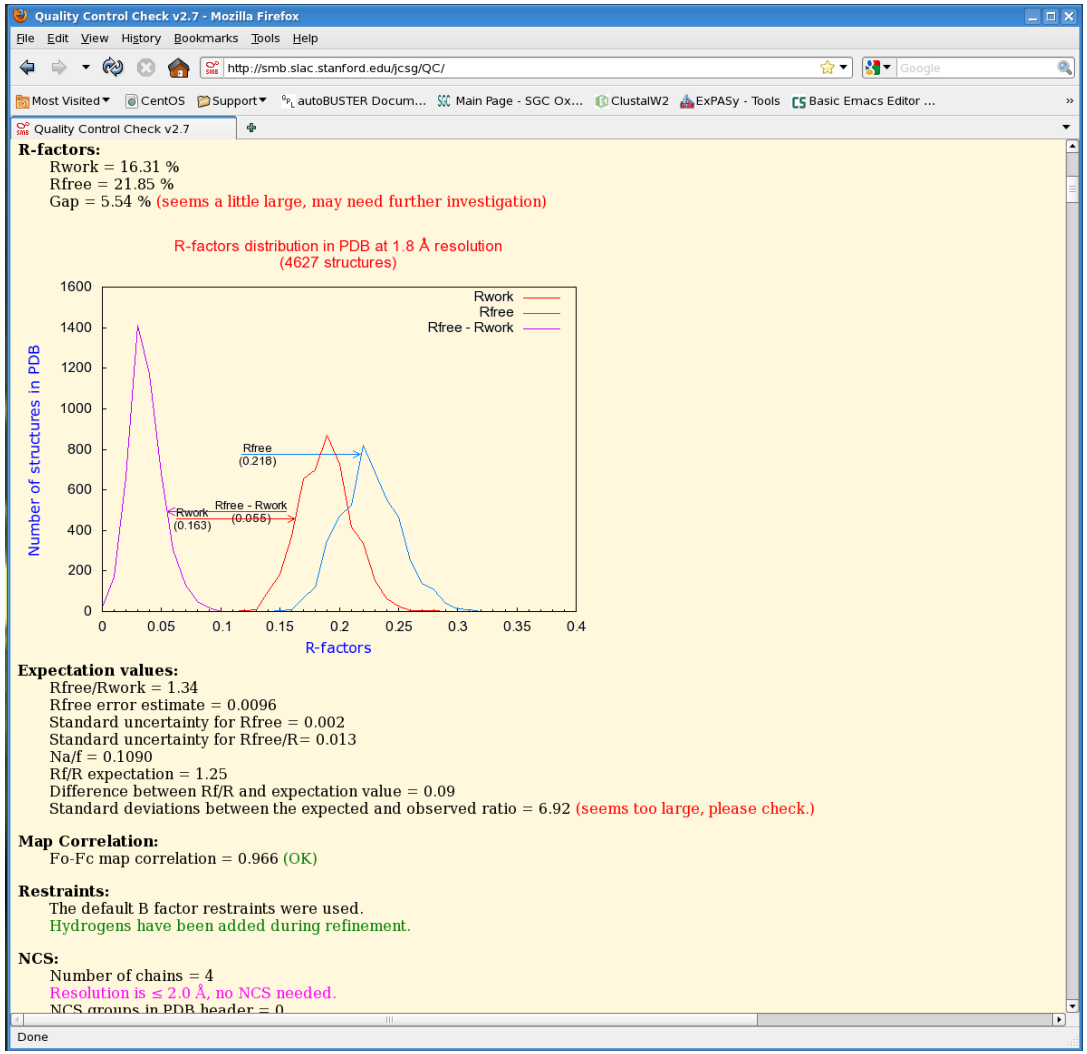
  http://smb.slac.stanford.edu/jcsg/QC/
  - Needs PDB file, MTZ file and log file from last refinement step, protein sequence
  - Only works for files refined with Refmac5



Checking geometry and more

- Quality Control Check from JCSG
  - The web server can be found here:

http://smb.slac.stanford.edu/jcsg/QC/

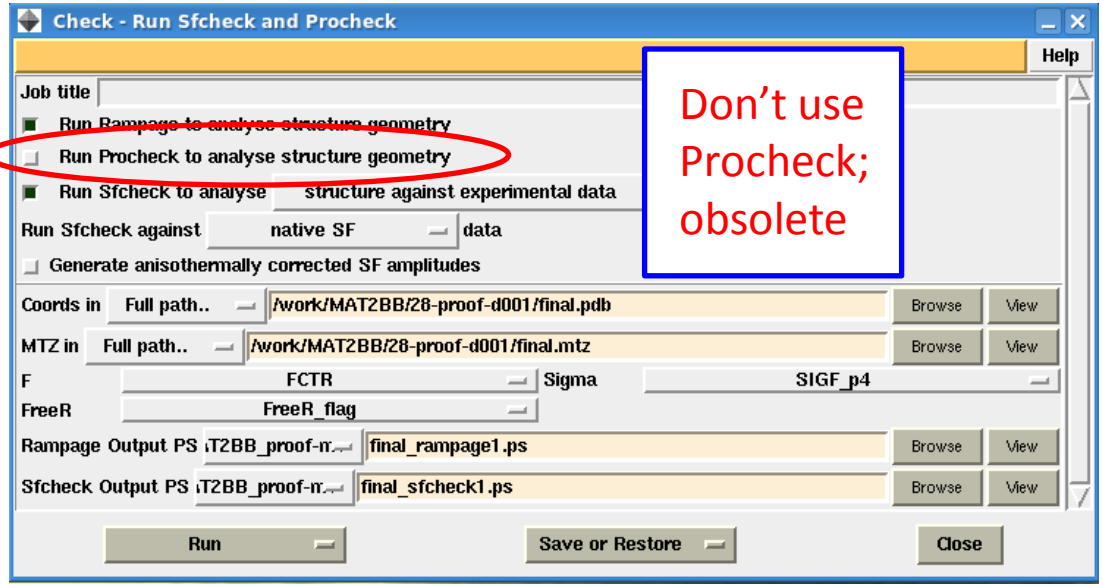


What to look out for:

- Geometry problems as in Molprobity check
- Flags on data quality
- Refinement settings
- Water molecule and ligand geometry
- Clashes between side chains or side chains and waters, side chains and ligand molecules, side chains and back bone …
- Sequence errors
- Unusual Bfactors and occupancy

Checking geometry and more

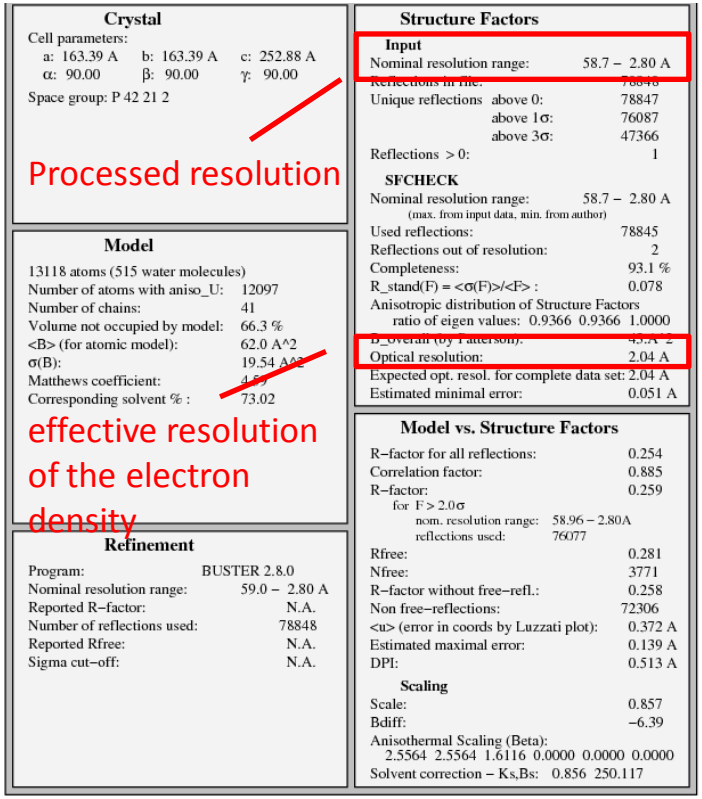
- Quality Control Check
  - Within CCP4
- Input
  - Refined PDB file
  - Refined MTZ file
- Output
  - Logfile with text presentation of results
  - Postscript files giving graphical presentation of results



Don't use Procheck; obsolete

What to look out for:

- Geometry
- Flags on data quality
- Refinement settings
- Water molecule and ligand geometry
- Clashes
- Sequence errors
- Unusual Bfactors and occupancy



Checking sequence

- Blast search to get reference sequence
  - Get sequence from model

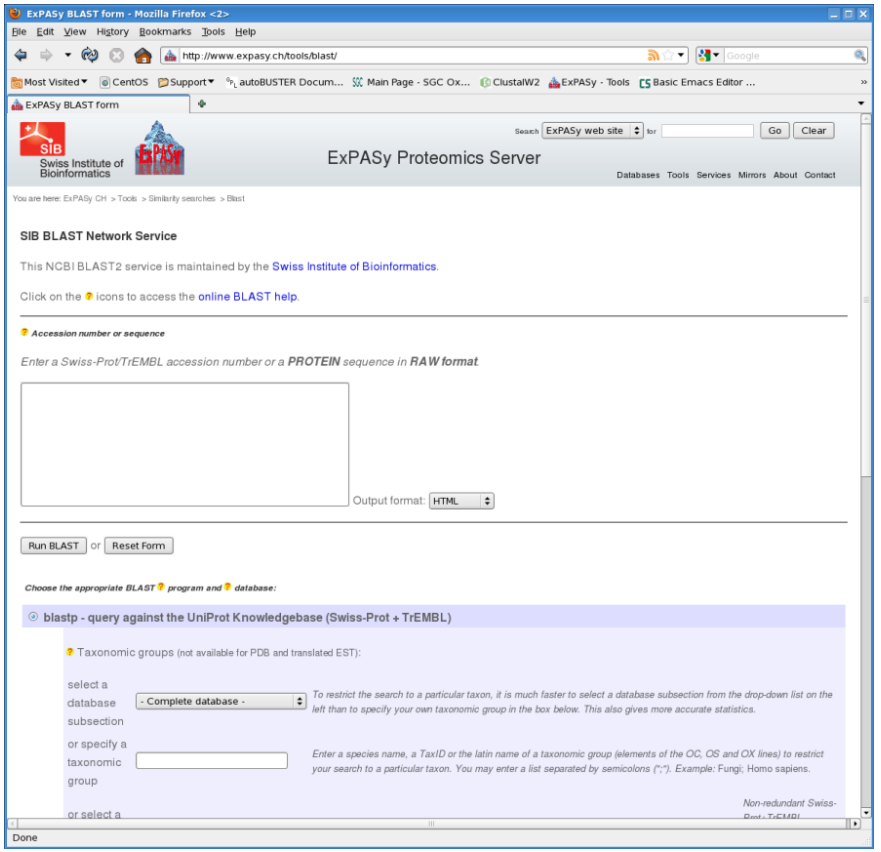
    phenix.print_sequence <PDBin.pdb> > <PDBin.seq>

    <PDBin.pdb> structure to get the sequence from

    <PDBin.seq> new file with printed sequence
  - The web server can be found here:

    http://expasy.org/tools/blast/
  - Get sequence from each chain of target protein in the ASU
  - Align the sequences of target protein with reference sequence
  - Sequence alignment tools

    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/



What to look out for:

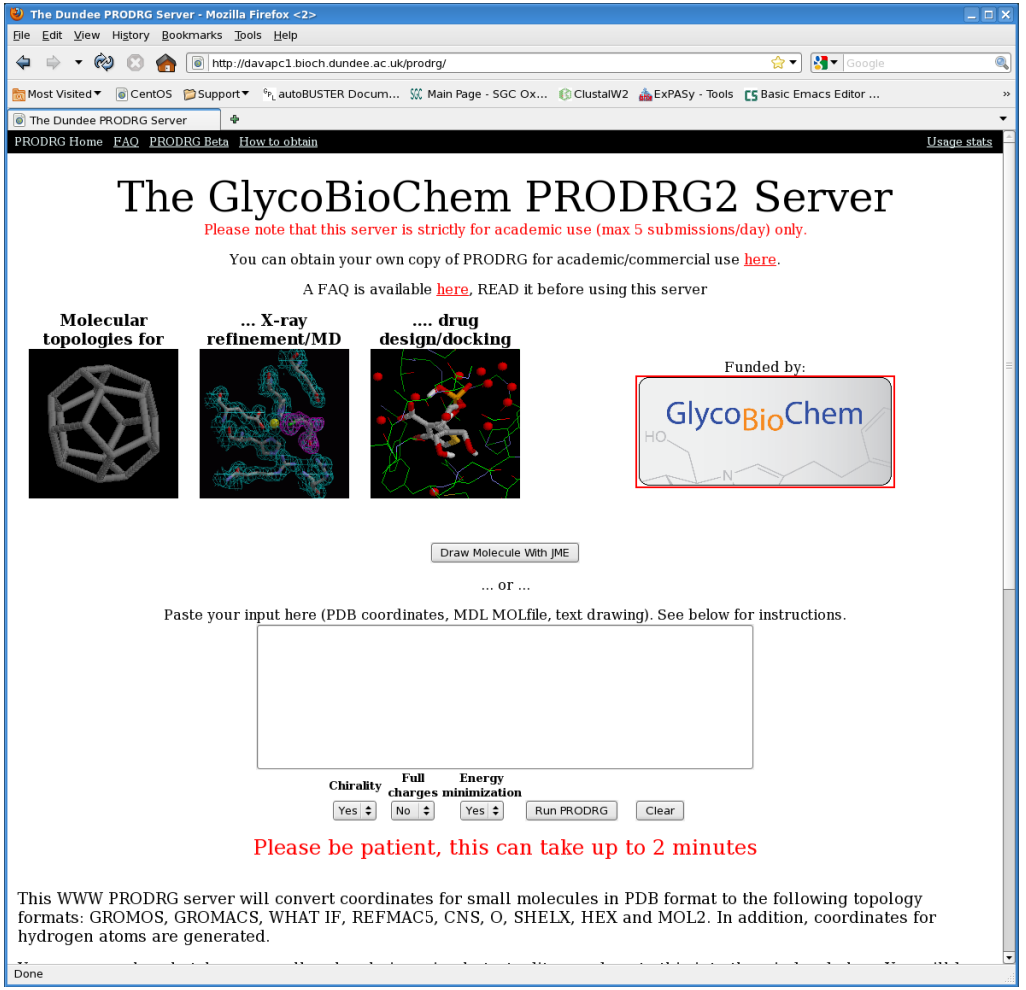
- Mismatch between reference sequence and model sequence due to building
- Mismatch between reference sequence and model sequence due to mutations
- Mismatch in residue numbering

Checking ligand geometry

- Turn non-standard ligand into PDB coordinates
  - The web server can be found here:

http://davapc1.bioch.dundee.ac.uk/prodrg/

http://www.ccdc.cam.ac.uk/support/documentation/mogul/mogul/mogul.3.3.html



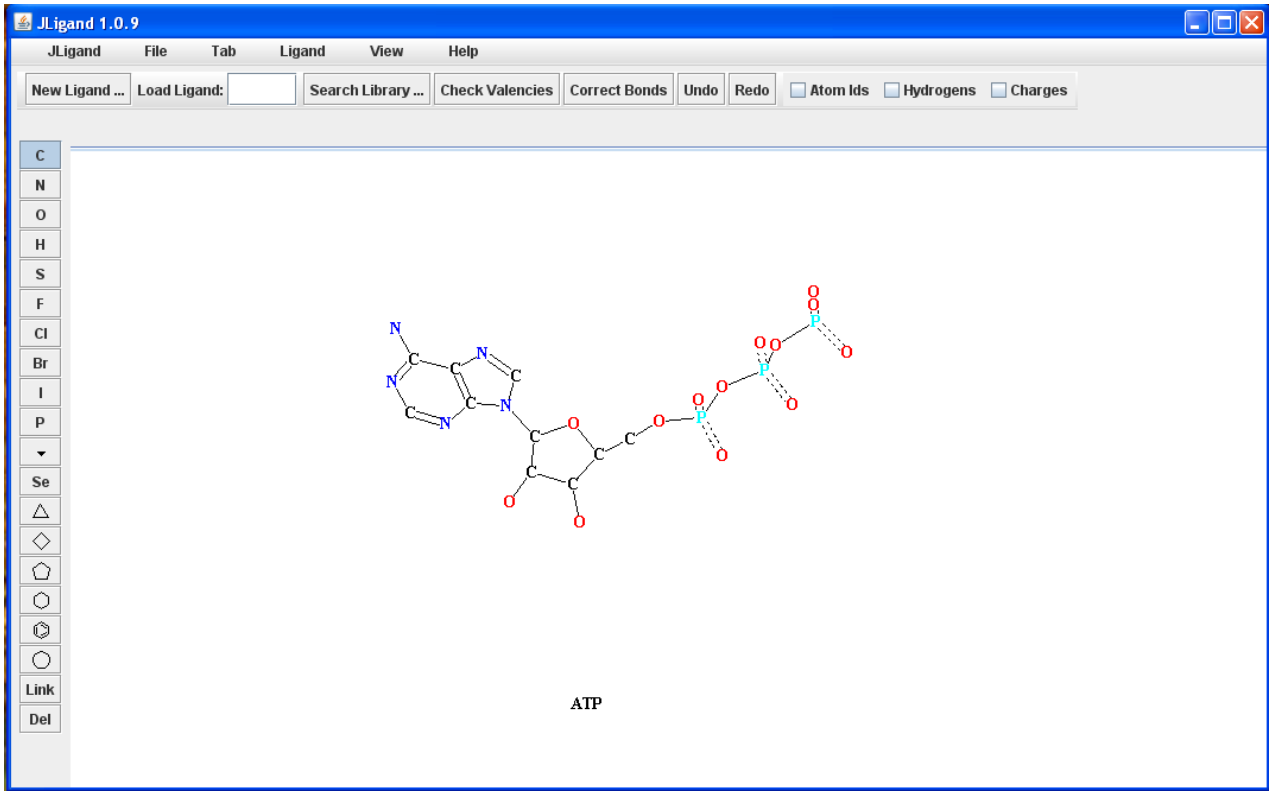
What to look out for:

- Correct geometry for ligand
- Correct bond length, bond angle, valence state
- Single, double, triple bond

Checking ligand geometry

- Turn non-standard ligand into PDB coordinates
  - software can be found here:

  http://www.ysbl.york.ac.uk/~pyoung/JLigand/JLigandInfo.html



What to look out for:

- Correct geometry for ligand
- Correct bond length, bond angle, valence state
- Single, double, triple bond

Walk structure

- Create “ghost” PDB structures of NCS chains
  - Script can be found here:

  ftp://hestia.sgc.ox.ac.uk/pub/px
- Input

  allOnChain.csh <pdbin> <first residue> <last residue> <chainID>

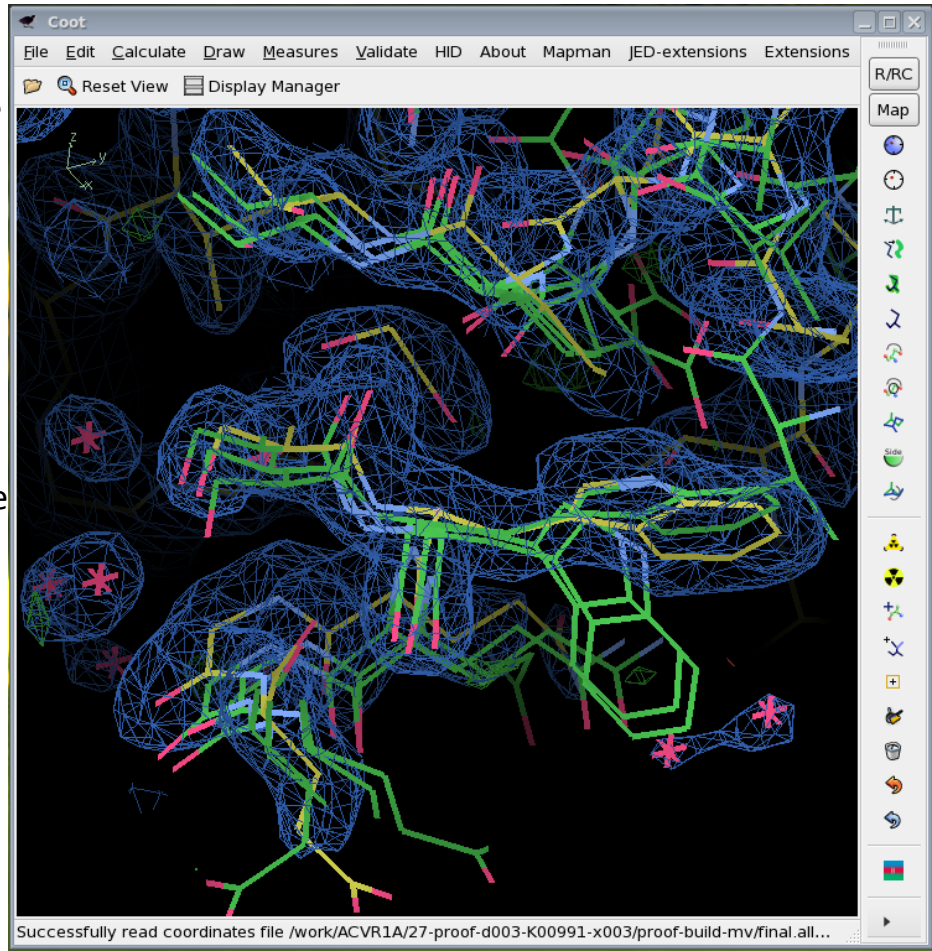
  where

  <pdbin> the PDB file you want to check the waters in

  <first residue> first common residue in all NCS chains of the model

  <last residue> last common residue in all NCS chains of the model

  <chainID> unique chain identifier
- Output
  - PDB file for each unique chain
  - NCS copies are superpositioned and last common atom for all chains is merged
  - Shows detailed differences in rotamers and backbone trace
  - Gives a full PDB file and not just a ghost image as in Coot



What to look out for:

- Differences in NCS chains
- Check if they are supported by density

Walk structure: overview

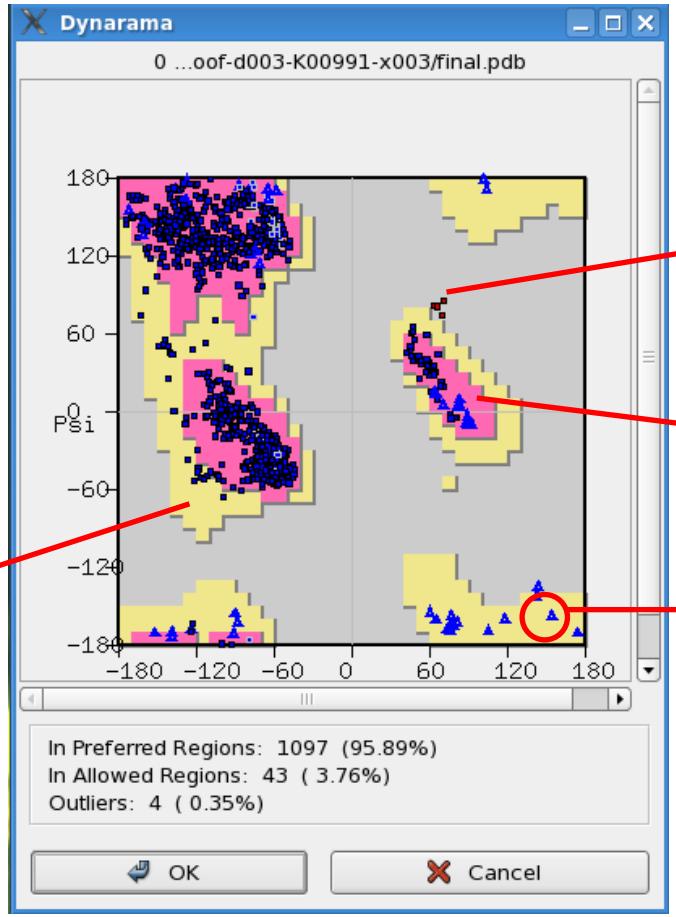
- Coot for visual inspection of structure
- Files needed:
  - Final PDB file to be checked
  - MTZ file from last refinement step
  - PDB files created by allOnChain.csh
- Create a views record file
  - This is a possibility to track changes suggested by the proofreader
  - Coot → Extensions → Views → Add view
    Coot → Extensions → Views → Save views
    Coot → Calculate → Run script → pick myproof-views.src
    Coot → Extensions → Views → Views panel
- Open Ramachandran plot
  - Coot → Validate → Ramachandran Plot
- Open Kleywegt plot
  - Coot → Validate → Kleywegt Plot
- Walk structure residue by residue for each unique chain
- Check waters from actit.csh and
  - Coot → Validate → Check/Delete Waters
- Check rotamers
  - Coot → Validate → Rotamer analysis
- Side chain building
  - Coot → Validate → Density fit analysis
- Check unmodeled density blobs
  - Coot → Validate → Difference Map Peaks

Walk structure: Ramachandran Plot

- Coot for visual inspection of structure
- Open Ramachandran plot
  - Coot → Validate → Ramachandran Plot

What to look out for:

- Ramachandran outlier
- And those in allowed region



Ramachandran outlier

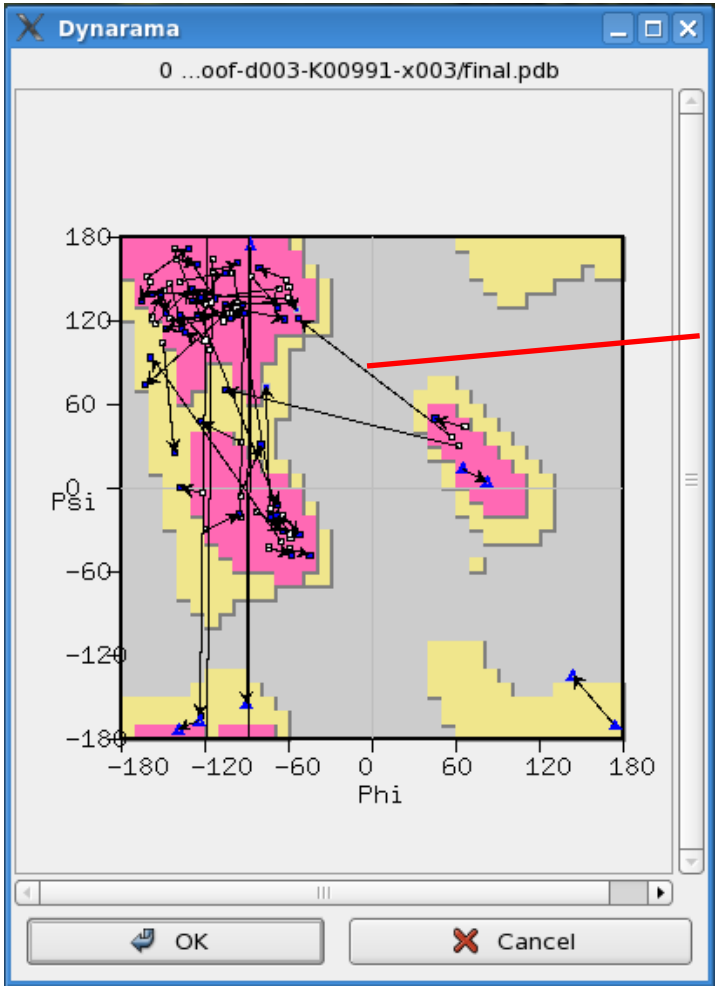
Favoured region

Allowed region

Glycine residues

Walk structure: Kleywegt Plot

- Coot for visual inspection of structure
- Open Kleywegt plot
  - Coot → Validate → Kleywegt Plot



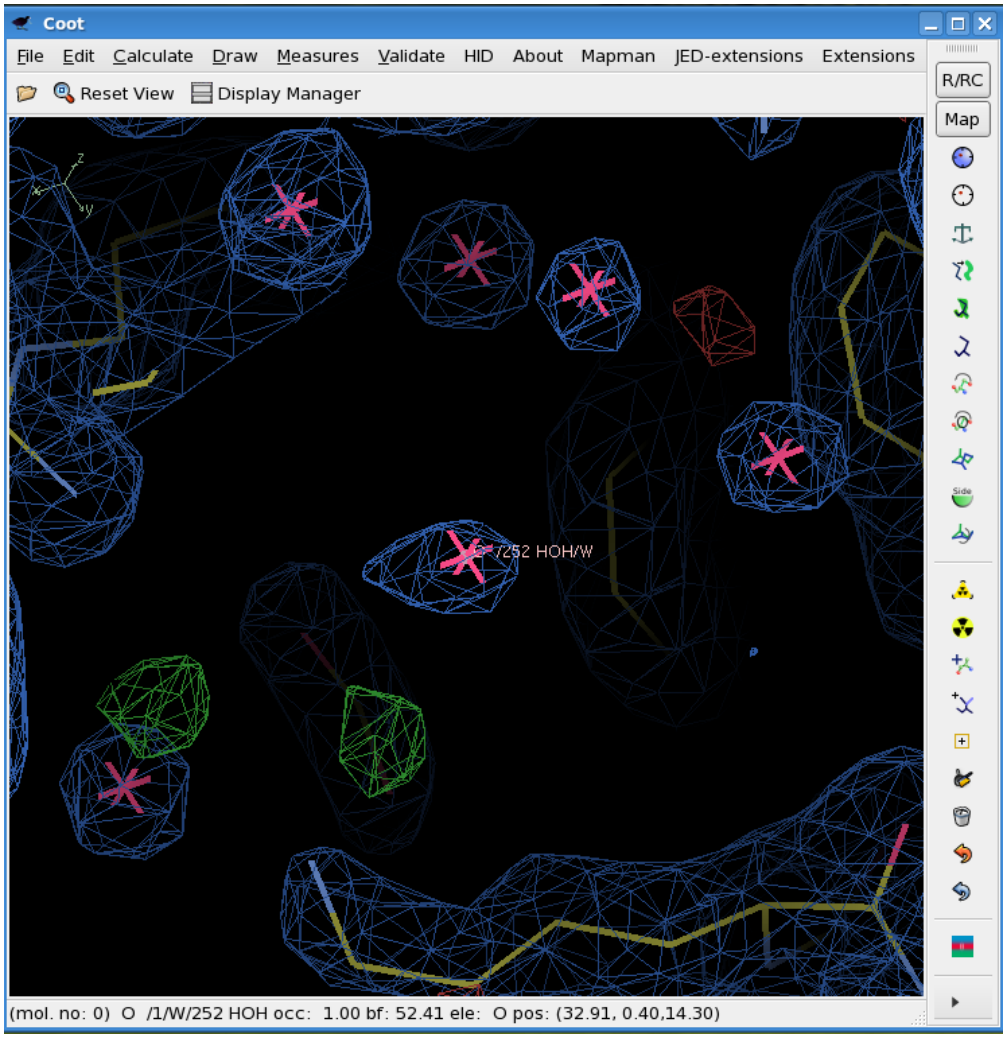
Deviation between the same amino acid in chain A and chain B

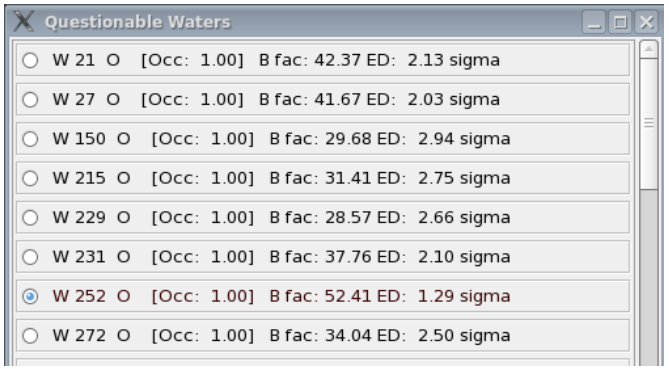
What to look out for:

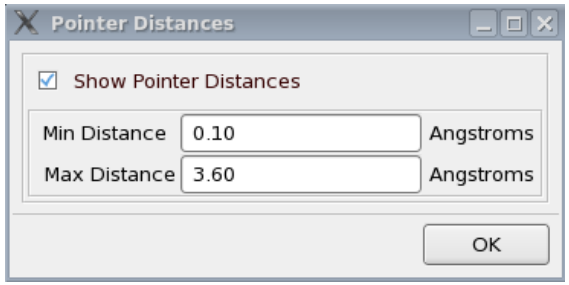
- Ramachandran deviation for the same residue but compared between two NCS molecules
- The longer the arrow the bigger the deviation
- Ramachandran plot of back bone should be as close as possible between NCS copies; side chains may vary

Walk structure: waters

- Coot for visual inspection of structure
- Check waters from actit.csh and
  - Coot → Validate → Check/Delete Waters







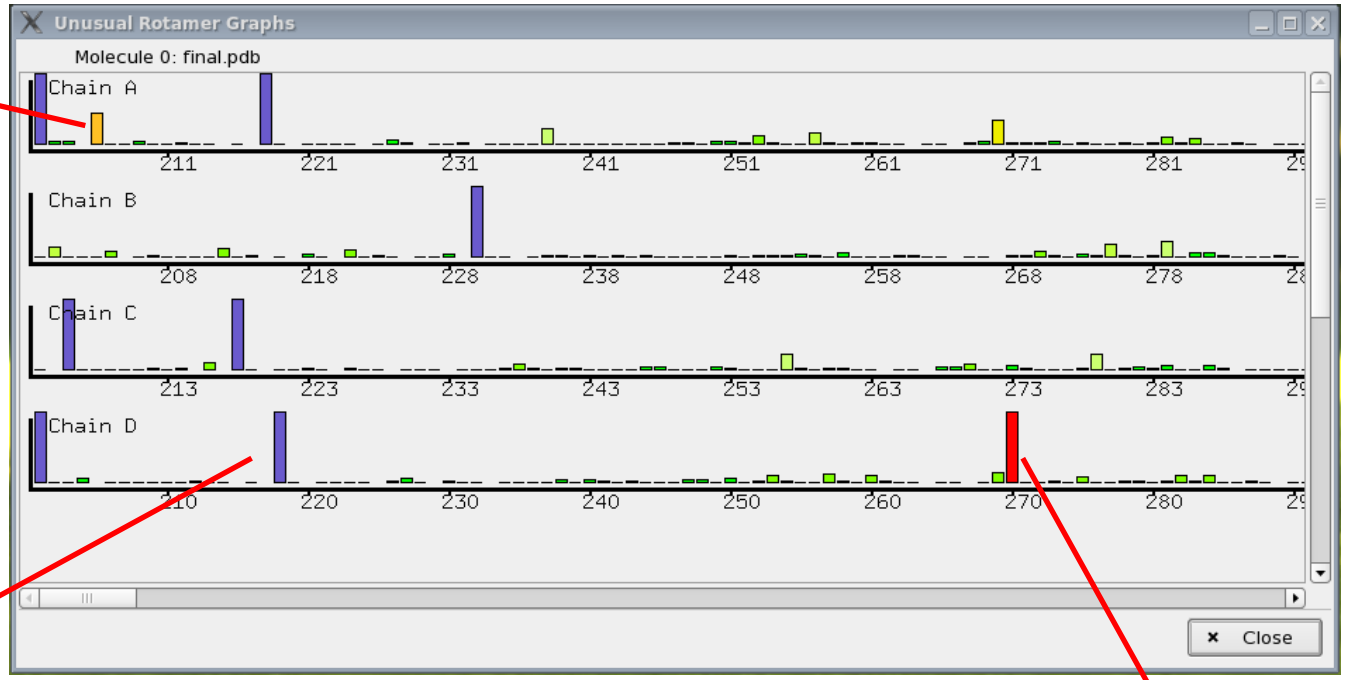
What to look out for:

- Coordination of water molecules
- Bfactors

Walk structure: rotamers

- Coot for visual inspection of structure
- Check rotamers
  - Coot → Validate → Rotamer analysis

Rotamer with low probablity



Rotamer with missing atoms

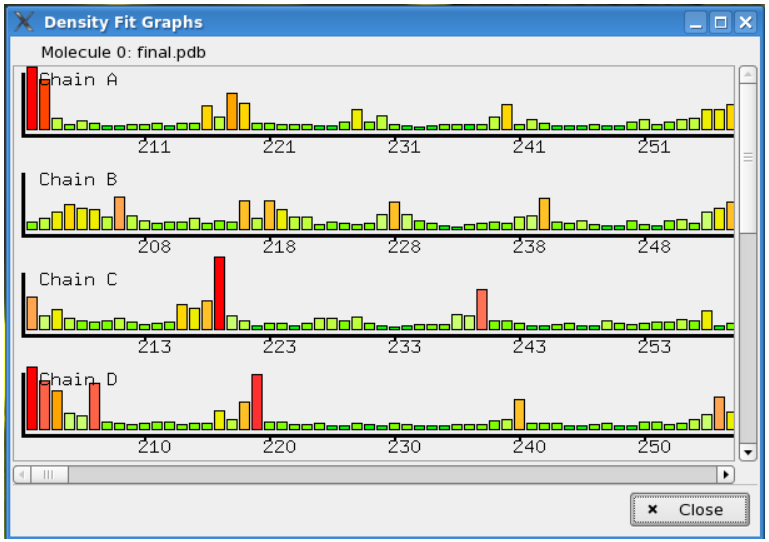
Unrecognised rotamer

What to look out for:

- Unusual rotamers
- Consistency between NCS chains

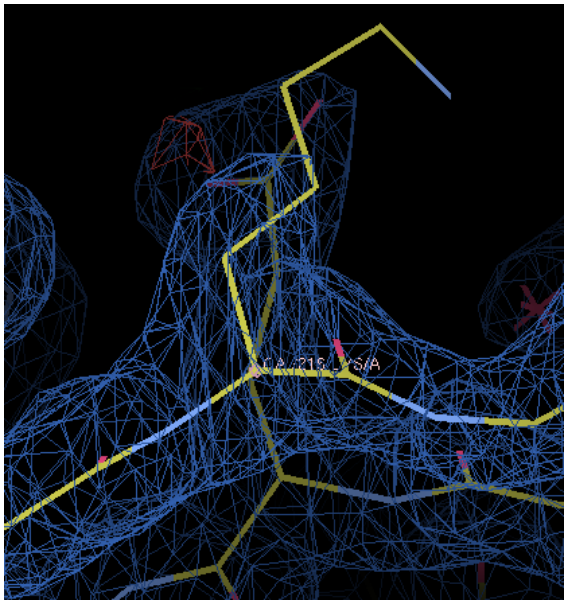
Walk structure: density fit

- Coot for visual inspection of structure
- Side chain building
  - Coot → Validate → Density fit analysis

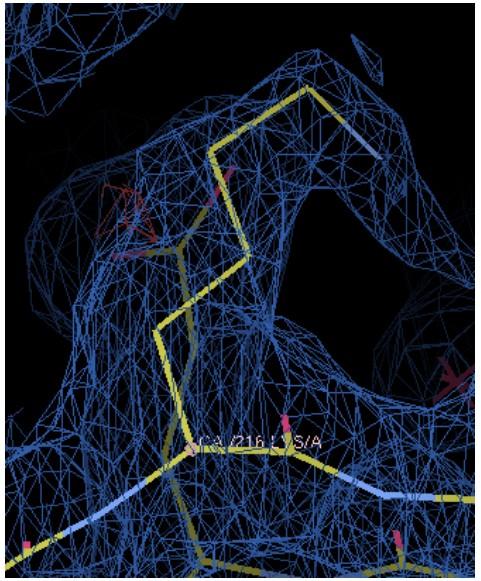


What to look out for:

- Atoms that can be placed in the map without doubt at sigma ≥ 0.5



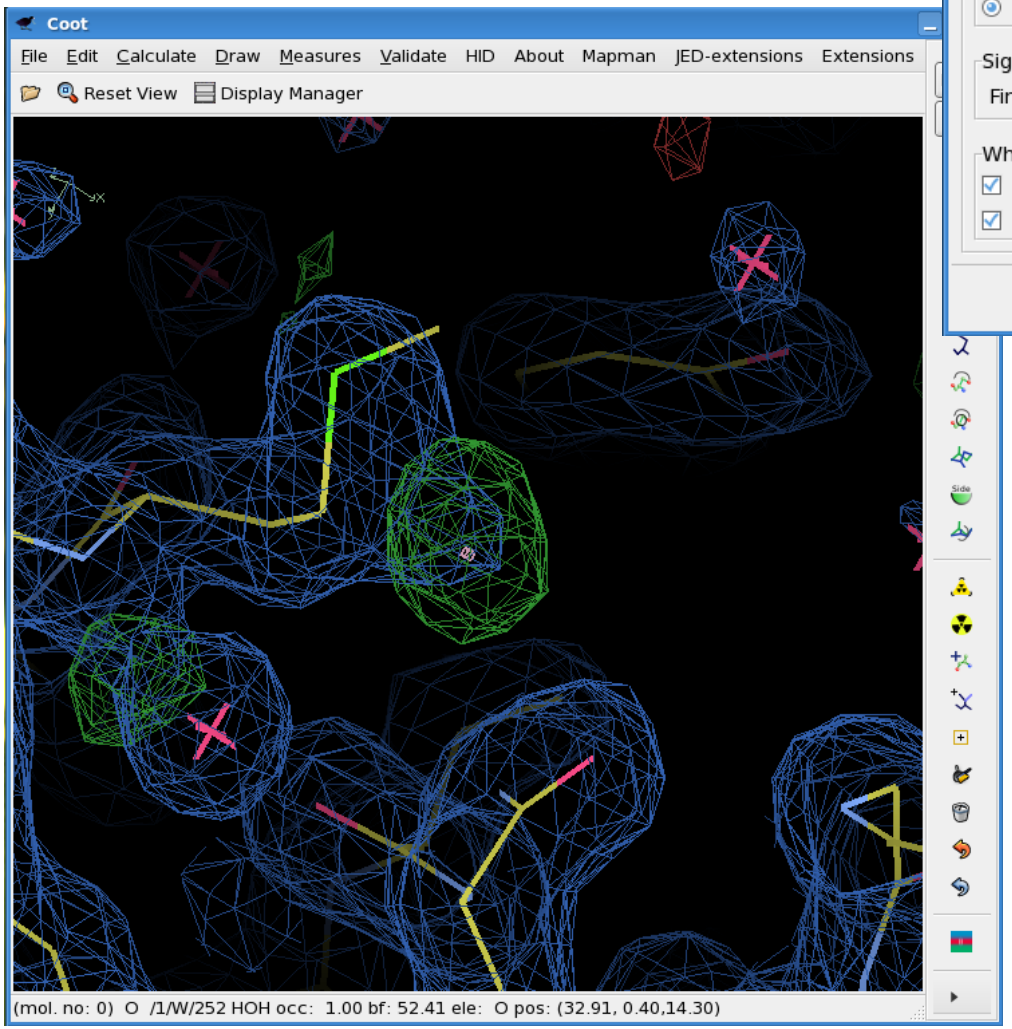
Lysine side chain @ sigma 1.0

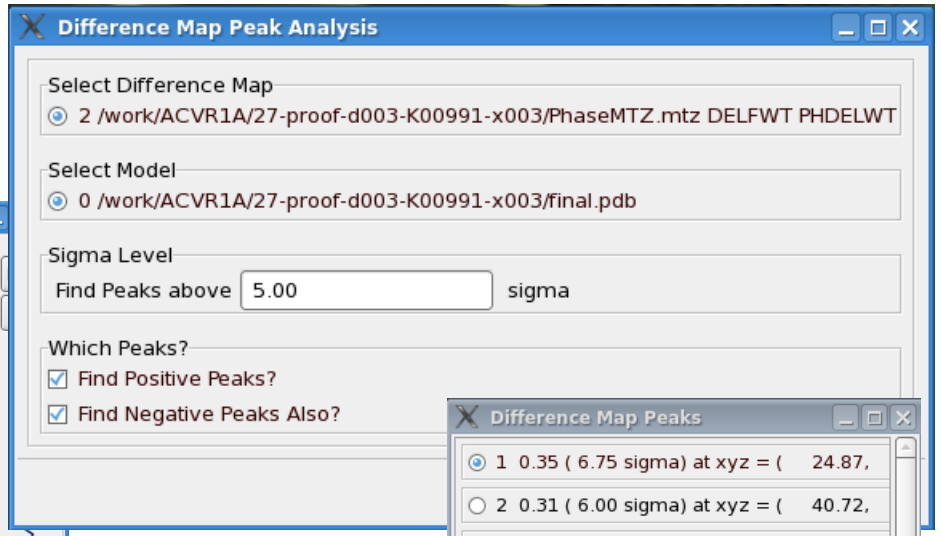


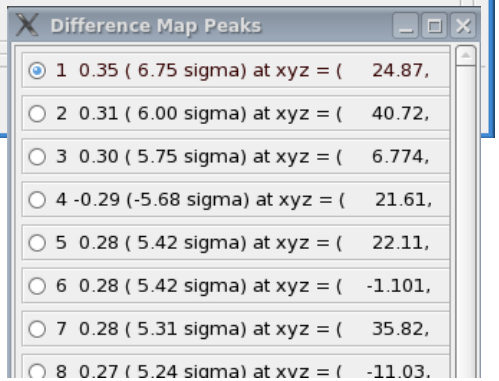
Same lysine side chain @ sigma 0.5

Walk structure: unmodeled density

- Coot for visual inspection of structure
- Check unmodeled density blobs
  - Coot → Validate → Difference Map Peaks







What to look out for:

- Missing water molecules
- Here: check rotamer fit; maybe alternative conformation

Walk structure: metal ions

- Coot for visual inspection of structure
- Check valence state and coordination of metal ions

Literature data for *d(M-X)* [if missing, see Table 2/3 in link below]

b 0.37 "universal constant", see Brese and O'Keeffe)

distances	Mg	Mn	Zn	Ca	Ni	Co	Cd	Fe
O	1.693	1.79	1.704	1.768	1.654	1.7	1.904	1.759
N	1.85	1.87	1.770	2.14	1.75	1.84	1.96	1.86
S	2.18	2.2	2.09	2.45	2.04	2.06	2.29	2.16
Valence	2	2	2	2	2	2	2	3

Observations - enter coordingating atom type, and Metal-ligand distances

Residue	H111	H113	H138	W111	W113	W138	Total valence (=2)
atom (X)	N	S	S	S	O	O	
d_{M-X} A	1.78	2.35	2.39	2.4			
d_{M-X} B	2.1	2.21	2.22	2.28	2.25	2.12	

What to look out for:

- Valence state of the ion
- Distances between the metal ion and coordinating residues/waters

Calculations - compare bold values to valence of metal (*i.e.* the charge, usually)

								Calculated valence
Mg	d0	1.85	2.18	2.18	2.18	1.693	1.693	
	vij-A	1.20827	0.631625	0.566903	0.551786	0	0	**2.96**
	vij-B	0.508813	0.922119	0.897531	0.763173	0.221927	0.315356	**3.63**
Mn	d0	1.87	2.2	2.2	2.2	1.79	1.79	
	vij-A	1.275379	0.666706	0.598389	0.582433	0	0	**3.12**
	vij-B	0.537073	0.973335	0.947381	0.805561	0.288447	0.40988	**3.96**
Zn	d0	1.77	2.09	2.09	2.09	1.704	1.704	
	vij-A	0.973335	0.495245	0.444498	0.432645	0	0	**2.35**
	vij-B	0.40988	0.723016	0.703736	0.598389	0.228624	0.324872	**2.99**
Ca	d0	2.14	2.45	2.45	2.45	1.768	1.768	
	vij-A	2.645799	1.310319	1.176051	1.144691	0	0	**6.28**
	vij-B	1.114168	1.912954	1.861945	1.583218	0.271796	0.386219	**7.13**
Ni	d0	1.75	2.04	2.04	2.04	1.654	1.654	
	vij-A	0.922119	0.432645	0.388312	0.377958	0	0	**2.12**
	vij-B	0.388312	0.631625	0.614783	0.522752	0.199726	0.283807	**2.64**
Cd	d0	1.96	2.29	2.29	2.29	1.904	1.904	
	vij-A	1.626591	0.850303	0.763173	0.742823	0	0	**3.98**
	vij-B	0.684971	1.241371	1.20827	1.027396	0.392533	0.557784	**5.11**
Co	d0	1.84	2.06	2.06	2.06	1.7	1.7	
	vij-A	1.176051	0.456675	0.40988	0.39895	0	0	**2.44**
	vij-B	0.495245	0.666706	0.648929	0.551786	0.226166	0.321379	**2.91**
Fe	d0	1.86	2.16	2.16	2.16	1.759	1.759	
	vij-A	1.241371	0.598389	0.537073	0.522752	0	0	**2.90**
	vij-B	0.522752	0.873598	0.850303	0.723016	0.265265	0.376938	**3.61**

Müller, Köpke & Sheldrick 2002, *Is the bond-valence method able to identify metal atoms in protein structures? Acta Cryst D 59, 32-3*

Brese & O'Keeffe 1991, Bond-valence parameters for solids, Acta Cryst B 47, 192-197

Summary

- Use web services and tools within CCP4 to check data quality and structure
- Objective analysis of known physical and chemical restraints
- Exceptions: biochemical knowledge and background information about target protein may explain outliers
- Check ligand geometry → it's the most interesting bit of your structure, e.g. catalytic mechanism, inhibition
- Second opinion helps to find errors and helps to get you out of a refinement loop
- Proof-reader not familiar with protein and therefore looks at it more objectively
- Speeding up to get structure ready for deposition
- Start using validation tools throughout refinement and especially towards the end

In the end:

A structure has to make biologically sense

Appendix: Checking phasing result

SGC

Using a text editor

SpaceGroup of Solution: P 1 21 1

SINGLE solution

Solution written to PDB file: /work/ACVR1A/3-phaser-d001-x003-aew/ACVR1A_9.1.pdb
Solution written to MTZ file: /work/ACVR1A/3-phaser-d001-x003-aew/ACVR1A_9.1.mtz
Solution log-likelihood gain: 5520.59
Solution annotation (history):

RFZ=16.1 TFZ=12.4 PAK=0 LLG=391
RFZ=14.9 TFZ=35.4 PAK=0 LLG=1383
RFZ=14.8 TFZ=45.3 PAK=0 LLG=2865
RFZ=14.5 TFZ=53.1 PAK=0 LLG=4684

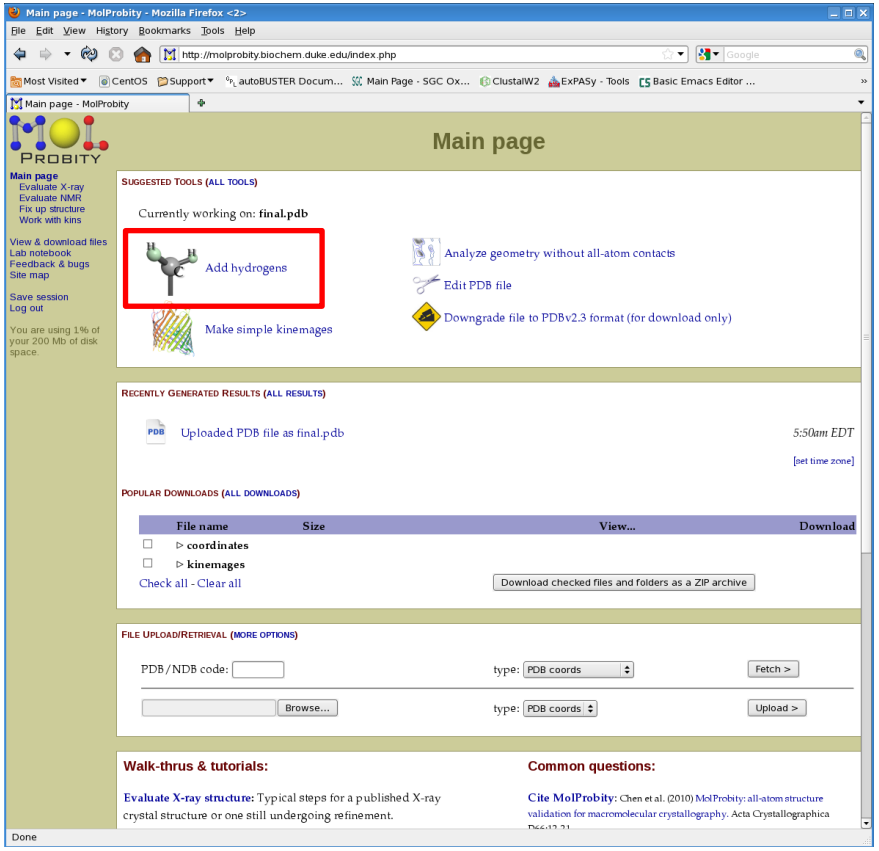
LLG=5521

What to look out for:

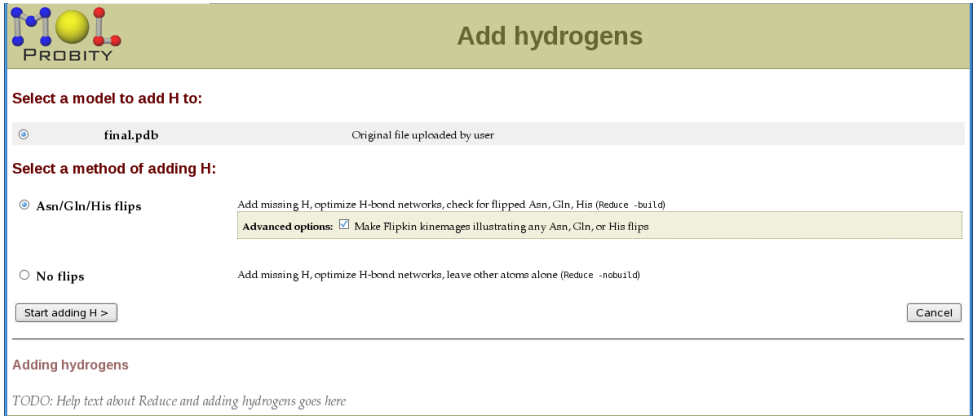
- Z-scores for rotation and translation function are expected to be > 6-7 for successful MR result
- LLG (log-likelihood gain) increases with every found molecule

Appendix: Checking geometry and more

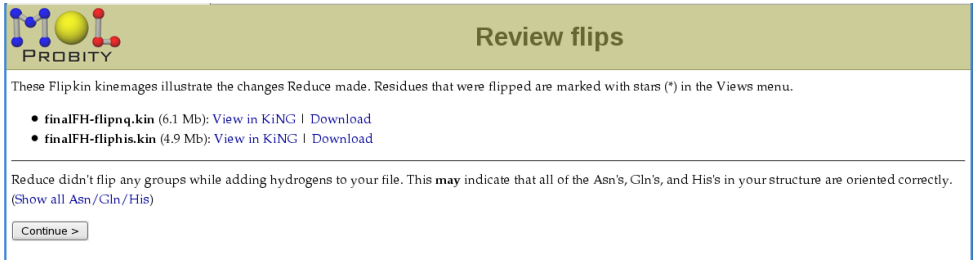
- Molprobity
  - The web server can be found here:
    http://molprobity.biochem.duke.edu/



Main page; chose file and analysis



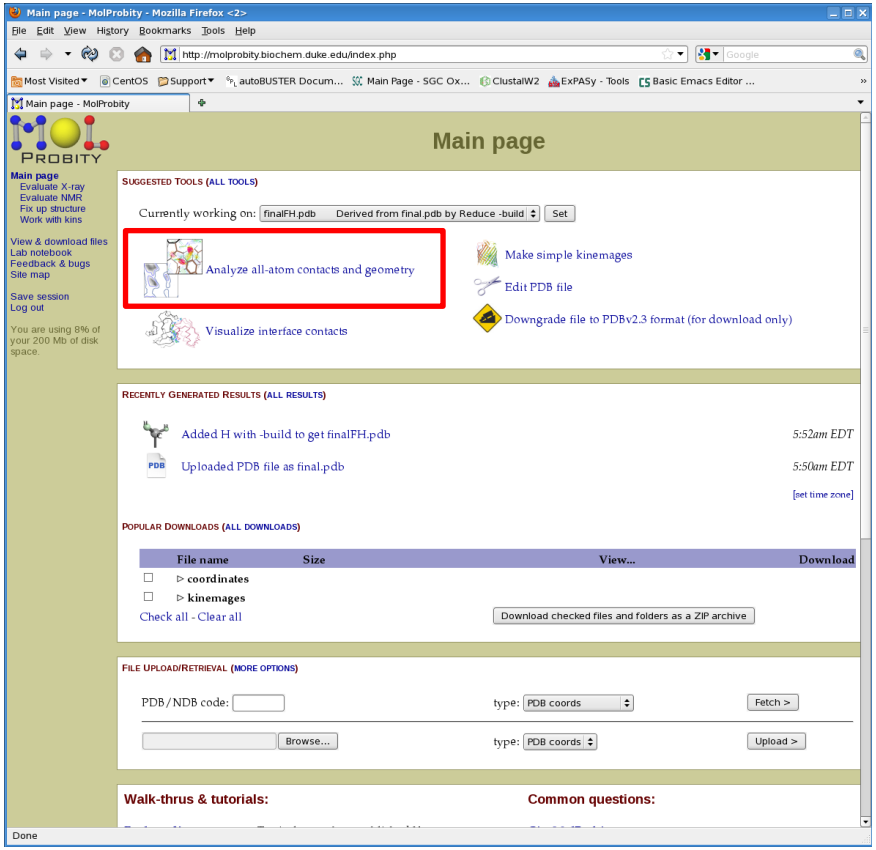
Check on atom flips by adding hydrogens



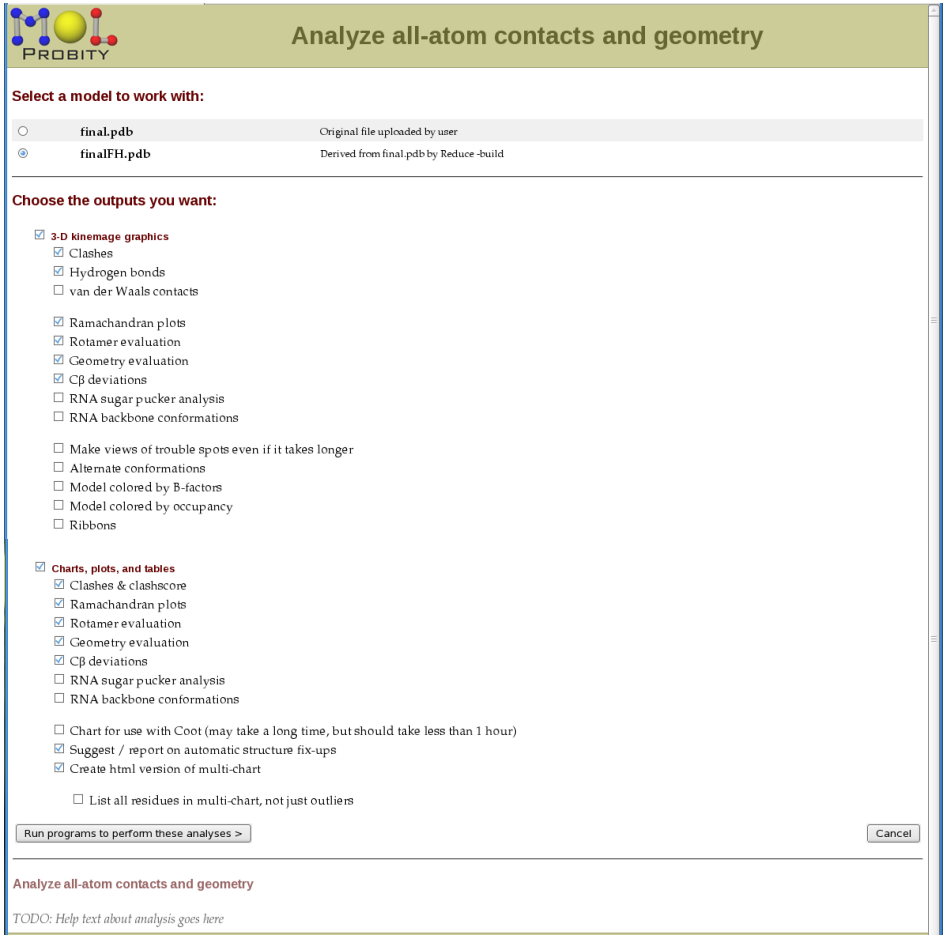
Flip suggestions and download possibility

Appendix: Checking geometry and more

- Molprobity
  - The web server can be found here: http://molprobity.biochem.duke.edu/



Main page; chose file and analysis

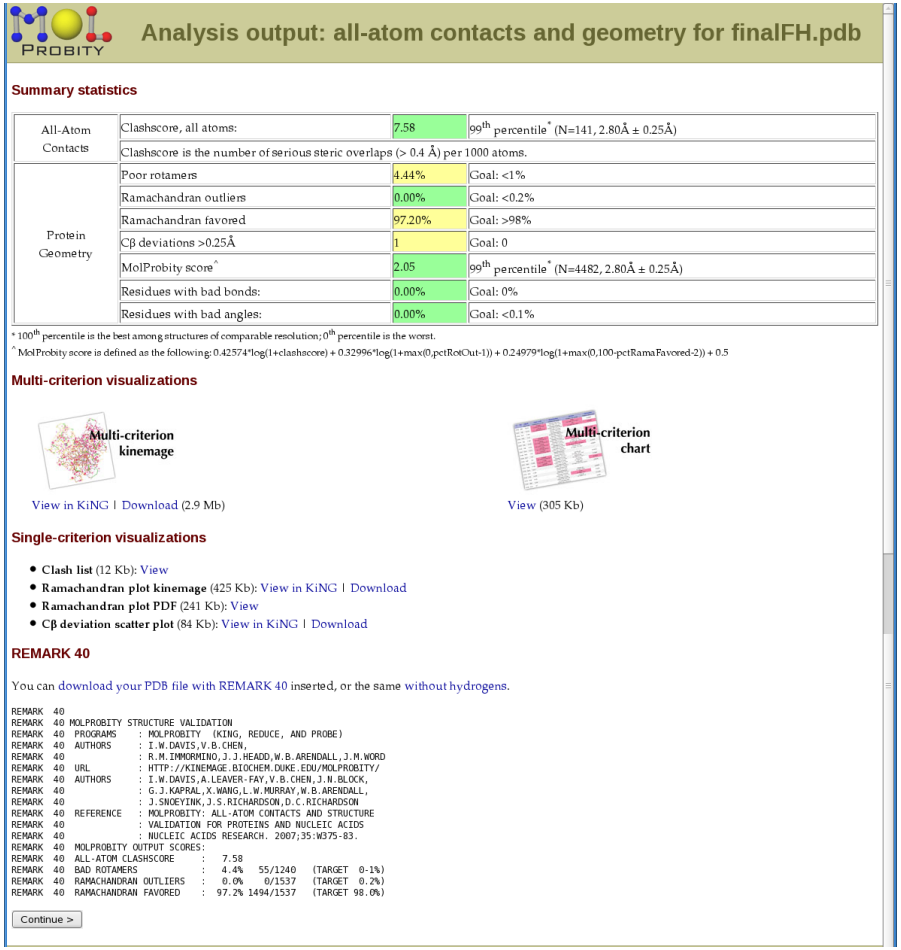


Options for geometry analysis

Appendix: Checking geometry and more

- Molprobity
  - The web server can be found here:

http://molprobity.biochem.duke.edu/



Analysis output: all-atom contacts and geometry for finalFH.pdb

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	7.58	99th percentile* (N=141, 2.80Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	4.44%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	97.20%	Goal: >98%
	Cβ deviations >0.25Å	1	Goal: 0
	MolProbity score^	2.05	99th percentile* (N=4482, 2.80Å ± 0.25Å)
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: <0.1%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.
^ MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5

Multi-criterion visualizations

Multi-criterion kinemage — View in KiNG | Download (2.9 Mb)

Multi-criterion chart — View (305 Kb)

Single-criterion visualizations

- **Clash list** (12 Kb): View
- **Ramachandran plot kinemage** (425 Kb): View in KiNG | Download
- **Ramachandran plot PDF** (241 Kb): View
- **Cβ deviation scatter plot** (84 Kb): View in KiNG | Download

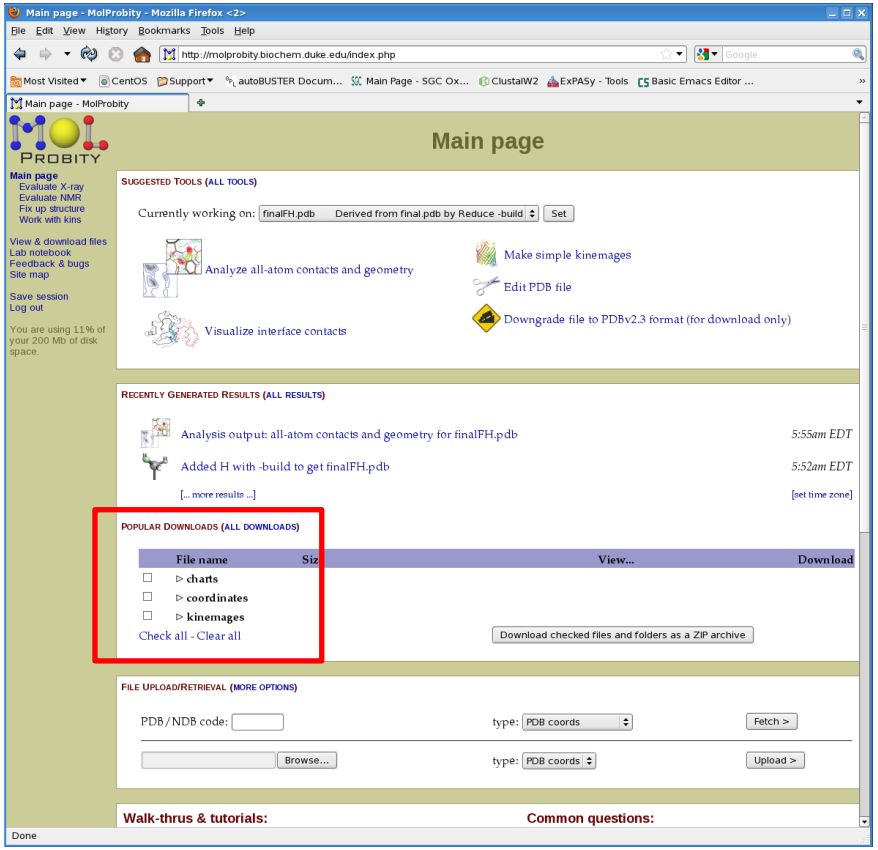
REMARK 40

You can download your PDB file with REMARK 40 inserted, or the same without hydrogens.

```
REMARK  40
REMARK  40 MOLPROBITY STRUCTURE VALIDATION
REMARK  40  PROGRAMS     : MOLPROBITY  (KING, REDUCE, AND PROBE)
REMARK  40  AUTHORS      : I.W.DAVIS,V.B.CHEN,
REMARK  40               : R.M.IMMORMINO,J.J.HEADD,W.B.ARENDALL,J.M.WORD
REMARK  40  URL          : HTTP://KINEMAGE.BIOCHEM.DUKE.EDU/MOLPROBITY/
REMARK  40  AUTHORS      : I.W.DAVIS,A.LEAVER-FAY,V.B.CHEN,J.N.BLOCK,
REMARK  40               : G.J.KAPRAL,X.WANG,L.W.MURRAY,W.B.ARENDALL,
REMARK  40               : J.SNOEYINK,J.S.RICHARDSON,D.C.RICHARDSON
REMARK  40  REFERENCE    : MOLPROBITY: ALL-ATOM CONTACTS AND STRUCTURE
REMARK  40               : VALIDATION FOR PROTEINS AND NUCLEIC ACIDS
REMARK  40               : NUCLEIC ACIDS RESEARCH. 2007;35:W375-83.
REMARK  40  MOLPROBITY OUTPUT SCORES:
REMARK  40  ALL-ATOM CLASHSCORE      :   7.58
REMARK  40  BAD ROTAMERS             :   4.4%   55/1240   (TARGET  0-1%)
REMARK  40  RAMACHANDRAN OUTLIERS    :   0.0%    0/1537   (TARGET  0.2%)
REMARK  40  RAMACHANDRAN FAVORED     :  97.2% 1494/1537   (TARGET 98.0%)
```

Continue >

Output after geometry analysis



Main page and download options after geometry analysis

Appendix: Walk structure

- Coot for visual inspection of structure
- Check waters from actit.csh and
  - Coot → Validate → Check/Delete Waters