==========
T0001 (TMV)

TARGET----------
Sequence: original uniprot P69687 - 159 residues; truncated - 158 res (see below).

Sequence of the official reference structures - 1ei7_A (XRay) and 4udv (EM) starts from 'SYS' and is shifted by 1 residue with respect to the uniprot sequence P69687 posted on the web (starts from 'MSYS').

The same concerns another potential reference structure - 4gqh (XRay).

1ei7 looks like a heteromultimer when judged from the pdb file format (two chains with different residue numbers), but is de-facto a homodimer: even though sequence of 1ei7_B starts from 201, it is identical to that of chain A.

All but one models follow residue numbering of 1ei7_A /4udv as they report residue 1 as S (while in the uniprot sequence res. 1 is M).

To remedy the situation, first residue M in the original uniprot sequence was deleted and the sequence file was renamed to seqnames/T0001..._adj.fasta. This way it was easier to bring models, target and the released sequence in accord (rather than renumbering all models and the target).

Structure
Reference 1ei7 (XRay) contains 2 homo chains in ASU and the whole disc (or multiple disc strcutrue) should be built from symmetry. Chains A and B have different numbering of residues even though correspond to the same sequence. Downloaded bioassembly from the PDB, which contains 34 NMR-like models forming 4 disks (each model contains chains A and B that are identical in sequence with rmsd=0 to each other).
Action item: SHOULD WE REFORMAT THIS 1ei7 BIOASSEMBLY SO THAT IT CONTAINS DIFFERENT CHAINS (INSTEAD OF MODELS) AND USE IT AS A MAIN REFERENCE FOR THIS TARGET?

Reference 4gqh (Xray) has 4 bioassemblies (1/4 arc) in the PDB; when all these 4 PDB models are concatenated, they form the whole disk (chains A-Z,a-i). Residues 93-106 and 155-158 are missing.

Reference 4udv (EM) contains 3 helical (out-of-plane) disks. Format of this pdb file is NMR-like (Models 1,2,...49); all models contain only chain A. Residues 154-158 are missing.
Action item: SHOULD WE REFORMAT THIS 4udv PDB ENTRY THAT IT CONTAINS DIFFERENT CHAINS (INSTEAD OF MODELS) AND USE IT AS A SECONDARY (EM) REFERENCE FOR THIS TARGET?

EM Map EMD-2842 seems to contain data on 9 'spiral' disks.

MODELS: 11----------
One model was deleted as group id could not be identified and Cathy said that a submitter withdrew the model (see email).

OK or minor issues:
T0001EM119_1 (+c) 1 chain (A); rmsd(M_A:T_A)=0.3
T0001EM123_1 (+) 37 structurally identical chains (A-Z,0-9,a) missing 154-158 as in 4udv; 3 disks; rmsd=0.6 to chains in 4udv
T0001EM123_2 (+c) 1 chain (A); rmsd(M_A:T_A)=1.1 ab initio
T0001EM133_1 (+c) 11 structurally identical chains (A-I,U,a); contain ~1/4 segments of 3 disks; rmsd=0.8 to chains in 4udv
T0001EM181_1 (+c) 1 chain (A) missing 154-158; rmsd(M_A:T_A)=2.9; ab initio; relatively high in clashes
T0001EM194_1 (+c) 1 chain (a) missing 154-158; rmsd(M_a:T_A)=0.9; ab initio

Problematic:
T0001EM130_1 (-fa) 2-symbol chain IDs - PHENIX can handle it, but other software not; ab initio; looks OK in pymol, but chains are fragmented in blocks like (SL - res.76-95; TL - res. 171-256 and sequence numbering does not correspond to the one in the target)
T0001EM130_2 (-fa) 2-symbol chain IDs - PHENIX can handle it, but other software not; ab initio; looks OK in pymol, but chains are fragmented in blocks like (BJ - res.39-50+79-88+114-135; CJ - res. 1000-1037 and sequence numbering does not correspond to the one in the target)
T0001EM164_1 (-fa) contains 136 models spread in space and forming 9 disks; numbering of residues starts from 1 goes to 3 (SYS), then restarts from 201 and contains non-amino acid entries for res. 201, 203, 205; afterwards contains HETATMS in place of regular sequence ATOM; at least needs cleanup and renumbering; *PHENIX fails - multiple models problem.
T0001EM181_2 (-ac) 1 chain (A); no seq. assigned - polyalanine; CA only; ab initio; PHENIX OK
T0001EM192_1 (-f) 2-symbol chain IDs (3 spiral disks); PHENIX OK.

==========
T0002 (Proteasome)

TARGET----------
Heterodimer. Two subunits alpha and beta have different sequences.
*
Two maps.
*
1yar (Xray) has 7 proteasome chains in 2 rings: A-G (alpha unit, in between outer ring and the

PA26 protein - see further) and H-N (outer ring, beta unit). 1yar also contains PROTEASOME ACTIVATOR PROTEIN PA26. Unit A misses residues 1-12; unit B misses residues 1-8.
*
3j9i (EM, mapA=EMD-5623 only) has 28 chains in 4 different rings:
COMPND 2 MOLECULE: PROTEASOME SUBUNIT ALPHA; 2 OUTER RINGS
COMPND 3 CHAIN: S, F, T, G, U, A, O, B, P, C, Q, D, R, E; (224 res)
COMPND 8 MOLECULE: PROTEASOME SUBUNIT BETA; 2 INNER RINGS
COMPND 9 CHAIN: Z, M, 1, N, 2, H, V, I, W, J, X, K, Y, L; (203 res)

Released sequence in beta-unit is shifted with respect to structures in the PDB: res.1-4 stretch in the structures (TTTV) is actually res.9-13 in the P28061 sequence. Models follow the released sequence - so we changed residue numbering in the reference experimental structures (so that TTTV is now res.9-13).

MODELS: 18----------
OK or minor issues:
T0002EM120_1 MapA (+) 28 chains; alpha:(A-G,O-U), 224 res ; beta:(H-N,V-Z,1,2), 203 res.
T0002EM131_1 MapA (+) 28 chains; alpha:(B-H,O-U), 224 res ; beta:(A,I-N,V-Z,1,2), 203 res.
T0002EM133_1 MapB (+) 28 chains; alpha:(B,I-N,P,W-Z,a-b), 229 res ; beta:(A,C-H,O,Q-V), 200 res
T0002EM133_2 MapA (+) 28 chains; alpha:(B,I-N,P,W-Z,a-b), 229 res ; beta:(A,C-H,O,Q-V), 200 res
T0002EM164_1 MapB (+) 28 chains; alpha:(A-G,H-N), 224 res; beta:(a-g,h-n), 203 res.
T0002EM181_1 MapA (+c) 1 chain; 203 res. (beta only); 6 Mutated residues: 206:LYS->ARG, 207:LEU->LYS, 208:GLY->LEU, 209:LEU->GLY, 210:ILE->LEU, 211:LEU->ILE; rmsd=4.9 to target chains H-N; ab initio
T0002EM183_1 MapA (+-fc) 1 chain; 10 models sitting on top of each other; 224 res. (alpha only); rmsd=5.75 to target chains A-G; ab initio; *PHENIX fails: Only one model allowed. Considering only the first model for the rest of the software.
T0002EM189_1 MapA (+) 28 chains; alpha:(A,C-O), 221 res; beta:(B,P-Z,1,2), 203 res.
T0002EM189_2 MapB (+) 28 chains; alpha:(A,C-O), 221 res; beta:(B,P-Z,1,2), 203 res.
T0002EM192_1 MapA (+) 28 chains; alpha:(A-G,O-U), 224 res ; beta:(H-N,V-Z,1,2), 203 res.

Problematic:
T0002EM123_1 MapA (+-f) 28 chains; alpha:(A-G,H-N), 224 res; beta:(a-g,h-n), 203 res.
*PHENIX, QMEAN, QSscore fail: duplicated atoms.
T0002EM123_2 MapB (+-f) 28 chains; alpha:(A-G,H-N), 224 res; beta:(a-g,h-n), 203 res.
*PHENIX, QMEAN, QSscore fail: duplicated atoms
T0002EM123_3 MapB (+-fc) 1 chain; 203 res. (beta unit only); rmsd=0.7 to target chains (H-N)
*PHENIX, QMEAN, QSscore fail: duplicated atoms
T0002EM130_1 MapB (-fa) 2-symbol chain names; 3*14 chains; third group of chains has res. numbers in 1000s (?): U, F, J, N, R, W, AA, EA, IA, MA, QA, UA, YA, CB;
T0002EM130_2 MapA (-fa) 2-symbol chain names; 3*14 chains; third group of chains has res.

numbers in 1000s (?): U, F, J, N, R, W, AA, EA, IA, MA, QA, UA, YA, CB;
T0002EM130_3 MapB (-fa) 2-symbol chain names; 3*14 chains; third group of chains has res. numbers in 1000s (?): U, F, J, N, R, W, AA, EA, IA, MA, QA, UA, YA, CB;
T0002EM130_4 MapA (-fa) 28 chains; 2-symbol chain names; wrong sequence; composed of 15res-chain + 232-res chain blocks; chains B,C,F,I,...: 15 res (159-173) correspond. to unit B; chains V,D,G: 232 res not fitting to target (starts from res. 241, while max. res. in either of chains - 233, sequence starts from YVYGK, which is not present in A or B unit)
T0002EM181_2 MapA (-ac) 1 chain; no seq. assigned; polyalanine; CA only; PHENIX OK; ab initio
*
For evaluation with QSs models are split in alpha and beta subunints separately

==========

T0003 (GroEL)

TARGET----------

Reference structure 1ss8 (Xray, 2.7A, P.Adams et al, 2004) has two one-residue conflicts compared to the uniprot sequence POA6F5 and other reference structures 1xck (Xray), 1svt(Xray) and 3cau(EM): 13:ARG(uniprot)->GLY(1ss8), 126:ALA(uniprot)->VAL(1ss8). 1ss8 contains 524/548 uniprot res (missing 1, 526-548).

Reference structure 1svt has two different sets of chain conformations (A-G and H-N). Evaluated separately.

MODELS: 8----------

OK or minor issues (6):
T0003EM119_1 (+) 14 structurally identical chains, 522 res; rmsd to 1ss8 chains in 1.6-1.7 range.
T0003EM123_1 (+) 14 structurally identical chains, 524 res; rmsd to 1ss8 chains in 1.0-1.1 range.
T0003EM133_1 (+) 14 structurally identical chains, 524 res; rmsd to 1ss8 chains in 1.7-1.8 range.
T0003EM164_1 (+) 14 structurally identical chains, 522 res; rmsd to 1ss8 chains in 1.7-1.8 range.
T0003EM164_2 (+) 14 practically structurally identical chains, 522 res; chains N is the most divergent and has 0.2 rmsd to other chains in the model; rmsd to 1ss8 chains in 1.1-1.32 range.
T0003EM192_1 (+) 14 structurally identical chains, 524 res; rmsd to 1ss8 chains in 0.8-1.0 range; .had the same residue names as in original 1ss8 (no mutations). This model was evaluated vs the original 1ss8 file (1ss8.pdb.orig), while all others were evaluated vs the adjusted target file (1ss8.pdb.seqfixed) with 13:ARG->GLY, 126:ALA->VAL mutations (back to the uniprot sequence).

Problematic (2):

T0003EM130_1 (-fa)
T0003EM130_2 (-fa)
Both models from group 130 contain data on 4 rings and contain 2-symbol chain IDs. In addition to this they are very fragmented (res. from different chains belong to different rings) and have wrong sequence - visualize in Pymol. Currently placed in the MODELS/T0003/problematic directory.

==========
T0004 (TRPV1 Channel)

TARGET----------
XRay - not identified
EM - 3j5p - 4 chains, each cover 581/838 residues in the associated uniprot sequence O35433. Residues (111-502,508-603,627-719) covered. Residues 752-762 - UNK.

MODELS: 13----------

OK or minor issues:
T0004EM119_1 (+) 4 chains, 619 res. (111-603,627-752)
T0004EM120_1 (+) 4 chains, 622 res. (111-603,627-755)
T0004EM123_1 (+) 4 chains, 489 res. (203-502, 508-603, 627-719)
T0004EM131_1 (+) 4 chains, 581 res. (same as in the target)
T0004EM133_1 (+) 4 chains, 512 res. (198-603, 627-719) #QMEAN problems
T0004EM164_1 (+) 4 chains, 581 res. (same as in the target)
T0004EM164_2 (+-) 4 chains, 315 res., res. (381-603) are OK; res. (604-695) present, but have wrong sequence; #QMEAN problems; went OK through PHENIX (bad scores)
T0004EM192_1 (+) 4 chains, 581 res. (same as in the target)
T0004EM193_1 (+-) 4 chains, 498 res. (199-603, 627-719), 7 mutated res. HIS->HSE; * PHENIX fails - no map-model FSC section (is it because of the HIS->HSE changes? or due to wrong box/symmetry parameters similarly to T0006 model from the same group:
T0006EM193_1 - will retry after Pavel's PHENIX fix.)
T0004EM194_1 (-?1) CA only, 1 chain 'a' with rmsd=11(!) to target chains, 530 res (199-225,227-271,276-733). #QMEAN failed (no dssp on ca only); PHENIX finished; ab initio

Serious issues:
T0004EM130_1 (-c) fragmented, 8 chains. Moved to 'problematic' directory; ab initio
T0004EM130_2 (-c) very fragmented, 8 chains. Moved to 'problematic' directory; ab initio.
T0004EM183_1 (-afc) wrong numbering, 1 chain, 315 res.; 10 models sitting on the top of each other; * PHENIX failed; ab initio

==========
T0005 (BMV)

TARGET----------
XRay reference structure: 1js9 (3.4A, year 2002). Long tail in chain C and a little bit shorter in chain B. Chain length: A:149/189; B:165/189; C:189/189. rmsd(T_A:T_B)=0.95.

EM reference: 3j7l

MODELS: 12----------

OK or minor issues:
T0005EM119_1 (+) OK, rmsd=1.1 between AB and BC chains of the model (AC similar)
T0005EM133_1 (+) OK
T0005EM194_1 (-+) 3 chains; only chain 'M_b' similar to T_A,B,C (rmsd ~1.6). Chains 'a' and 'c' look mostly like spaghetti (practically no ss elements in pymol). Chain 'b' is just a little bit better. Ab initio.

Different problems:
T0005EM123_1 (-) monomer; wrong residue numbering; ab initio
T0005EM123_2 (-f) 60 chains; 477 res; model check program reports no structurally matching chains. How to evaluate?
T0005EM130_1 (-f) 2-symbol chain IDs. Moved to 'problematic' directory; ab initio.
T0005EM130_2 (-f) 2-symbol chain IDs. Moved to 'problematic' directory; ab initio.
T0005EM164_1 (-f) The file is formatted as 60 different models (NMR like) with applied symmetry operators, so that all the models together form a biounit 'ball-like' structure. RMSD between different chains of the model 1 is in the range 1.0-1.2. *PHENIX fails
T0005EM181_1 (-) wrong residue numbering (seems like 40 res. shift); ab initio
T0005EM181_2 (-!) no sequence assigned (polyalanine chain); only CAs; ab initio
T0005EM183_1 (-) multiple models on top of each other; 149 res; wrong residue numbering. Moved to 'problematic' directory. *PHENIX failed: multiple models; ab initio
T0005EM192_1 (-f) 2-symbol chain IDs. Moved to 'problematic' directory.

==========
T0006 (BGal)

TARGET----------
(4 chains, 1011 res.)
Sequence of the reference Xray structure 1jz7 is shifted by 1 residue with respect to the uniprot sequence of the target P00722 posted on the web.

All models seem to use residue numbering of 1jz7 as they reported residue 13 as R, while in the uniprot sequence res. 13 was Q (one residue shift of the whole sequence).

To remedy the situation, first residue M in the original uniprot sequence

T0006_BetaGalactosidase_P00722.fasta was deleted (T0006_BetaGalactosidase_P00722_adj.fasta). This way it was easier to bring models, target and the released sequence in accord (rather than renumbering all models and the target).

Res. 247 is changed back from modified CSO to CYS (extra oxygen deleted).

MODELS: 16 (10 on MapA; 6 on MapB)----------

12 OK (one monomer - T0006EM194_1); 4 problematic

OK or minor issues:
T0006EM119_1 MapA (+) 4 chains, 1022 res.
T0006EM119_2 MapB (+) 4 chains, 1022 res.
T0006EM123_1 MapB (+) 4 chains, 1022 res.
T0006EM123_2 MapA (+) 4 chains, 1022 res.
T0006EM128_1 MapA (+) 4 chains, 1011 res.
T0006EM133_1 MapB (+) 4 chains, 1011 res.
T0006EM133_2 MapA (+) 4 chains, 1011 res.
T0006EM164_1 MapA (-f) 4 chains, 1022 res.; *PHENIX fails: Conflicting scattering type symbols:
  initial symbol: "G" (from pdb element column)
      new symbol: "MG" (with residue name MG)
  atom:
    "HETATM32825 MG   MG A3001 .*.   G "
AK fix: deleted all HETATMs towards the end of prediction and then PHENIX ran OK.
T0006EM192_1 MapA (+) 4 chains, 1022 res.
T0006EM192_2 MapB (+) 4 chains, 1022 res.
T0006EM193_1 MapA (-f) 4 chains, 1022 res.; *PHENIX fails: TypeError: float argument required, not NoneType (wrong symmetry/box parameters - Pavel is looking inot it.)
T0006EM194_1 MapA (+c) 1 chain (a), 1018 res.

Problematic
T0006EM130_1
T0006EM130_2
T0006EM130_3
T0006EM130_4
For all 4 models from group EM130 cannot assign target seq. with >90% correspondence. Currently sitting in the MODELS/T0006/problematic directory. Ab initio.
*
Lists of models that were built on MapA or MapB are in models/MODELS/T0006/MapX.models files (files needed for PHENIX run). All models from MapB (except for two from group 130) were evaluated normally with PHENIX. Two models from mapA (and additionally two models from group 130) failed with PHENIX.

==========

T0007 (GSec)

TARGET----------

No crystal structure. 2 different maps.

4 different sequences in the complex.

Target EM reference structure 4upc corresponds to the 4th sequence (nicastrin, 709 residues) and contains only 409 res. out of 709 in uniprot seq.

Another EM reference structure 5a63 (year 2015, not listed in the target spreadsheet) is a heterotetramer with chain A correponding to the target. It is longer that 4upc - 665 out of 709 res available.
Chain length: A:665/709 res.; B:215; C:243; D:100.
!!! WARNING: rmsd on 409 corresponding atoms in two EM structures is 11.1 (!). Which one we can trust? Only 270 atoms out of 409 in 4upc fits under 4A (LGA SUMMARY).

CLARIFIED at the teleconf: 4upc is a wrong structure. Use only 5a63 for the assessment (not on the official reference structure list).

MODELS----------

T0007EM118_1 MapA (-c) 1 chain; 697 res.; rmsd(M_A:T_A)=18.4
T0007EM119_1 MapA (+) 4 chains: chain A 665 res, chain B 215 res, chain C 243 res, chain D 100 res. All chains have low rmsd (around 1.0) to corresponding chains in 5a63.
T0007EM119_2 MapB (+) 4 chains: chain A 665 res, chain B 215 res, chain C 243 res, chain D 100 res. All chains have low rmsd (<0.5) to corresponding chains in 5a63 (but high rmsd=10.93 to chain A of 4upc).
T0007EM120_1 MapB (+) 4 chains: chain A 665 res, chain B 215 res, chain C 243 res, chain D 100 res. All chains have low rmsd (<0.5) to corresponding chains in 5a63.
T0007EM123_1 MapA (+) 4 chains: chain A 665 res, chain B 215 res, chain C 243 res, chain D 100 res. All chains have average rmsd (around 2.0) to corresponding chains in 5a63.
T0007EM123_2 MapB (+) 4 chains: chain A 665 res, chain B 215 res, chain C 243 res, chain D 100 res. All chains have low rmsd (around 1.0) to corresponding chains in 5a63.
T0007EM133_1 MapB (+) 4 chains: chain A 243 res (corr. T_C), chain B 665 res(corr. T_A), chain C 100 res (corr. T_D), chain D 210 res (corr. T_B). All chains have relatively low rmsd (1.0-1.4) to corresponding chains in 5a63.
T0007EM133_2 MapA (+) 4 chains: chain A 243 res (corr. T_C), chain B 665 res(corr. T_A), chain C 100 res (corr. T_D), chain D 210 res (corr. T_B). All chains have relatively low rmsd (1.0-1.4) to corresponding chains in 5a63.

T0007EM164_1 MapA (-c) 1 chain; 409 res., rmsd(M_A:T_A)=10.9. ***[DELETED as built on wrong template]***
T0007EM164_2 MapB (+) 4 chains: chain A 665 res, chain B 215 res, chain C 243 res, chain D 100 res. All chains have low rmsd (<0.3) to corresponding chains in 5a63.
T0007EM181_1 MapB (-c) 1 chain; 665 res.; rmsd=3.52.
T0007EM181_2 MapB (-ac) 1 chain 'a', 665 res; no seq. correspondence; polyalanine, CA only.
T0007EM183_1 MapB (-f) 1 chain, 10 models on top of each other. 665 res; poor rmsd=14.49 to 4upc_A and also to 5a63_A (rmsd=11.43). *PHENIX fails: multiple models.
T0007EM185_1 MapB (+) 4 chains, A 590 res, B 168 res, C 147 res, D 42 res. In chain A 9 mutated res.: 701:PRO->LEU, 702:ARG->ILE, 703:GLU->THR, 704:PRO->LEU, 705:GLY->THR, 706:ALA->VAL, 707:VAL->GLY, 708:SER->PHE, 709:TYR->GLY.; rmsd(M_A-T_A)=6.06(5a63) and rmsd=10.37(4upc).
T0007EM189_1 MapA (-c) 1 chain; 618 res.; 1 Mutated residue :261:LYS->ASN; rmsd=3.7.
T0007EM192_1 MapA (-c) 1 chain, 409 res; 8 mutated res. :193:LYS->ASN, 194:GLN->LEU, 195:CYS->SER, 196:TYR->GLN, 197:GLN->ASN, 198:ASP->GLY, 199:HIS->SER, 261:LYS->ASN; poor rmsd=11.1 for M_A:T_A in 5a63, but quite low rmsd=1.86 to 4upc (how possible? - probably 409 res. in 4upc are messed up the same way as in this model?).***[DELETED as built on wrong template]***
T0007EM192_2 MapB (+) 4 chains, A 665 res, B 215 res, C 243 res, D 100 res; All chains have low rmsd (<0.74) to corresponding chains in 5a63 (but high rmsd=11.03 to chain A of 4upc)
T0007EM194_1 MapB (-ac) 2 chains: c 243 res, d 100 res. Sequences correspond to chains C and D in 5a63 (no piece corresponding to the subunit of the main chain A).

Problematic
T0007EM130_1 MapA (-) 4 chains: A 18 res, B 22 res, D 37 res, X 727 res. No seq. fit to target. Many GGG... or AAA... stretches.
T0007EM130_2 MapB
T0007EM130_3 MapA
T0007EM130_4 MapB

==========
T0008 (Ribosome)

--- TARGET

X-ray reference 4u26 is available only in mmcif format.

EM reference 5afi is a ribosome-EF-TU complex. Protein molecules start at orig. atom 33031 and are represented as chains b-u plus z (starting with atom 56946).

--- MODELS