

SET8 F@h data analysis

Rafal Wiewiora, Chodera Lab

The simulation setup

1ZKK

4IJ8



Strip of everything (peptide, co-factors etc.), leave
only chain A of protein



Consensus sequence

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
RKSKAELQSE	ERKRIDELIE	SGKEEGMKID	LIDGKGRGVI	ATKQFSRGDF	VVEYHGDLIE
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
ITDAKKREAL	YAQDPSTGCY	MYYFQYLSKT	YCVDATRETN	RLGRLINHSK	CGNCQTKLHD
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>		
IDGVPHLILI	ASRDIAAGEE	LLYDYGDRSK	ASIEAHPWLK	H	

(positions in Uniprot: Q9NQR1-1: 233-393)

solvate

equilibrate (15 ns
simulation -
discard)



production run on
F@h -
500 CLONEs started
from each
of two starting
configurations

Simulation output being analyzed here

aggregate **0.98** ms dataset, in **969** trajectories
(not all trajectories had been active by
that point)

485 started from the RUN0 configuration
484 started from the RUN1 configuration

Whether the two sets have connected in conformational space, and hence whether we can analyze them together, is still being analyzed (and v. important for building the Markov State Model)

RMSD analysis

- 1) C-alpha RMSD of all frames in a trajectory vs. frame 0
- 2) C-alpha RMSD vs. reference structures (1ZKK, 4IJ8, Apo, Inhibitor)
- 3) RMSF (average RMSD over time) / residue - still thinking / calculating

RMSD analysis - sequence analyzed

Wanted to find a common sequence, where we have all the C-alpha's present, for all 5 structures - simulation data + 4 reference

Took chains 0 (using 0-indexing here) of all reference structures. 4IJ8 has one mutation (C343S), Apo has 3 mutations (K297A, K298A, E300A) and 1 deletion (T309). We will not worry about mutations - i.e. consider those C-alphas regardless of exact amino acid. For the deletion, we will not consider that C-alpha in any structures.

Analyzed sequence:

KSKAELQSEERKRIDELIESGKEEGMKIDLIDGKGRGVIATKQFSRGDFVVEYHGDLIEITDAKKREALYAQDPSGCYMY
YFQYLSKTYCVDATRETNRLGRLINHSKCGNCQTKLHDIDGVPHLILIASRDIAAGEELLYDYGDRSKASIEAHPWL

Uniprot Q9NQR1-1: 234 to 391 (without T309)
In our analysis, we 0-index this, to: 0 to 157 (without T75)

RMSD analysis - separation into domains



Not sure where this numbering comes from (Shi please correct?), but comparing the SET domain range here with the numbering on Uniprot, I assumed the numbering here = Uniprot numbering - 40.

So in our analyzed sequence, with 0-indexed residues, I have these domain ranges:

N-flank: 0 - 22 (so first 2 residues from the picture above missing)

SET1: 23 - 56

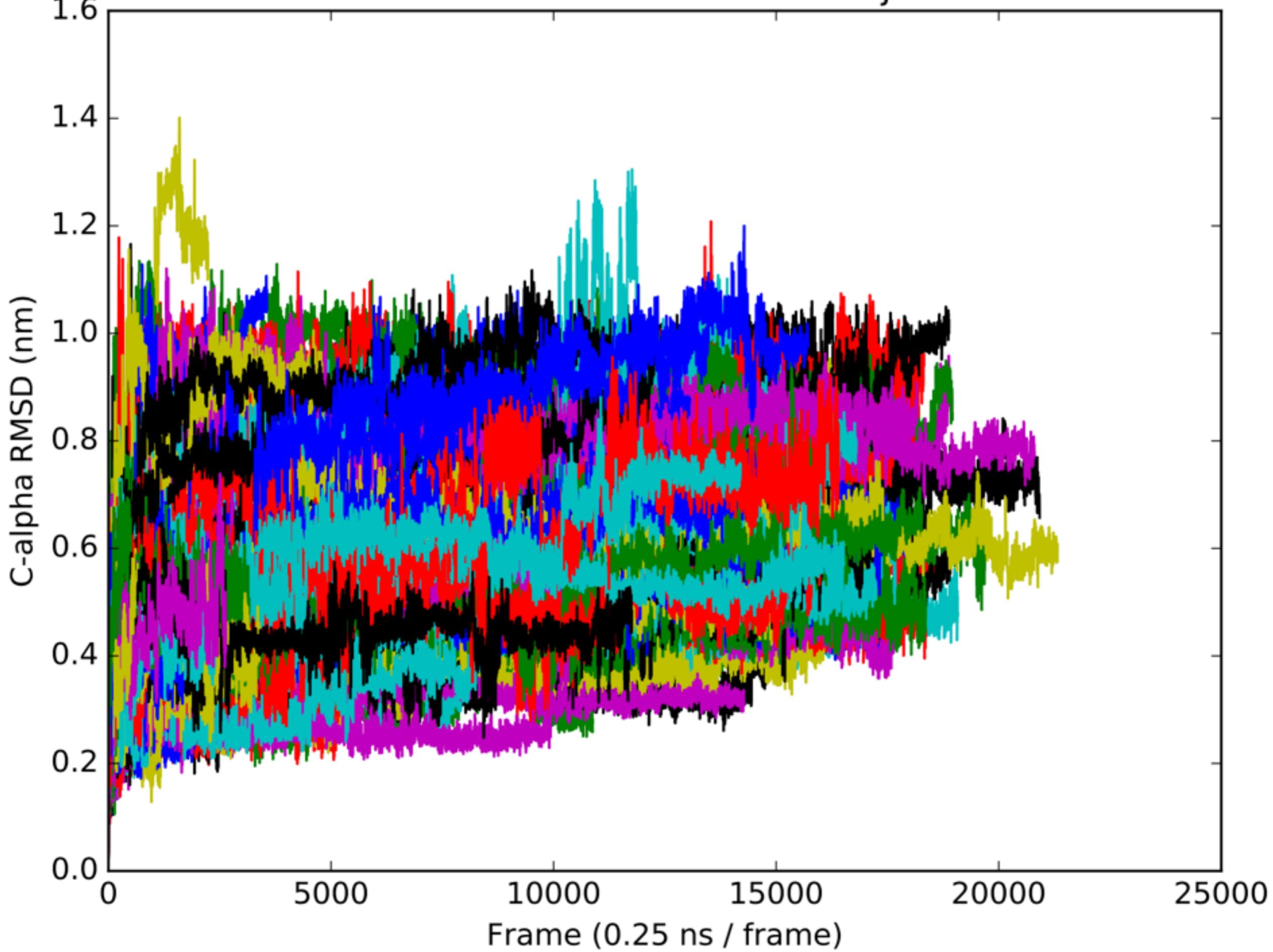
SETI: 57 - 92

SET2: 93 - 142

C-flank: 143 - 157 (1 last residue missing)

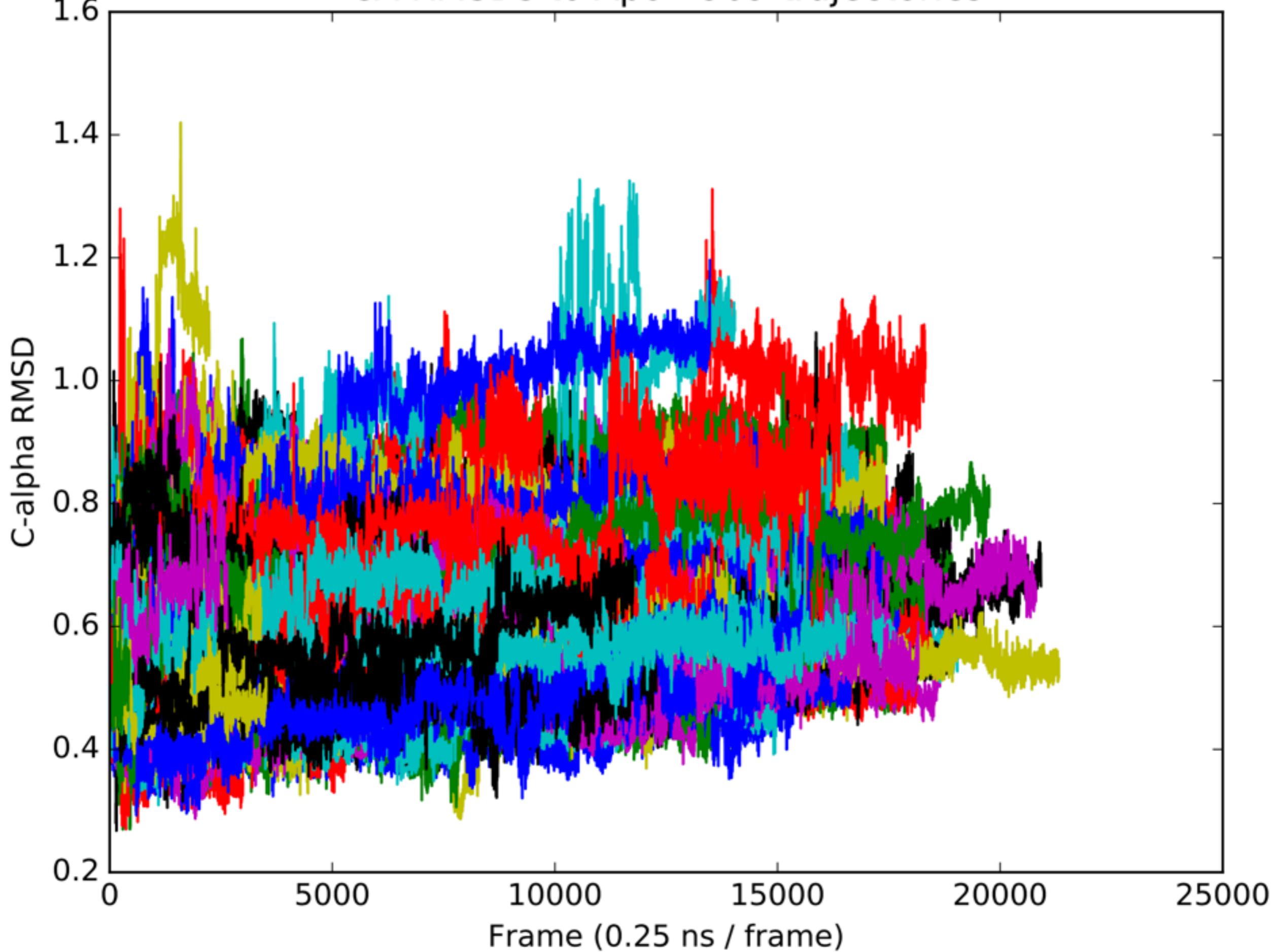
Full protein C-alpha RMSD to first frame of a trajectory

CA RMSDs to 1st frame - 969 trajectories

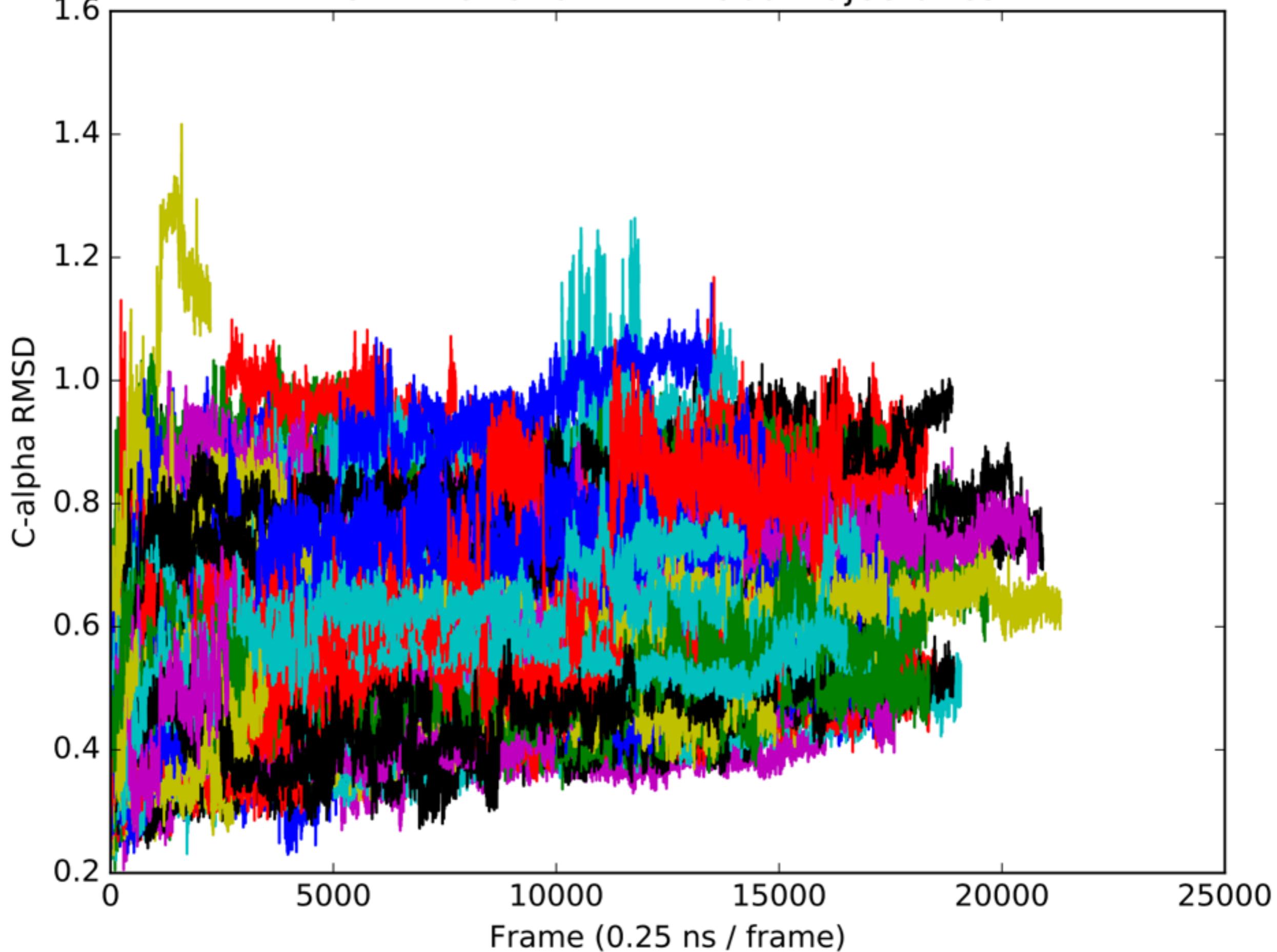


Full protein C-alpha RMSD to reference crystal structures

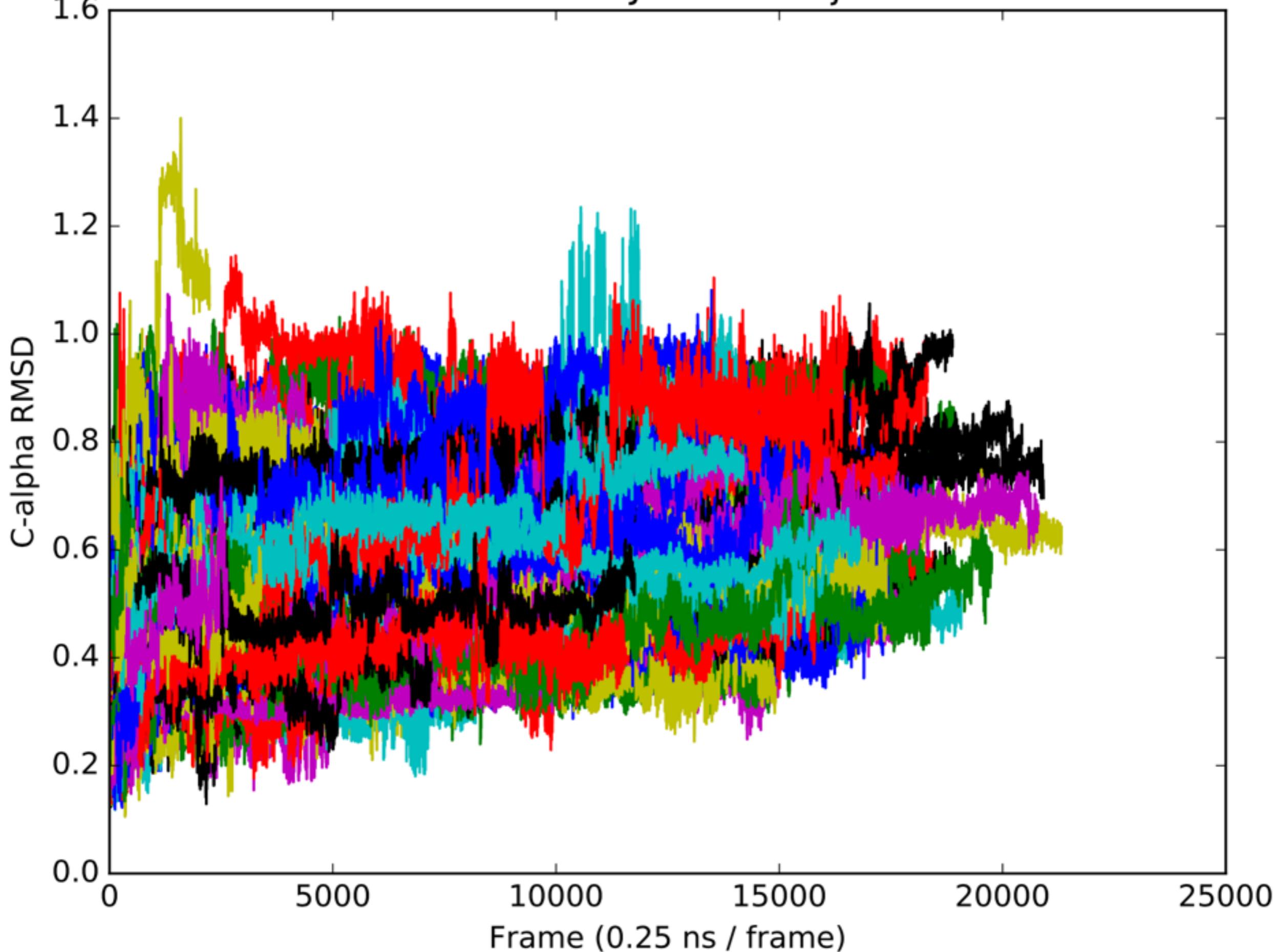
CA RMSDs to Apo - 969 trajectories



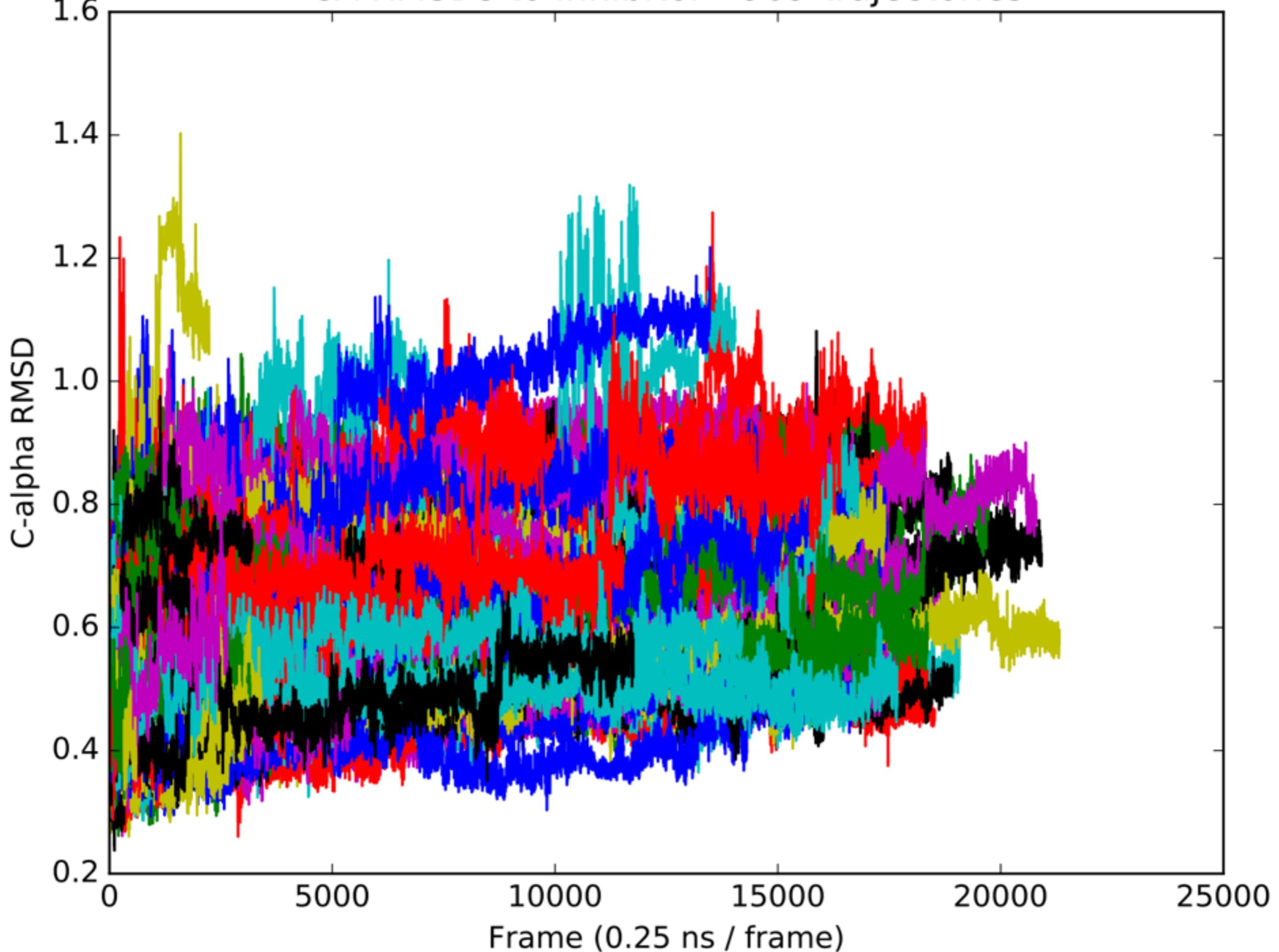
CA RMSDs to 1ZKK - 969 trajectories



CA RMSDs to 4IJ8 - 969 trajectories

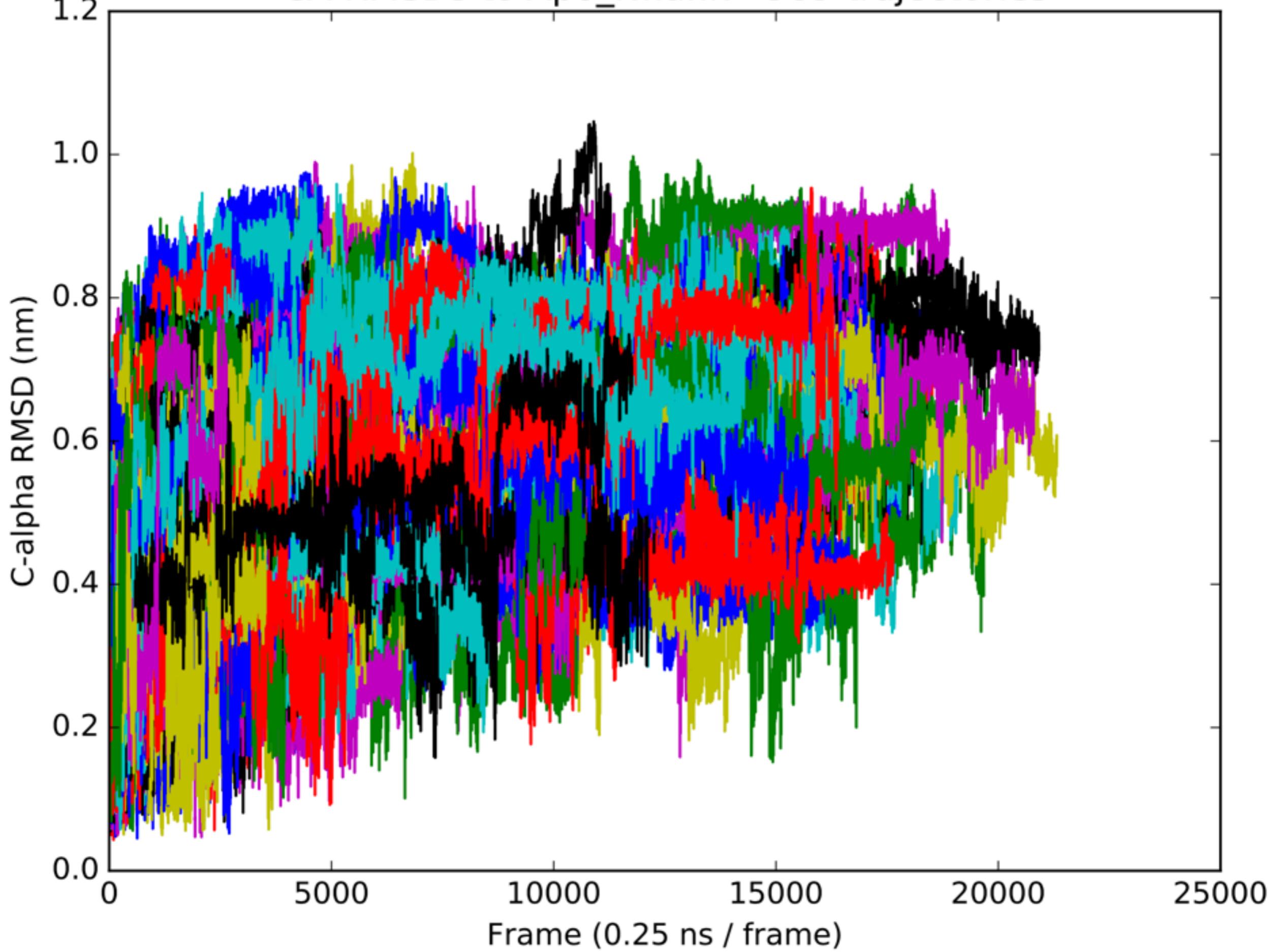


CA RMSDs to Inhibitor - 969 trajectories

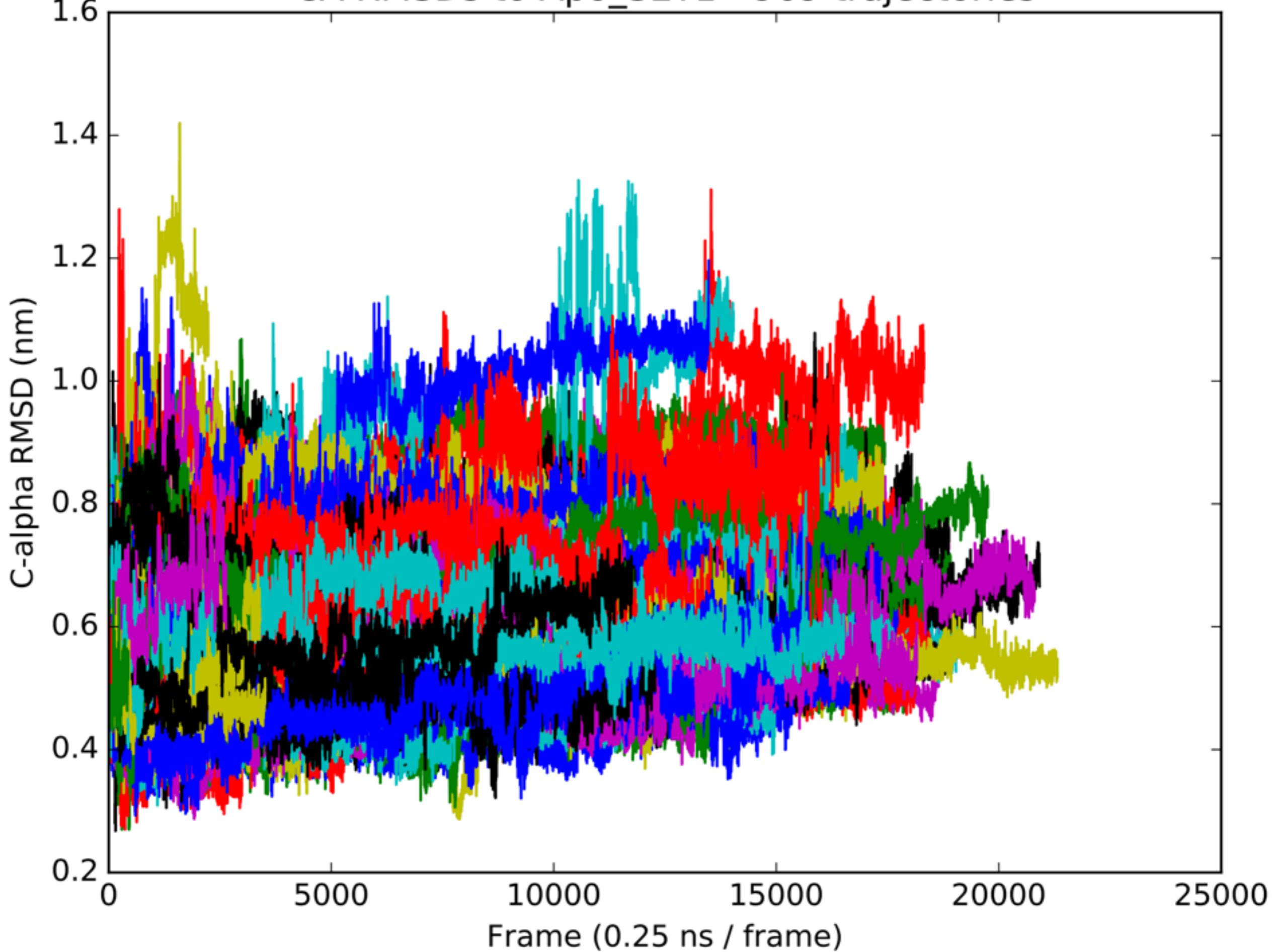


C-alpha RMSDs by domain, to reference crystal structures

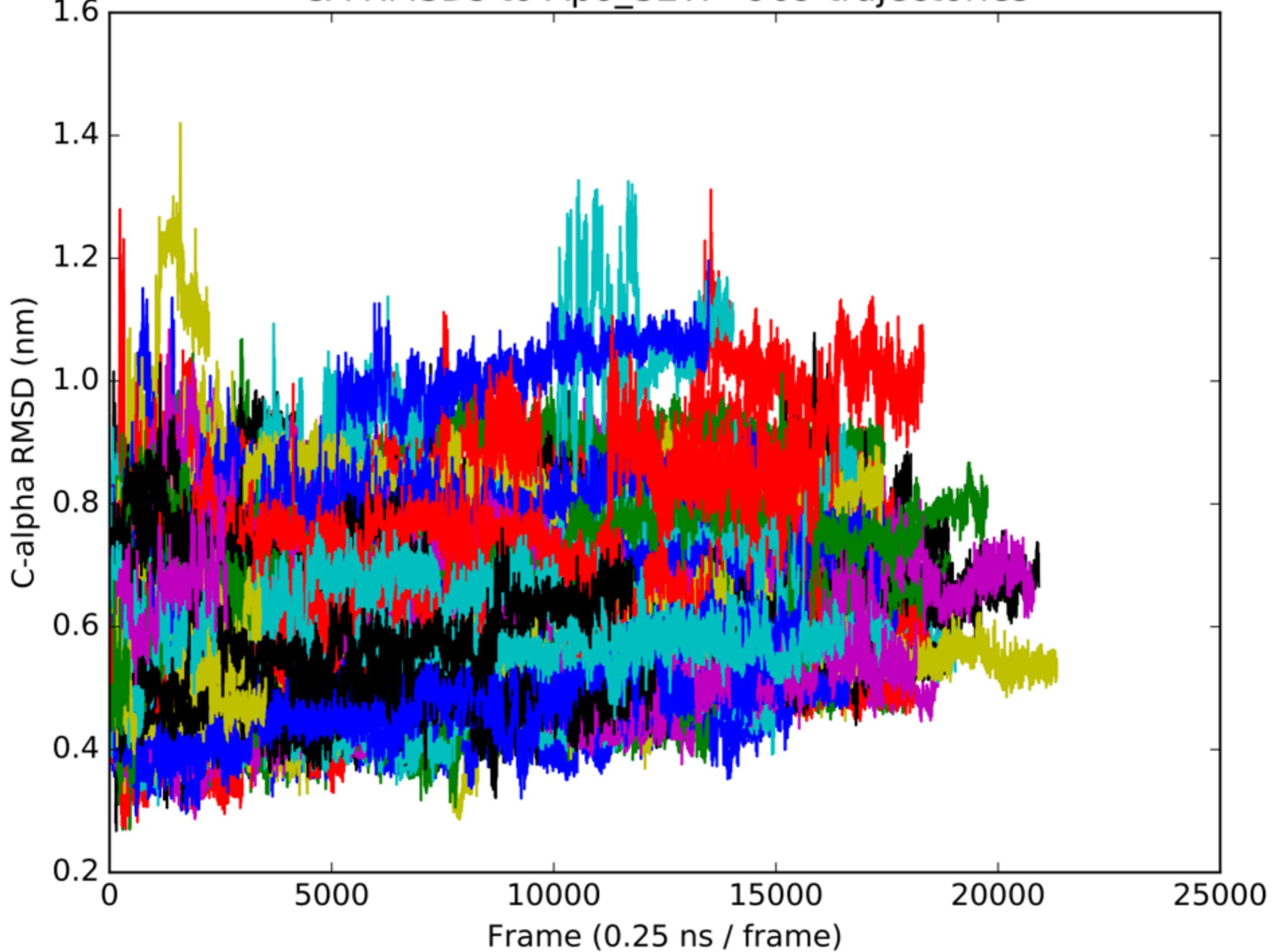
CA RMSDs to Apo_Nflank - 969 trajectories



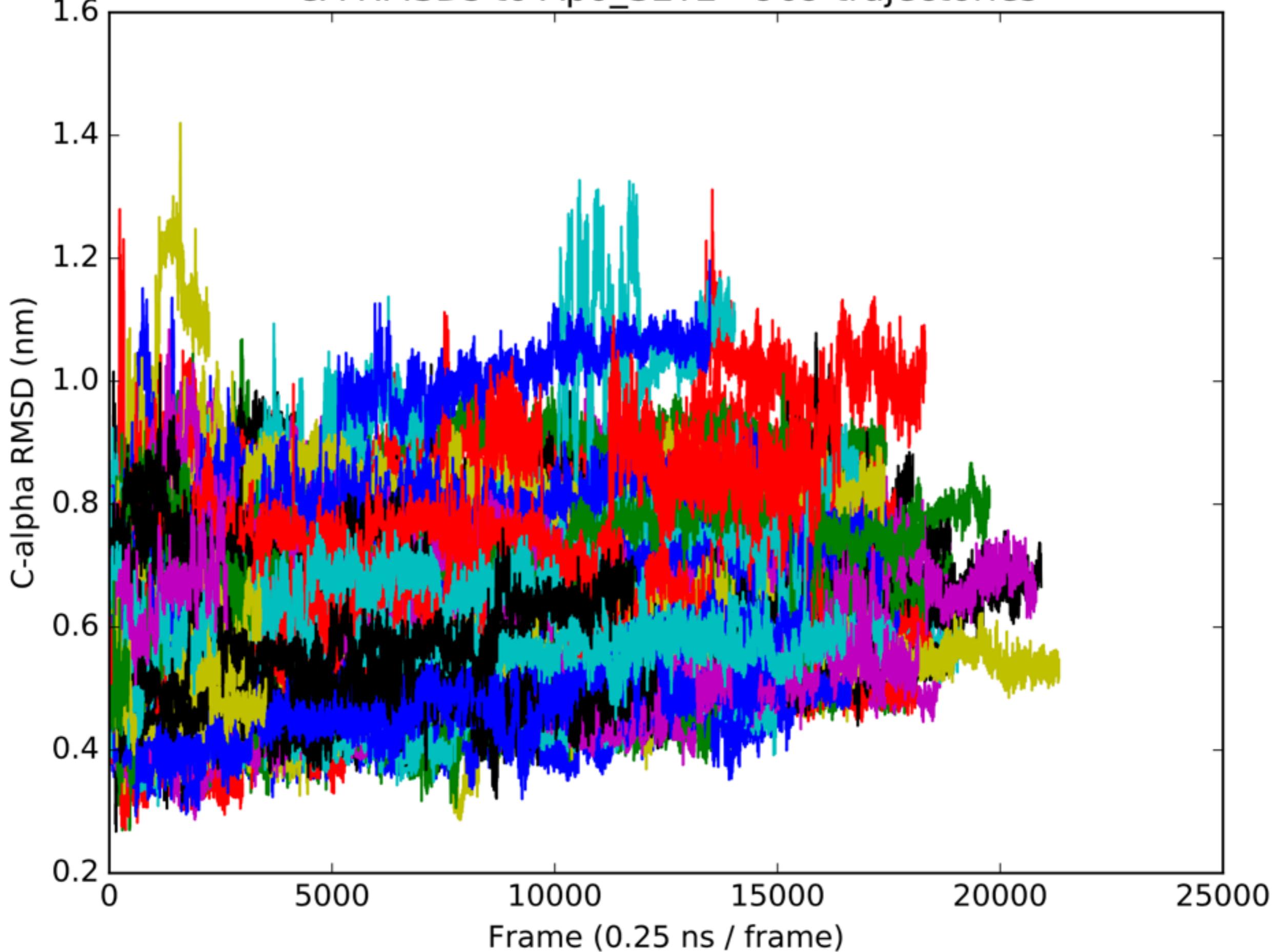
CA RMSDs to Apo_SET1 - 969 trajectories



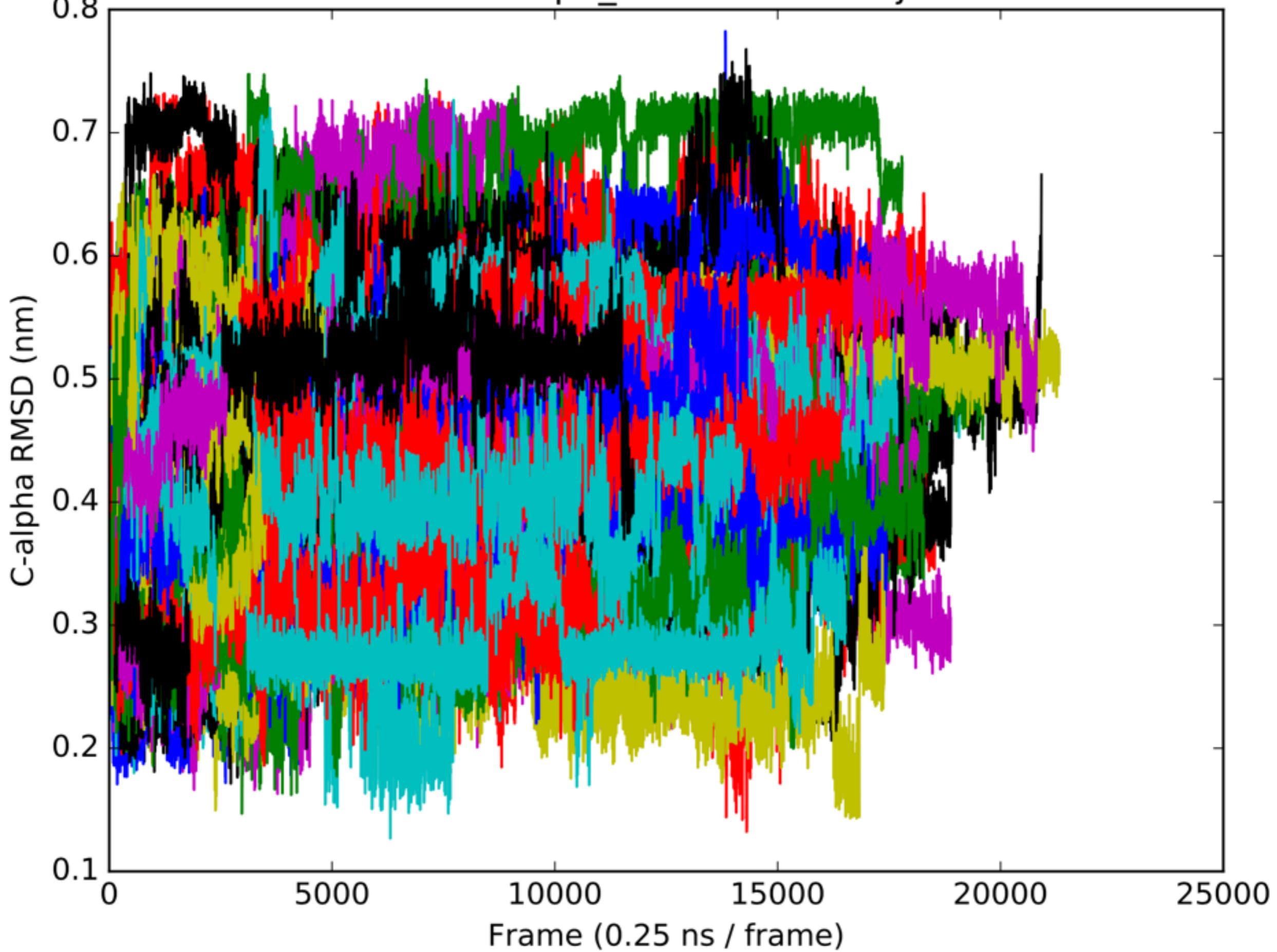
CA RMSDs to Apo_SETI - 969 trajectories



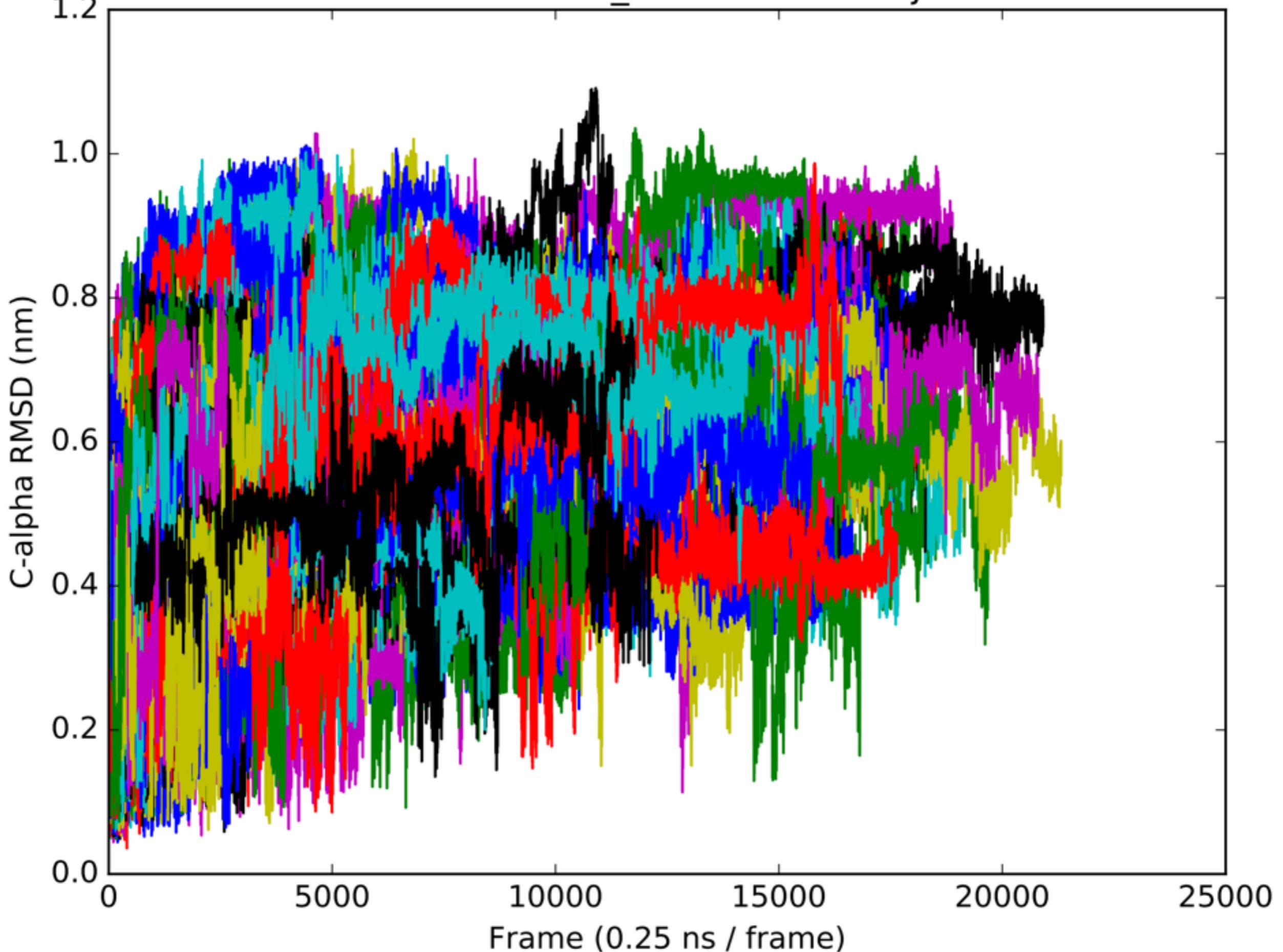
CA RMSDs to Apo_SET2 - 969 trajectories



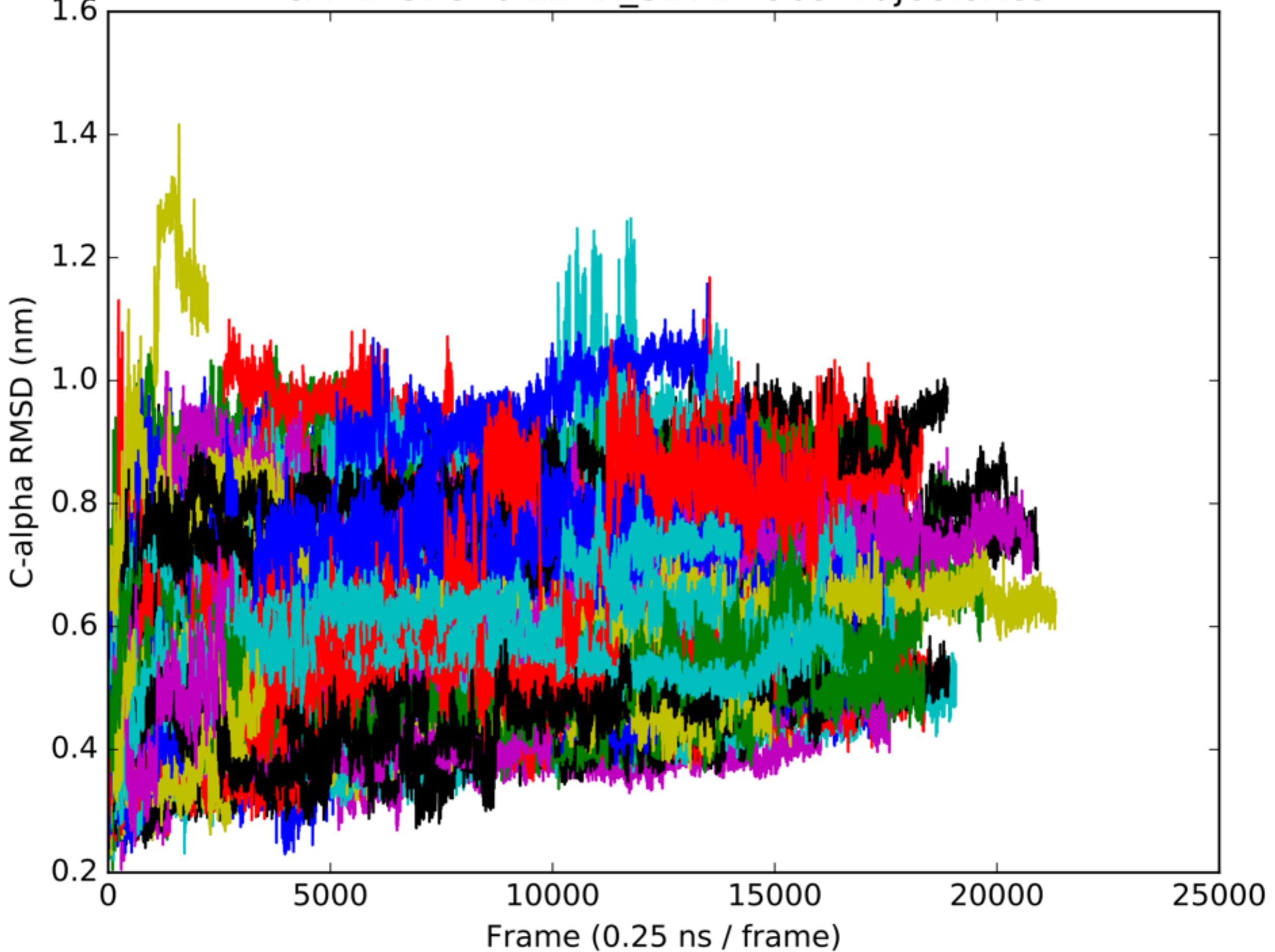
CA RMSDs to Apo_Cflank - 969 trajectories



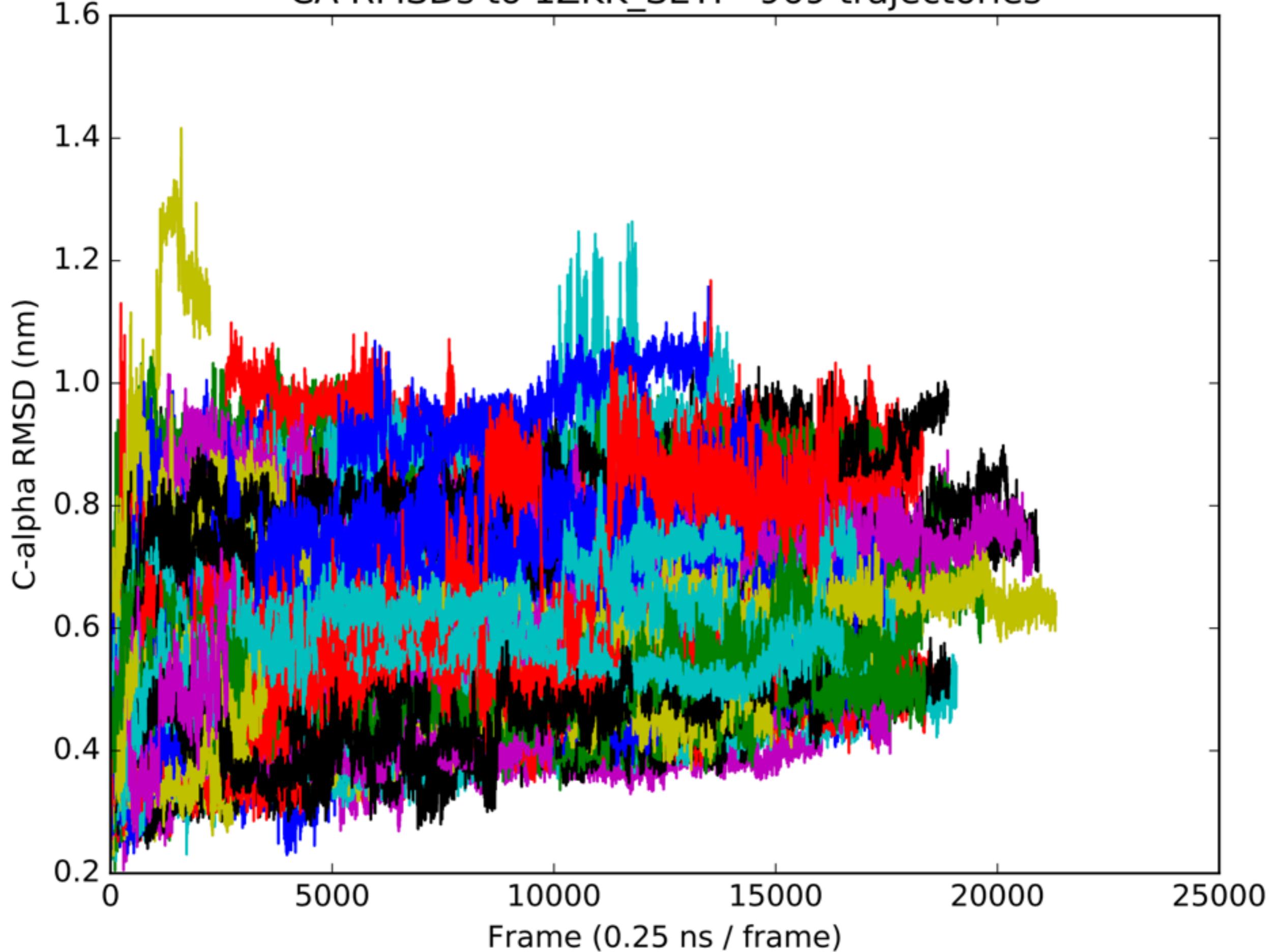
CA RMSDs to 1ZKK_Nflank - 969 trajectories



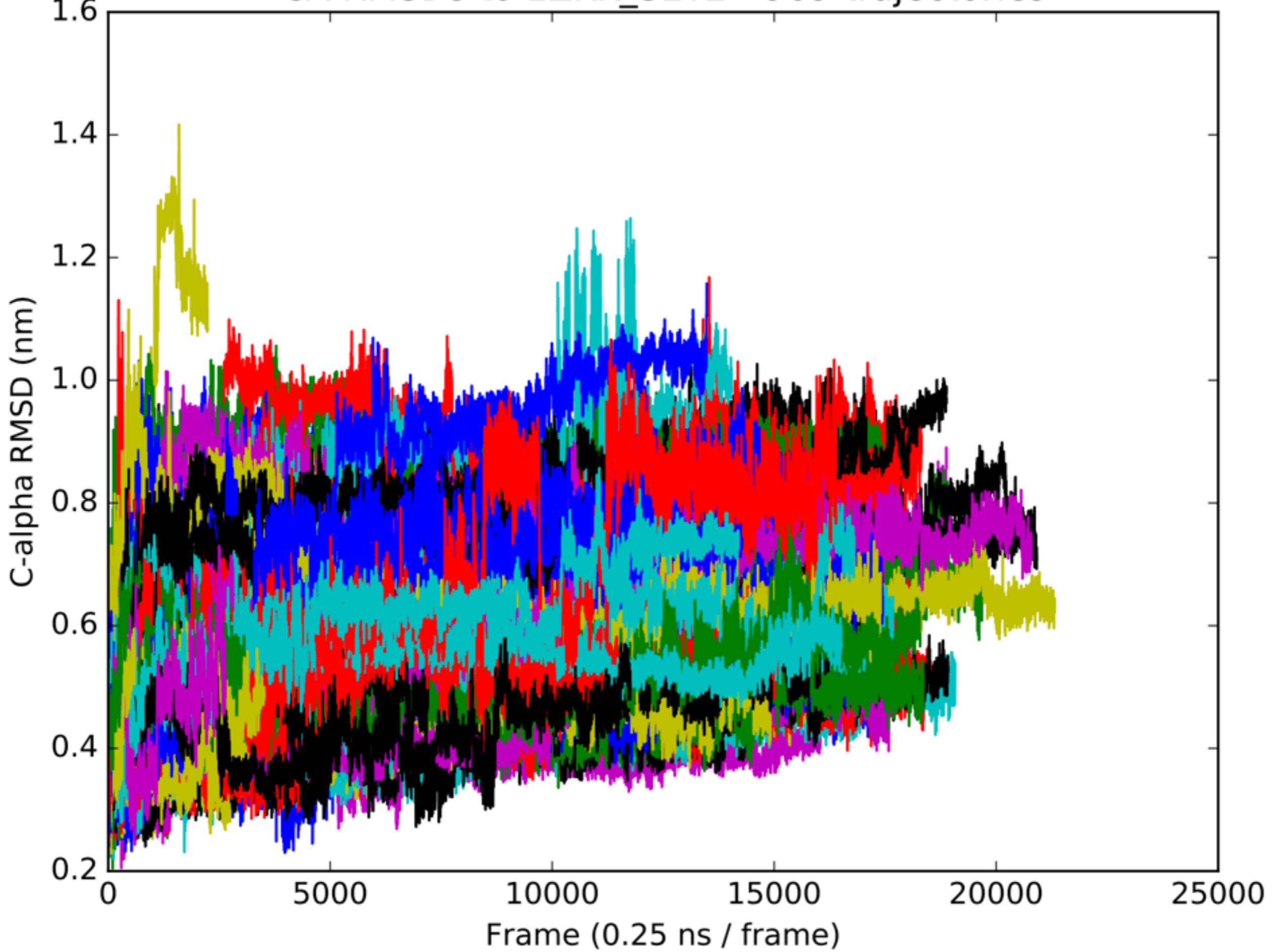
CA RMSDs to 1ZKK_SET1 - 969 trajectories



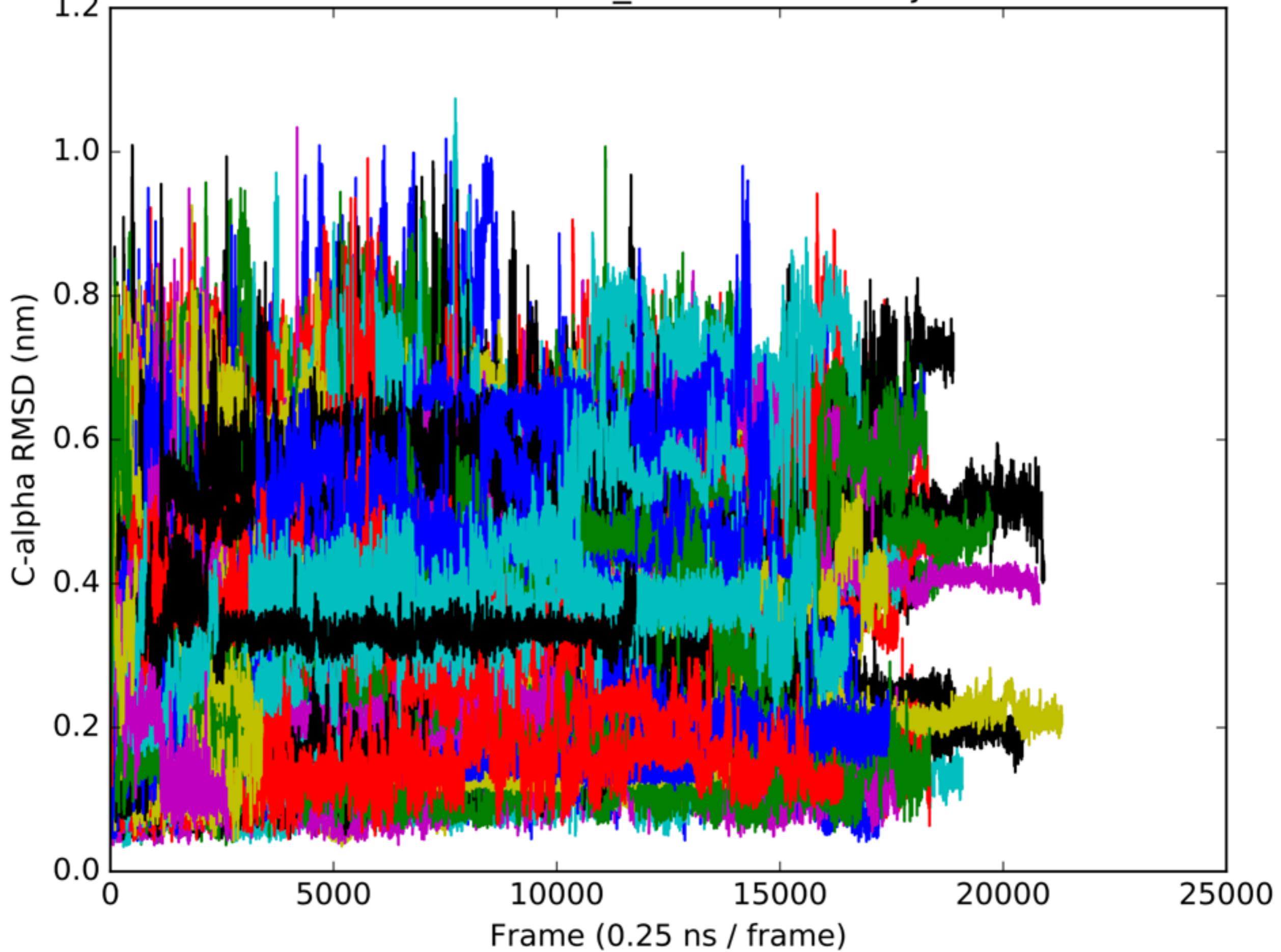
CA RMSDs to 1ZKK_SETI - 969 trajectories



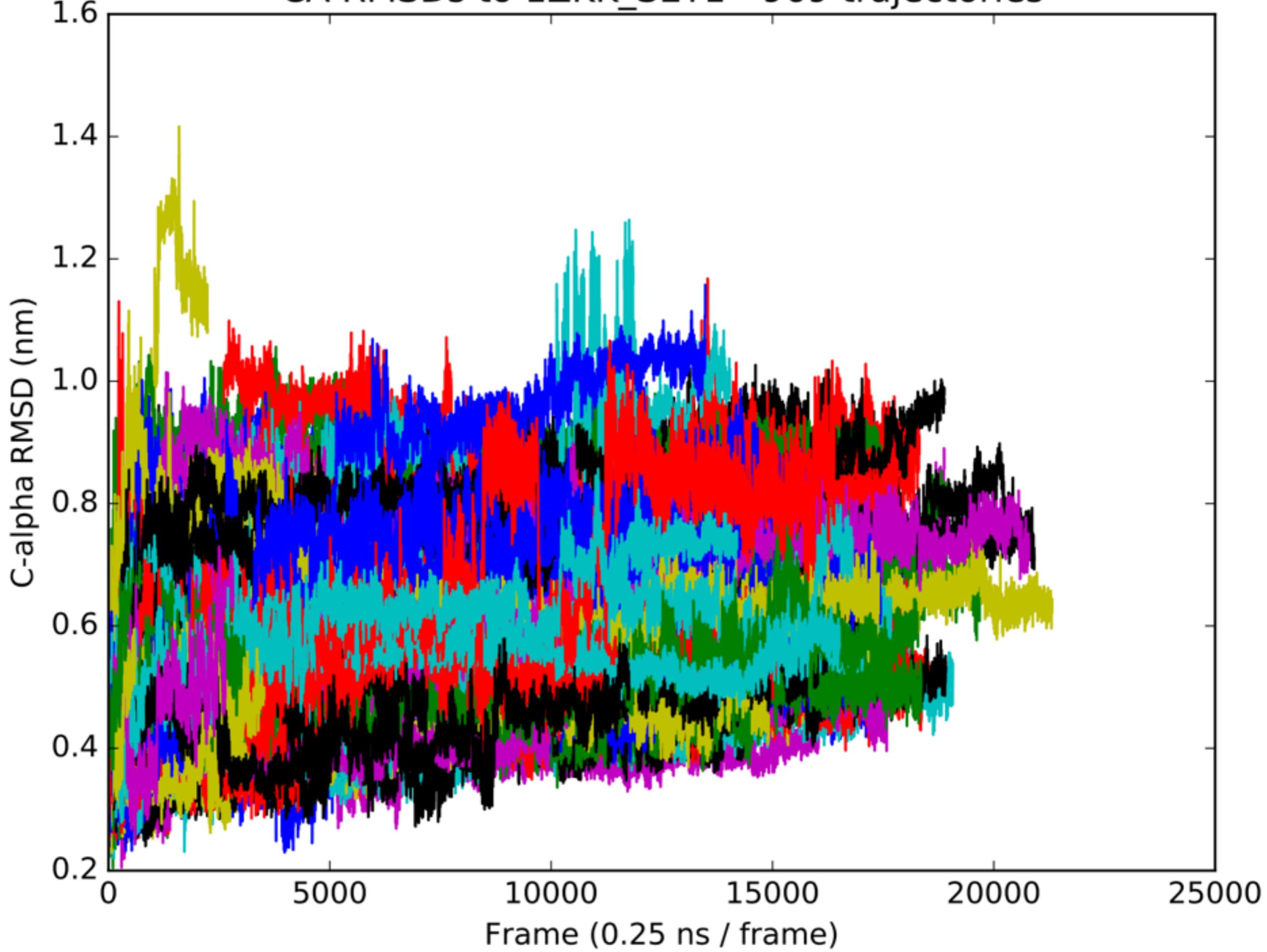
CA RMSDs to 1ZKK_SET2 - 969 trajectories



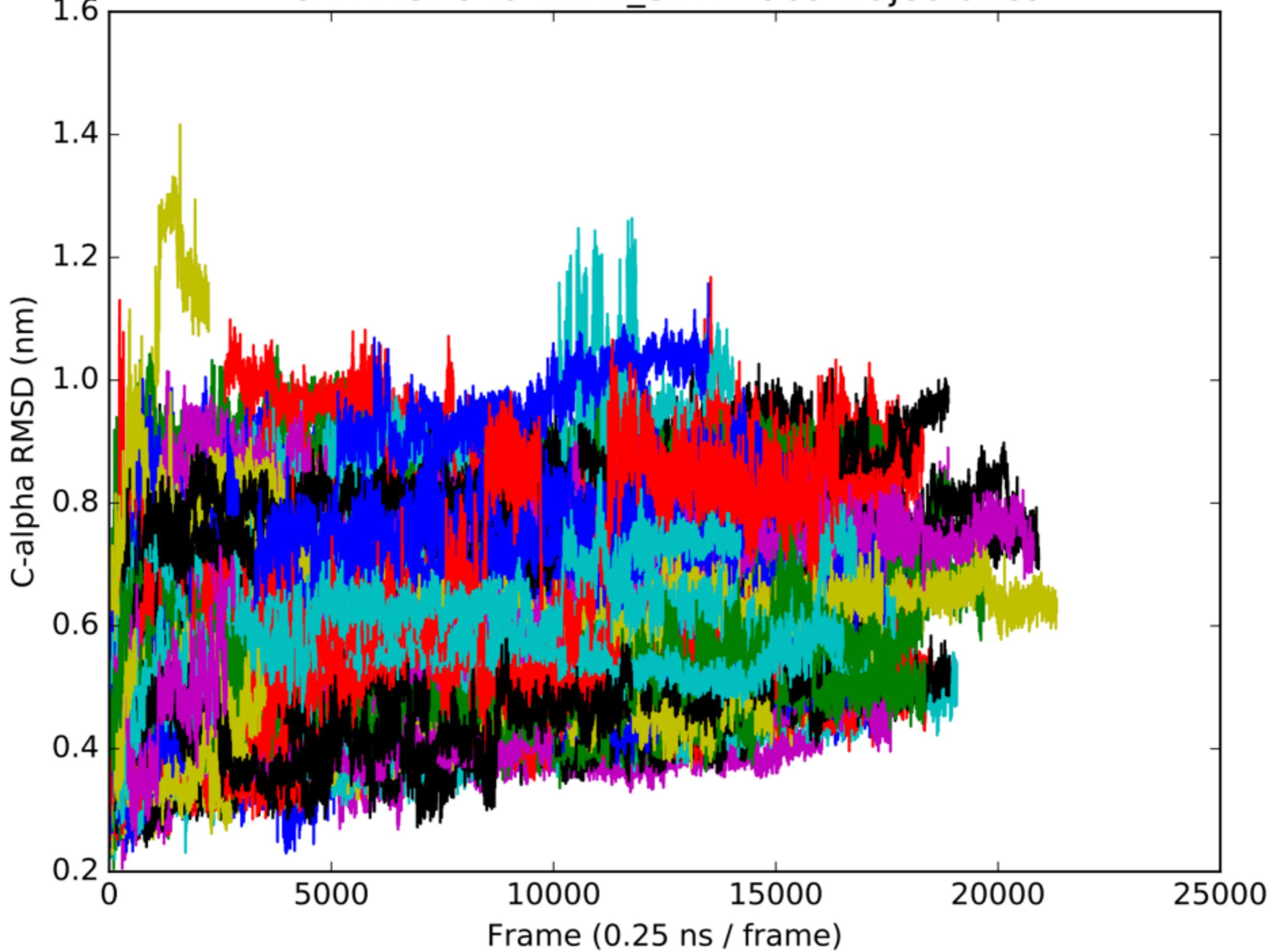
CA RMSDs to 1ZKK_Cflank - 969 trajectories



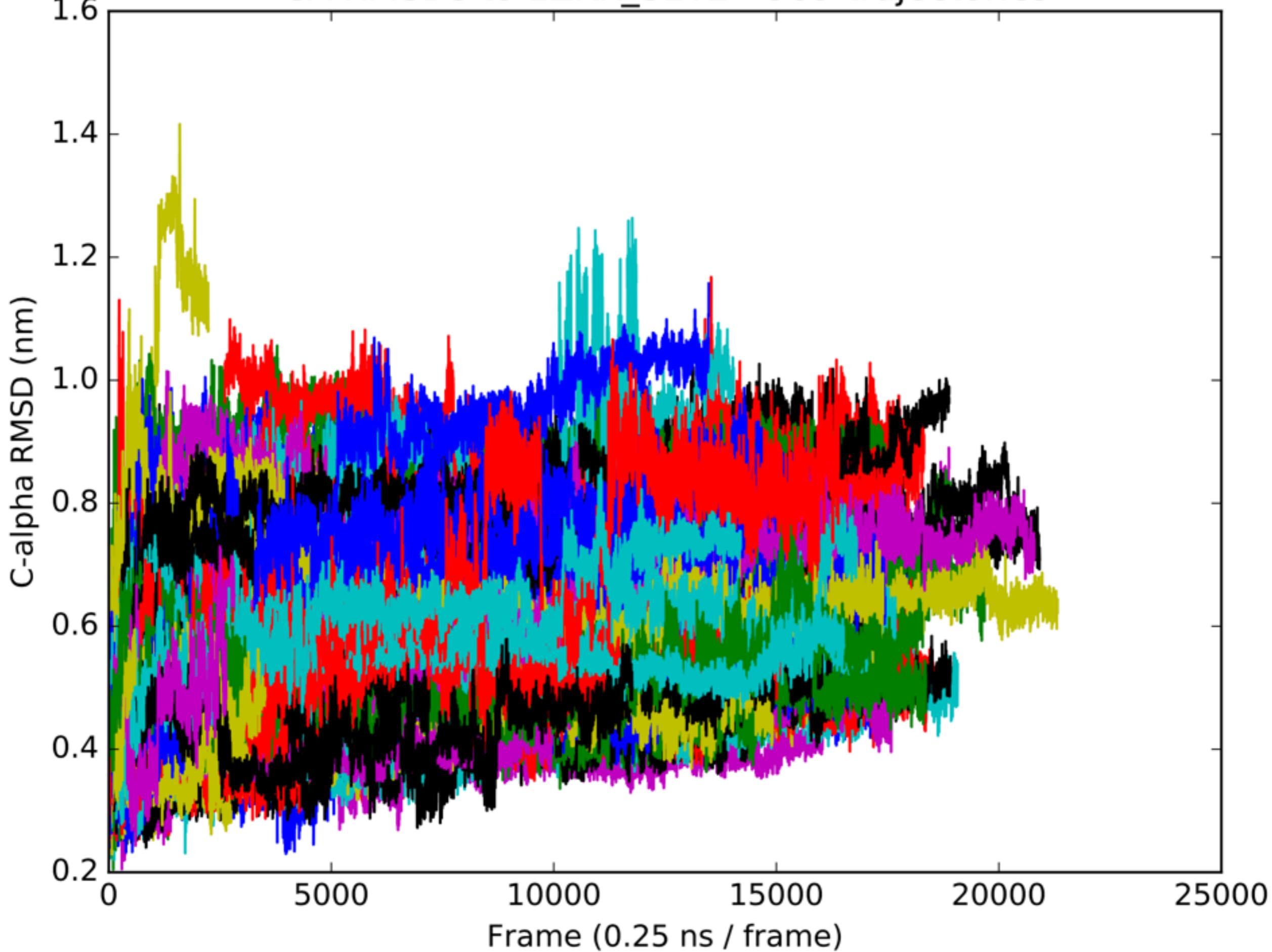
CA RMSDs to 1ZKK_SET1 - 969 trajectories



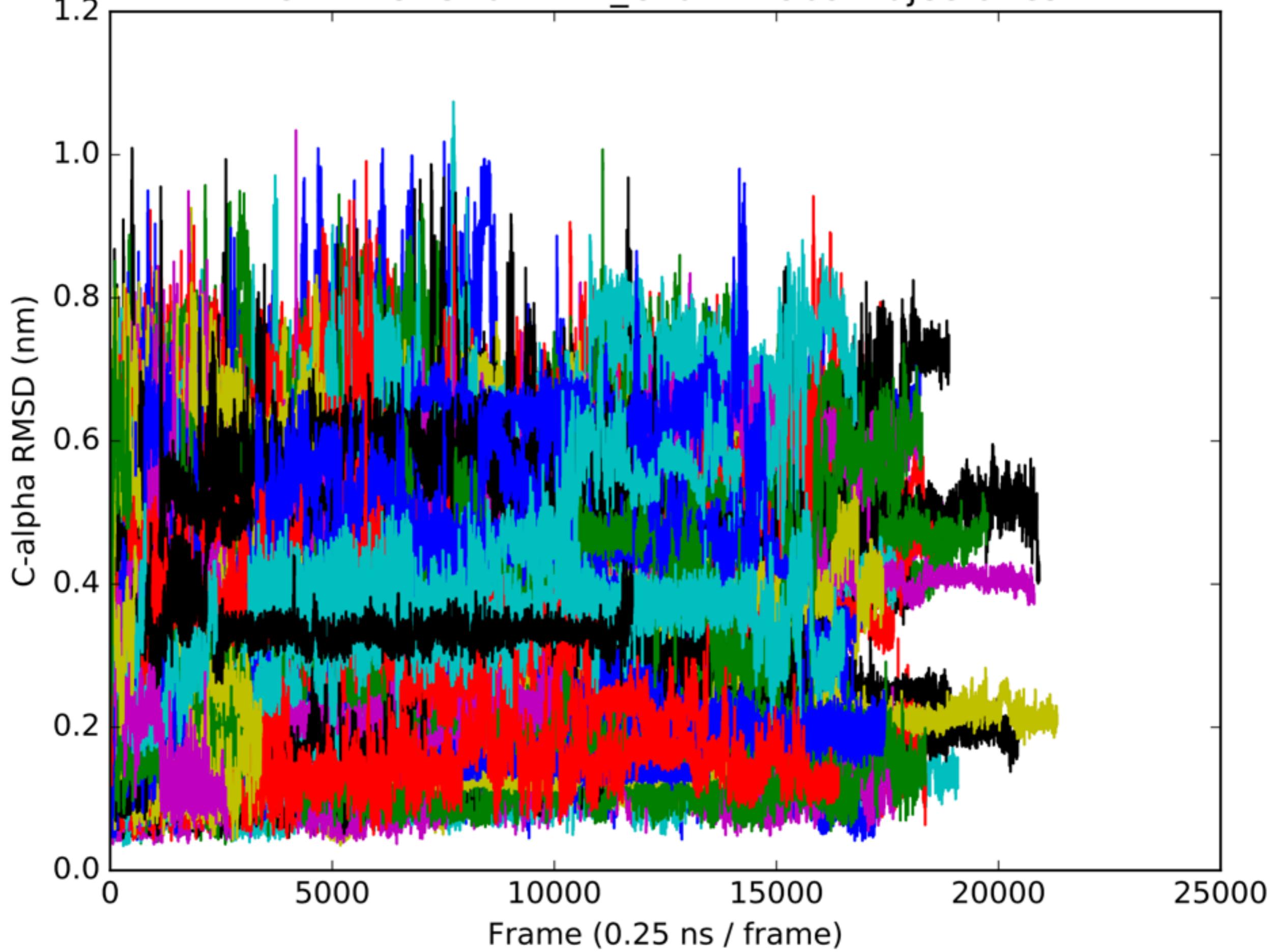
CA RMSDs to 1ZKK_SETI - 969 trajectories



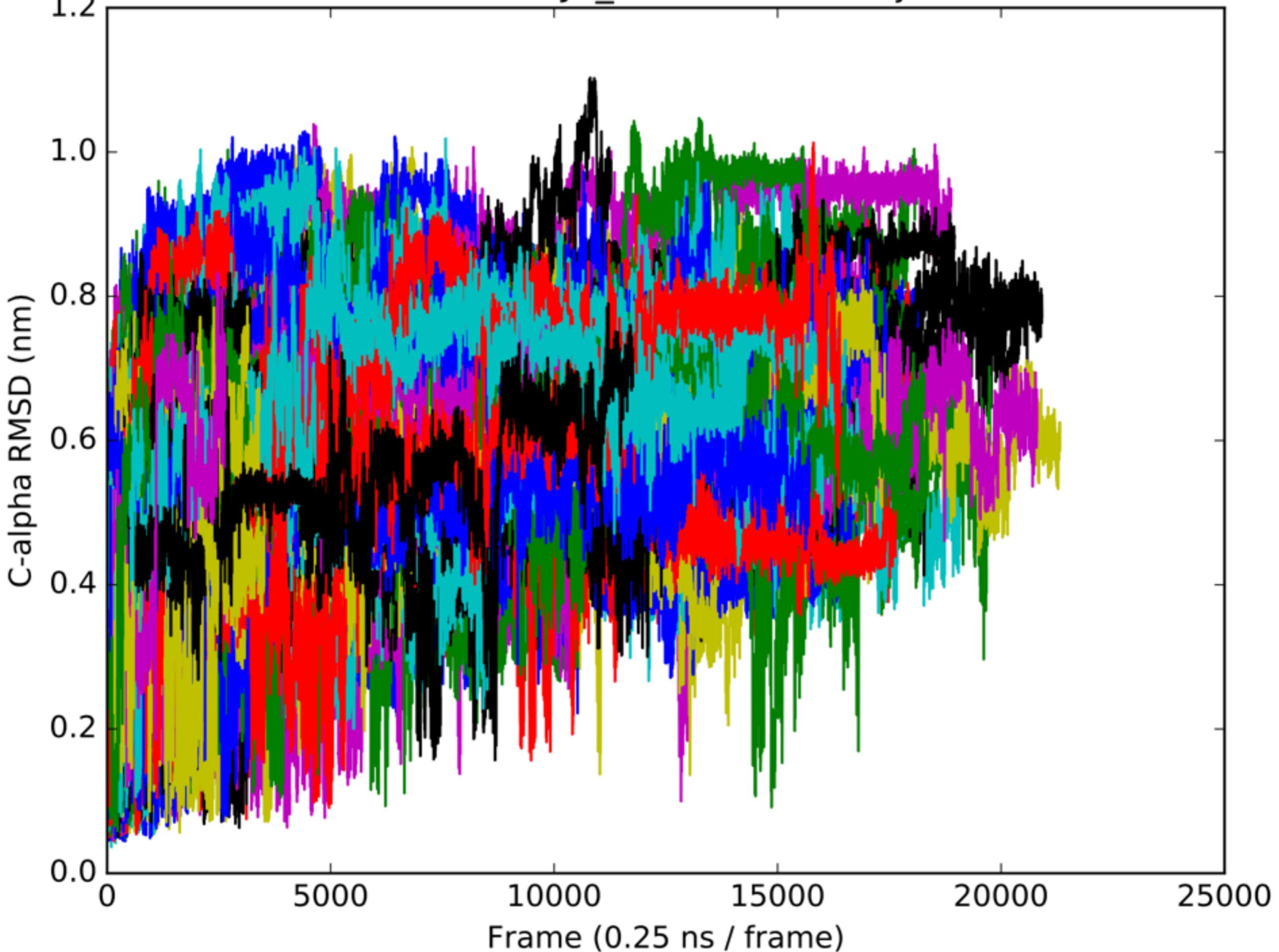
CA RMSDs to 1ZKK_SET2 - 969 trajectories



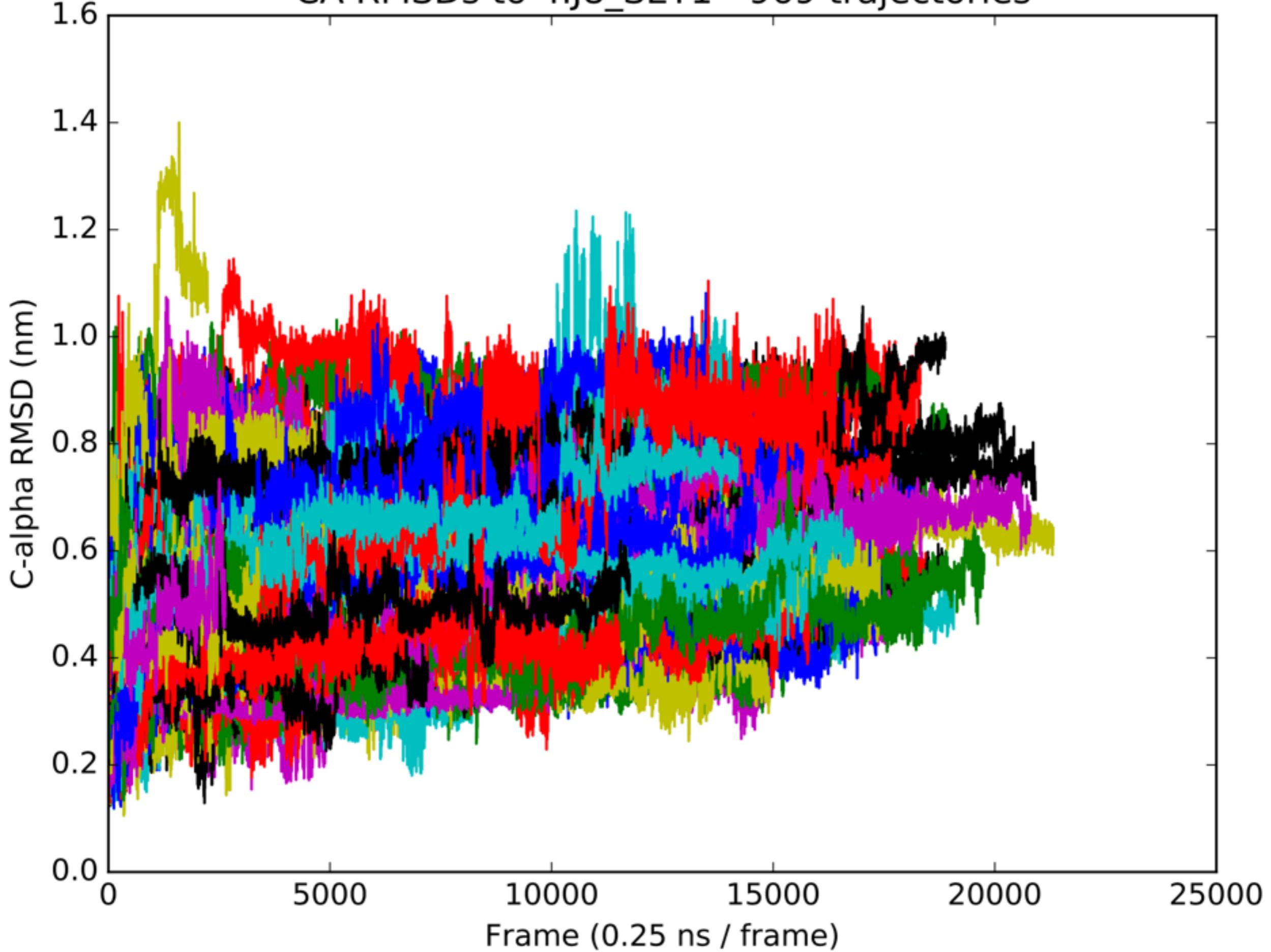
CA RMSDs to 1ZKK_Cflank - 969 trajectories



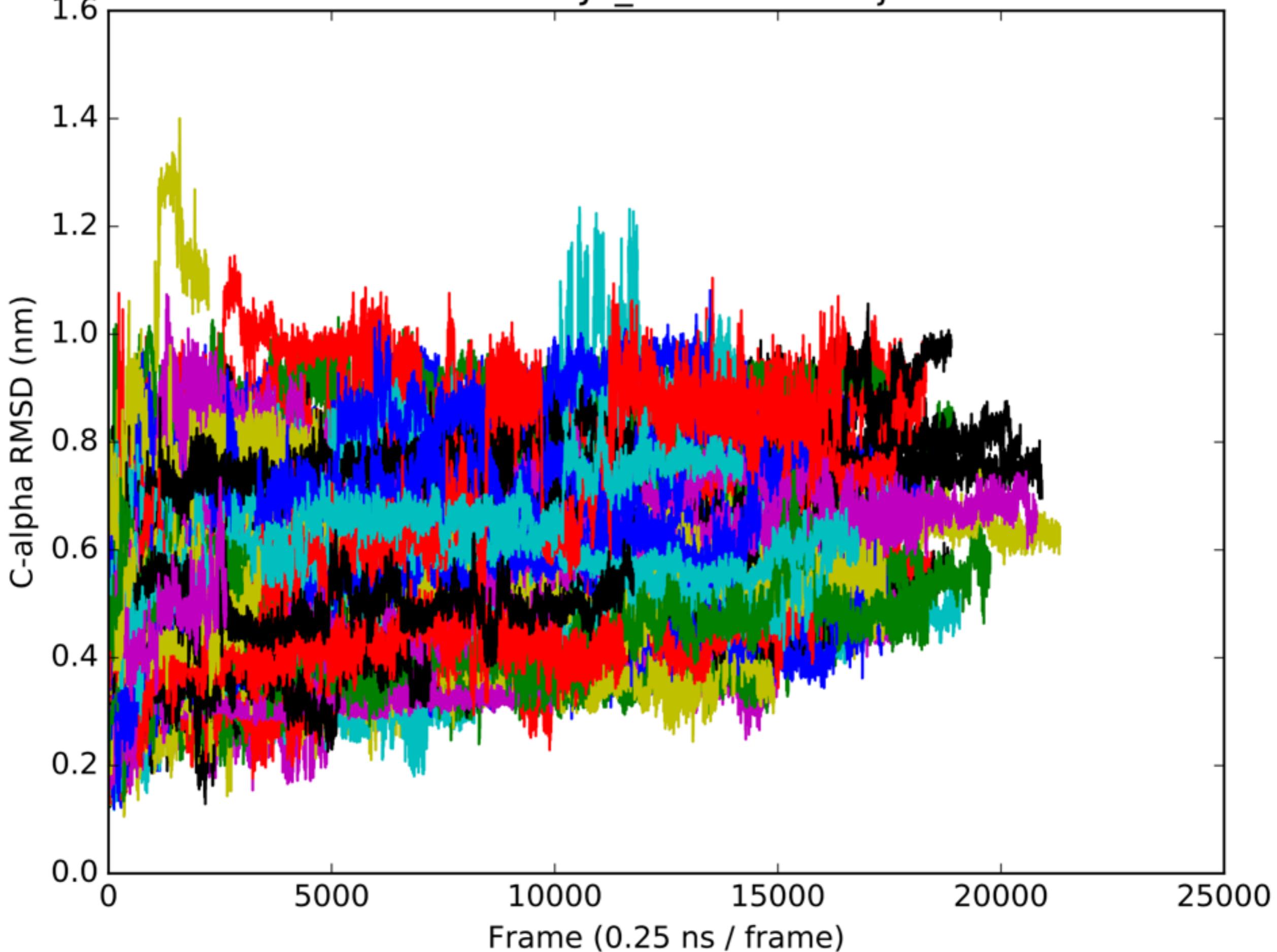
CA RMSDs to 4IJ8_Nflank - 969 trajectories



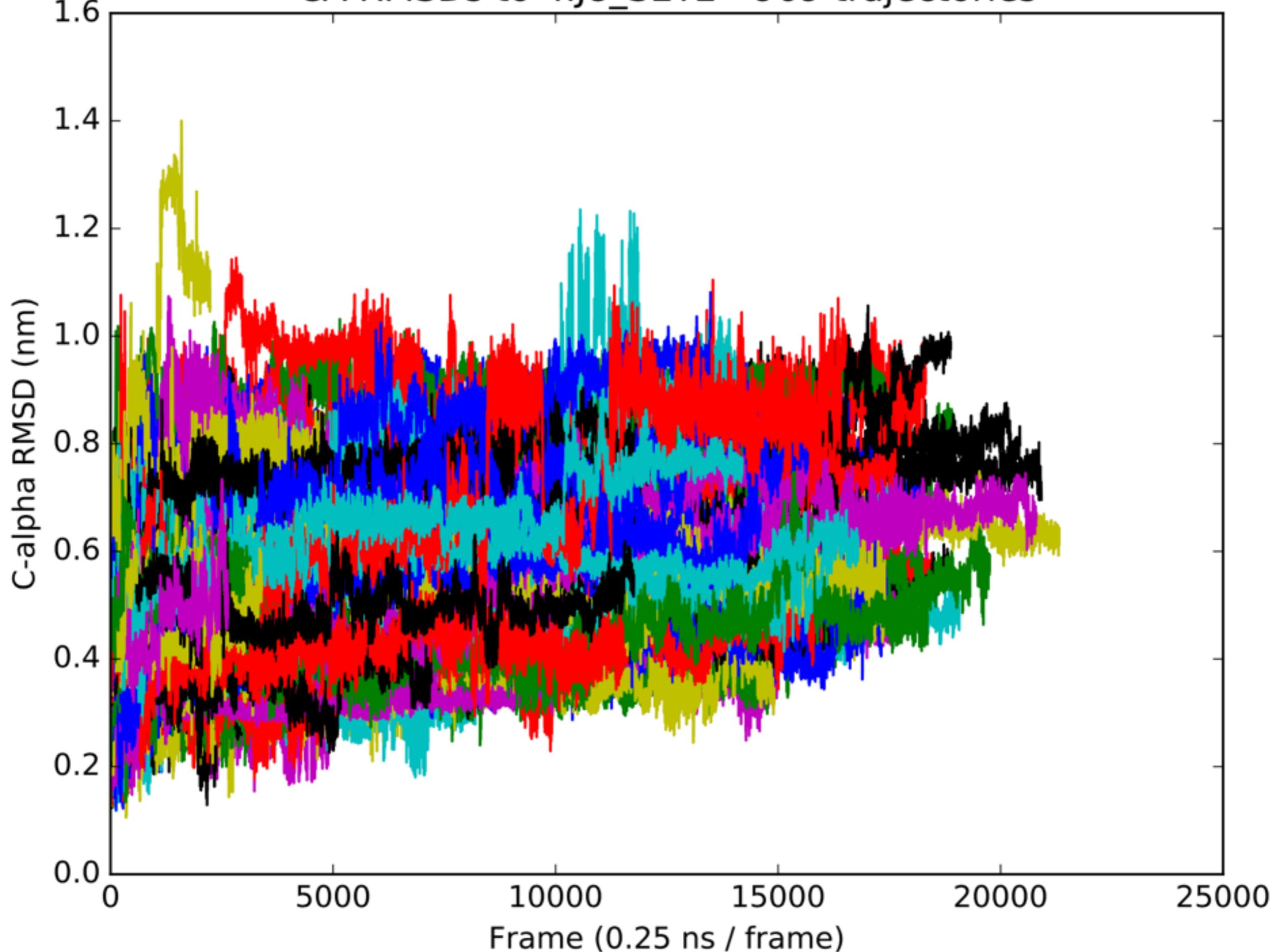
CA RMSDs to 4IJ8_SET1 - 969 trajectories



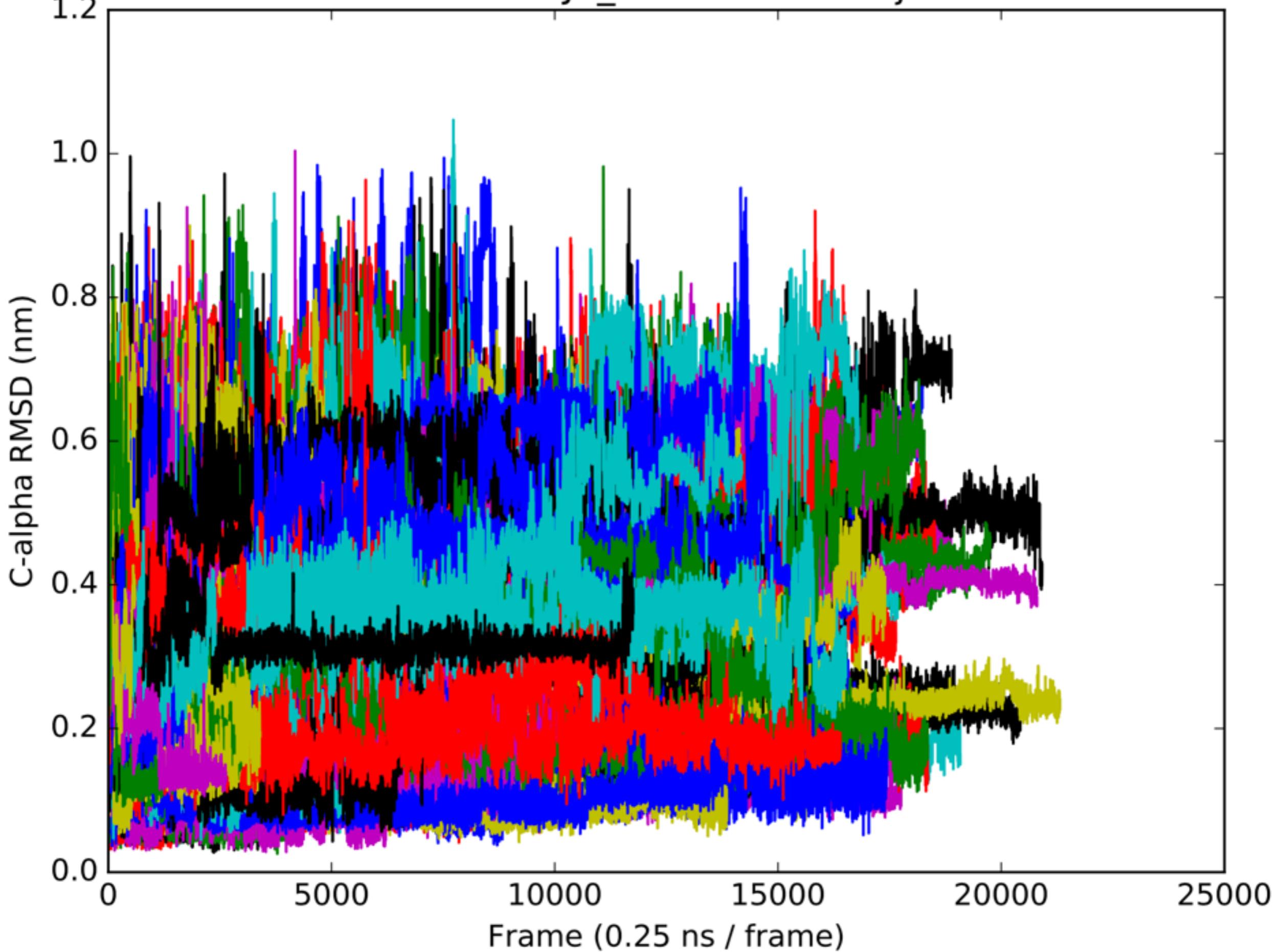
CA RMSDs to 4IJ8_SETI - 969 trajectories



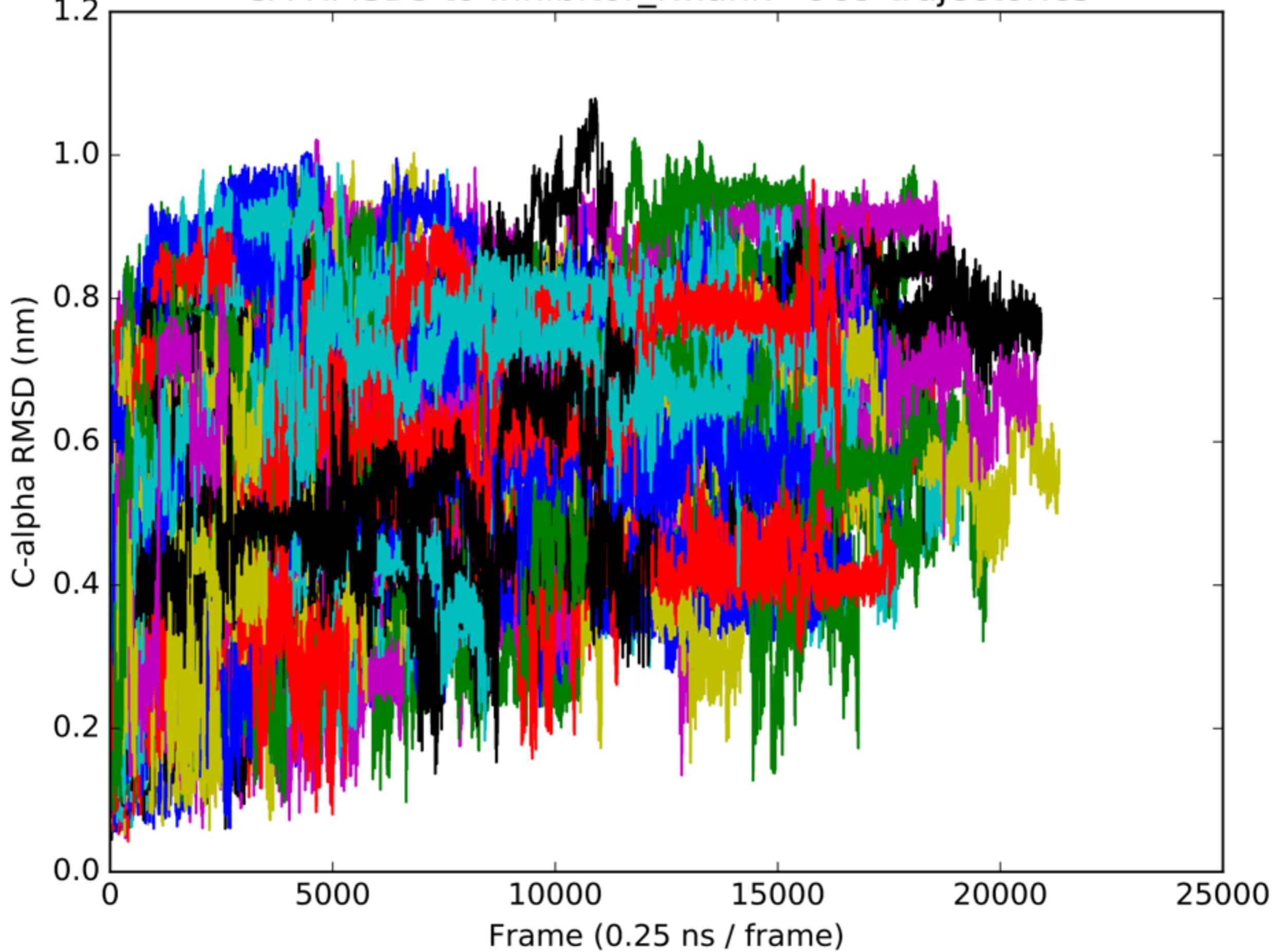
CA RMSDs to 4IJ8_SET2 - 969 trajectories



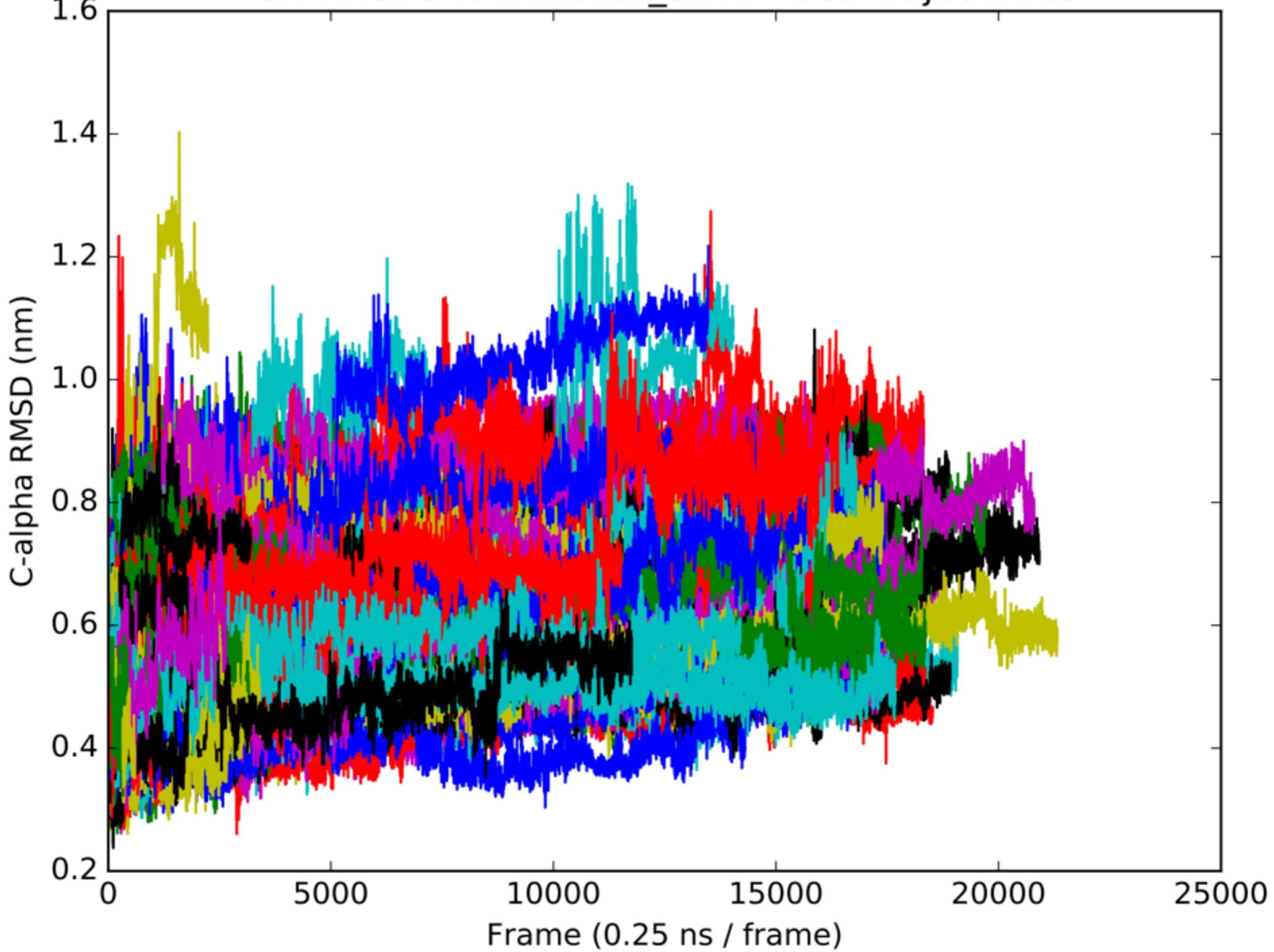
CA RMSDs to 4IJ8_Cflank - 969 trajectories



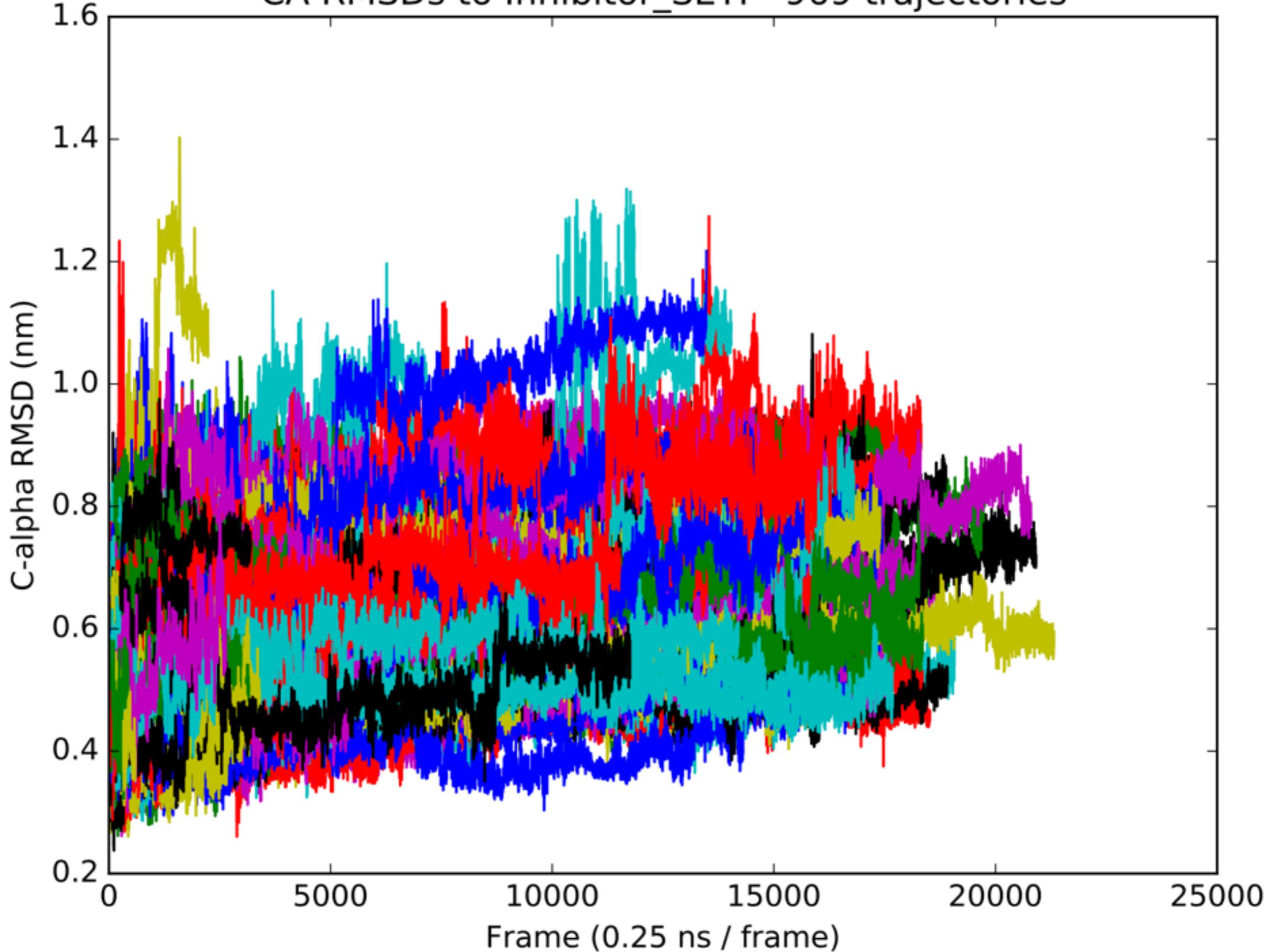
CA RMSDs to Inhibitor_Nflank - 969 trajectories



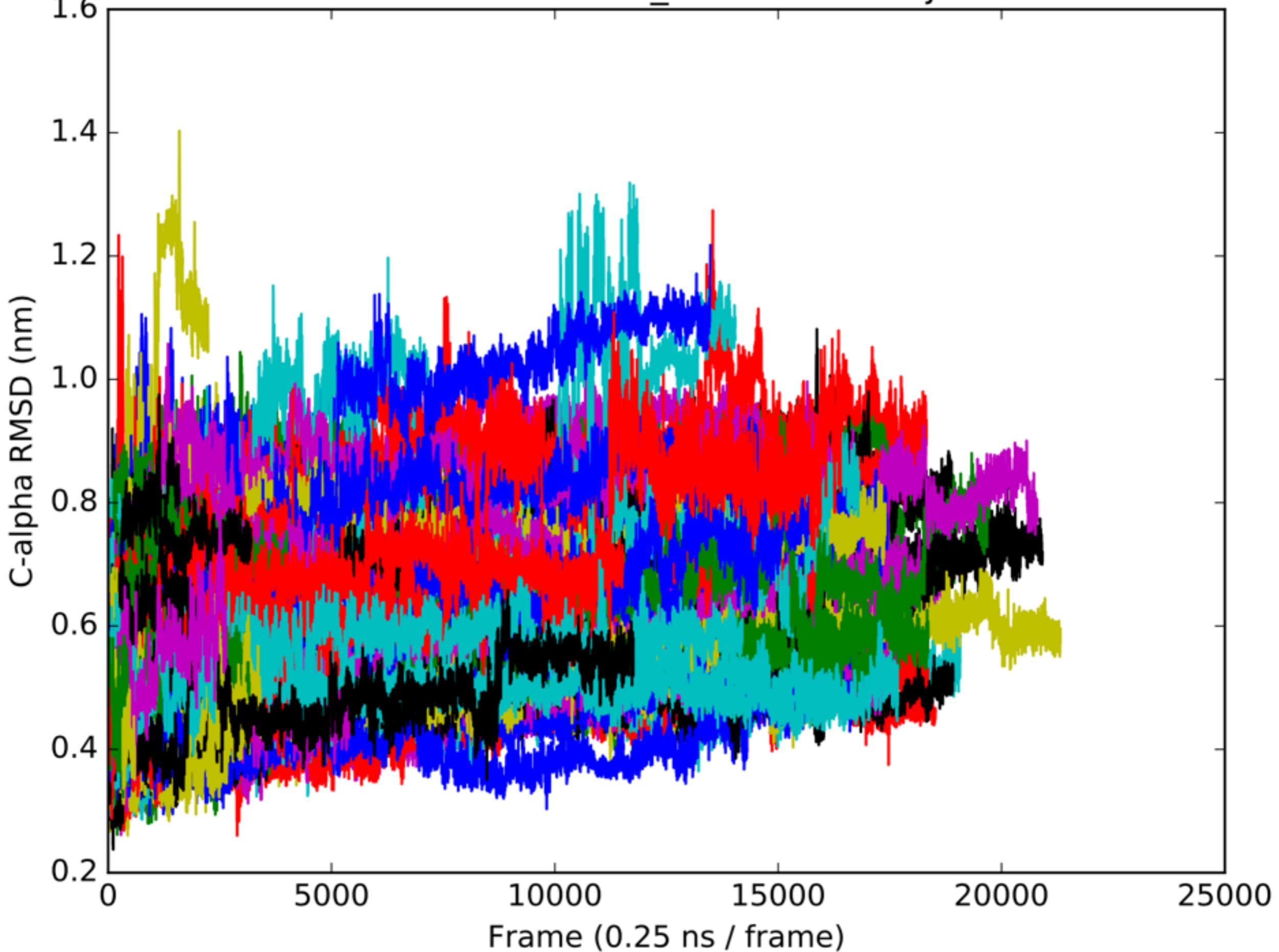
CA RMSDs to Inhibitor_SET1 - 969 trajectories



CA RMSDs to Inhibitor_SETI - 969 trajectories



CA RMSDs to Inhibitor_SET2 - 969 trajectories



CA RMSDs to Inhibitor_Nflank - 969 trajectories

