A Computational Structural Approach to Compare, Contrast, & Conceptualize Human Disease Models in Drosophila

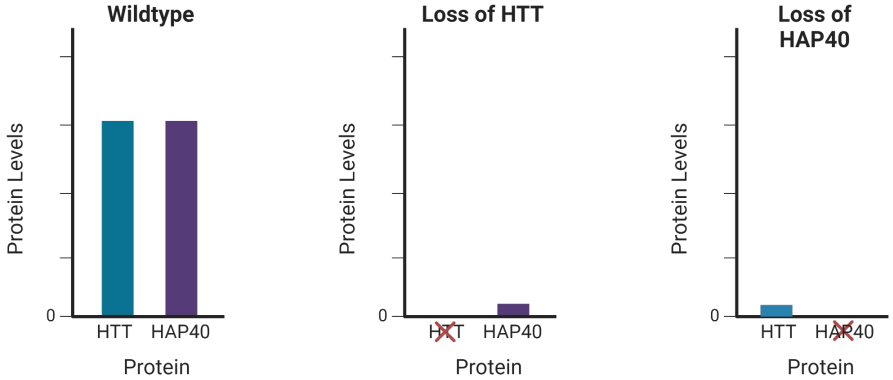
A Presentation on Stephen M. Farmer's Paper (2025)
"*Structural-functional analyses of the huntingtin/ HAP40 complex in Drosophila and humans*"

What is Huntington’s disease?

- A progressive neurodegenerative disease with the following symptoms
  - Jerky uncoordinated movement
  - Cognitive and behavioral deficits
- Linked to CAG repeat expansions that encode poly q tract in HTT protein
  - Leads to alterations in the striatum degenerating the basal ganglia’s modulation of motor programs.
- Has no effective cure so discovering relevant drug targeting of gene products and regulatory elements is critical.
  - Especially since HTT is known to bind to numerous HTT associated proteins (HAPs) involved in numerous cellular pathways.

HAP40 is a prime suspect

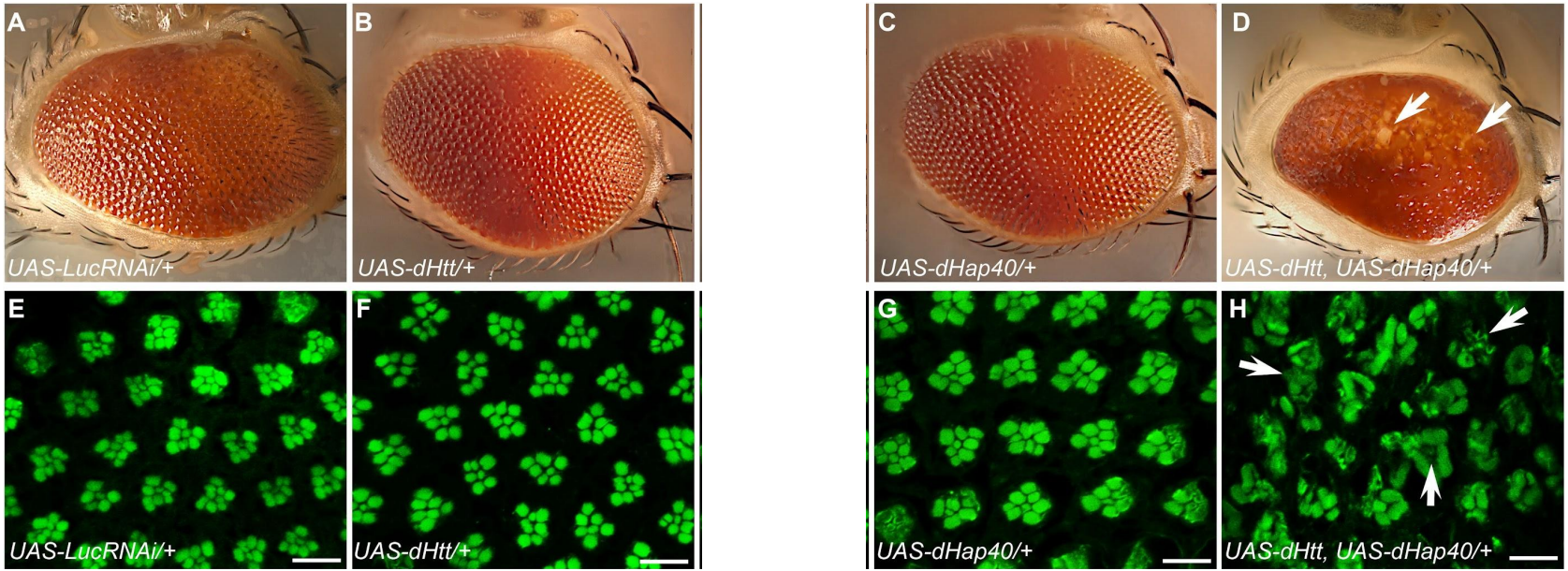
- In rat/mouse homogenates and proteomic studies of Drosophila (dHAP40), HAP40 significantly correlates with HTT
- Interacts in 1:1 molar ratio with HTT
- HAP40 changes the conformation of HTT to a more favorable globular form with 3 distinct domains
    - (1) N-HEAT (2) Bridge domain (3) C-HEAT



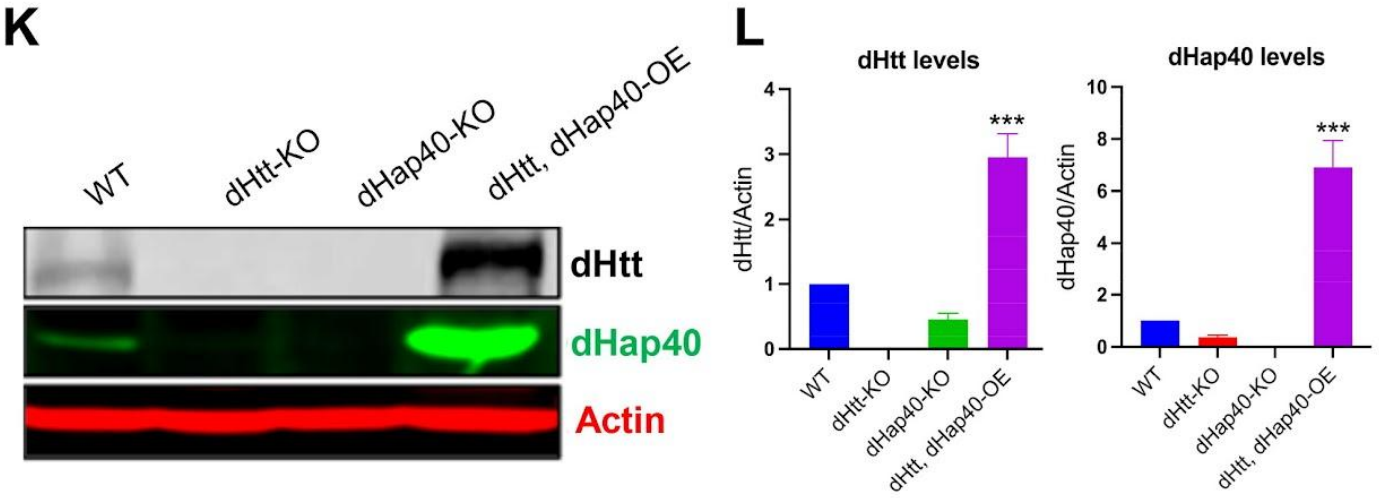
Graphs not representative of actual data. Made to visualize concepts below.

- **Previous research shows the loss of HTT nearly depletes HAP40 levels through degradation**
- **While loss of HAP40 leads to significant reduction of HTT levels**
- **Highly indicating a dependence on each other for its own existence, and the regulation control of HAP40 over HTT**

HAP40 expression with HTT is deleterious to fly phenotype




I
UAS-dHtt, UAS-dHap40/
UAS-dHtt, UAS-dHap40


K
WT
dHtt-KO
dHap40-KO
dHtt, dHap40-OE
dHtt
dHap40
Actin
L
dHtt levels
dHtt/Actin
***
WT
dHtt-KO
dHap40-KO
dHtt, dHap40-OE
dHap40 levels
dHap40/Actin
***
WT
dHtt-KO
dHap40-KO
dHtt, dHap40-OE

HAP40 is interesting. Can bioinformatics reveal more information about it in complex with HTT?

Generating protein structure for binding/interaction analysis

Obtain Experimental Data:

UniProt: Hosts DNA/Protein sequences along with numerous annotations regarding localization, function, binding partners, pathways and much more.

PDB hosts hundreds of thousands of experimentally elucidated folded protein structures.

Predict Structure Folding:

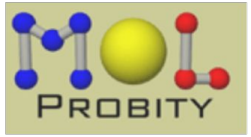
Google's DeepMind AlphaFold can confidently predict the folded structures of proteins using comparative sequences, existing structures(PDB), geometrics, and thermodynamics.

Energy Minimization:

Lowers energy of inputted PDB to energetically stabilize the structure for later modeling.

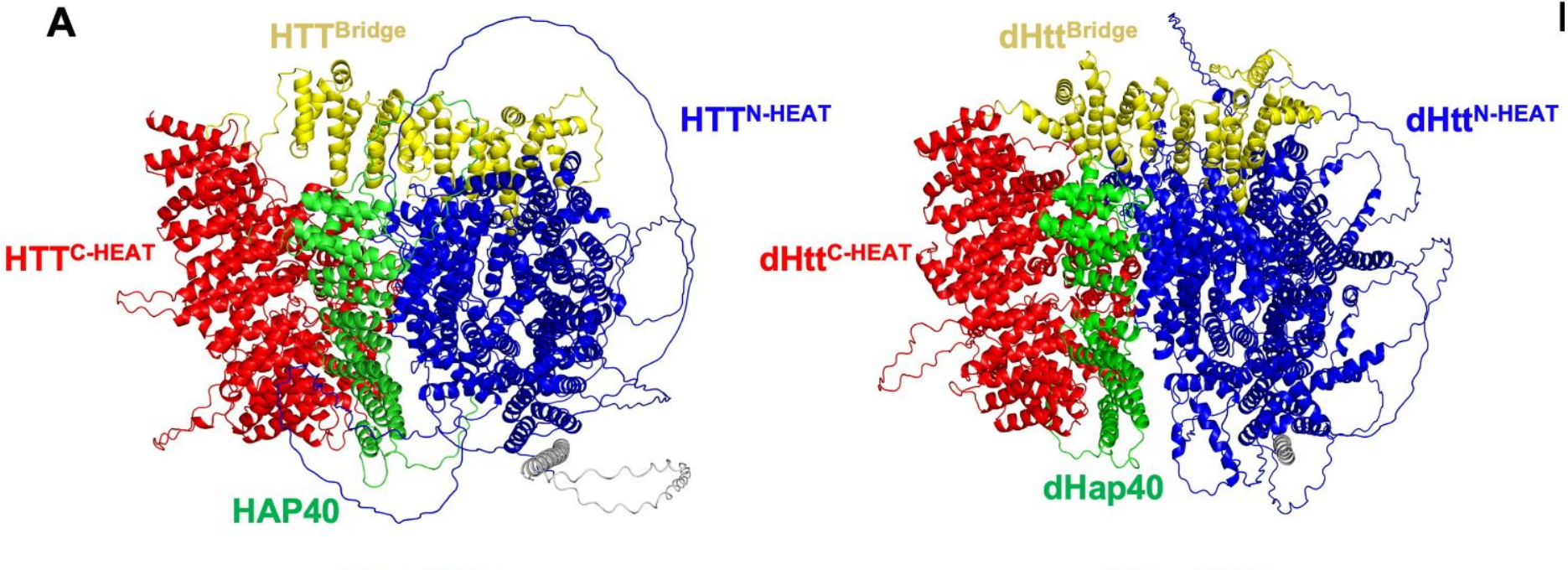
Structural Quality Checks:

MolProbity: Analyzes dihedral and rotamer angles through Ramachandran Plot Analysis to asses the realism of the folded structure.



Structural Quality Checks:

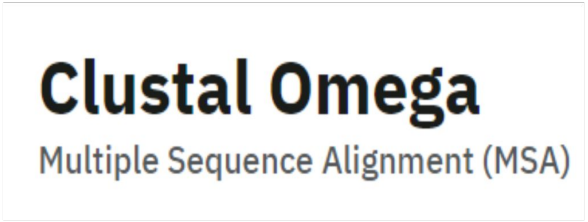
IGP-ERRAT: Verifies errors in crystallography data based on non bonded atom-atom interactions and comparison to higher resolution structure.



AF3 Predicted complex of HAP40/HTT. TM-score assesses the topological similarity of protein structures. Cryo-EM(not-shown) TM-score = 0.875, AF3 TM-score = 0.740 (Score >0.5 generally indicates correct topology)
Left: Human HTT/HAP40 complex. Right: *Drosophila* HTT/HAP40 complex.

Red = HTT's C-HEAT domain
Yellow = HTT's Bridge domain
Blue = HTT's N-HEAT domain
Green = HAP40

Comparing predicted complexes

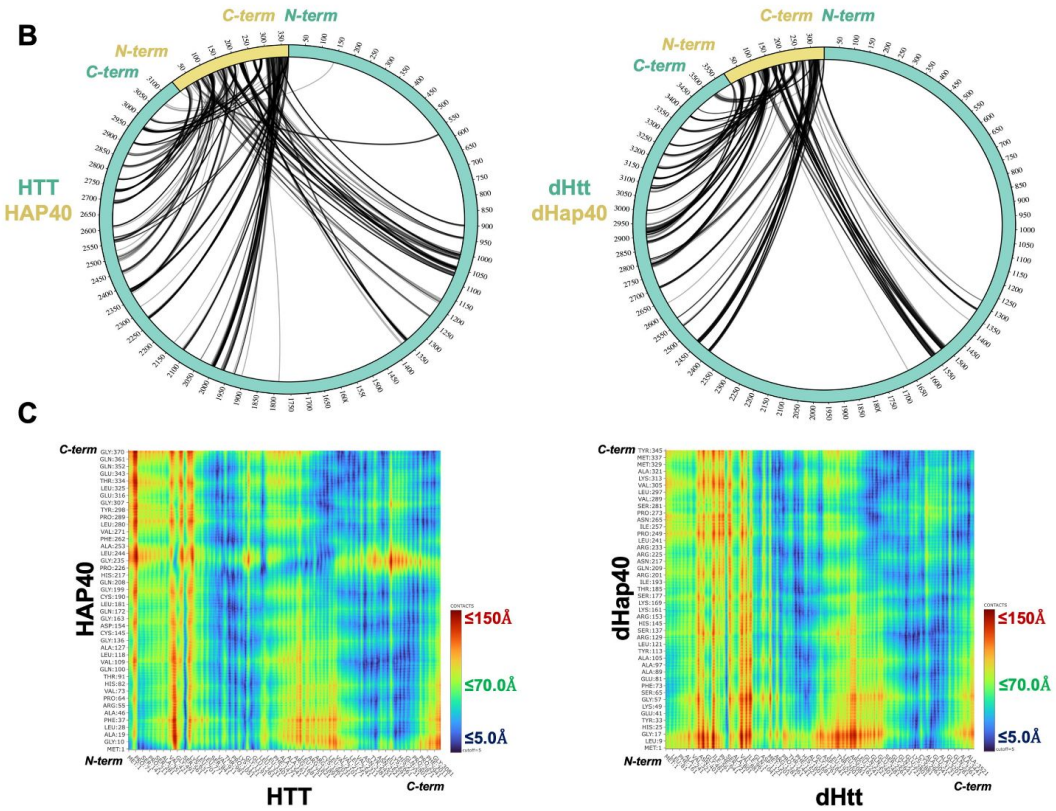


Compare Sequences:
Allows the comparison of protein sequences between species to uncover orthologous domains

Compare folded structures:
Allows the comparison of protein structures to uncover shared similar global topology using Template Modeling (TM) scores and Root Mean Square Deviation (RMSD) values.

Annotate structure interaction points:
Allows the comparison of protein sequences between species to uncover orthologous domains

Calculate binding affinities :
Predicts the binding affinities in Gibbs free energy (ΔG) and disassociation constant(K_d).



B: circular map shows the sequence of the 2 proteins in complex
Black lines = interacting residues (~5 Å)

C: Heat map representation of B.
Blue = interacting residues (~5 Å)

Found similar interacting residue patterns.
Made with MAPIYA

Table 1. Summary of protein-protein interfacial contacts and PRODIGY-predicted binding affinities.

Complex	*ΔG (kcal/mol)*	*Kd (M)*	*Charged-charged*	*Charged-polar*	*Charged-apolar*	*Polar-polar*	*Polar-apolar*	*Apolar-apolar*	*NIS- charged %*	*NIS-apolar %*
dHap40-dHtt	−34.1	9.8e-26	68	67	114	24	70	69	25.92	38.96
dHap40-dHtt^{NTD} (627-2236)	−10.1	1.6e-08	8	18	24	7	17	23	22.46	42.13
dHap40-dHtt^{CTD} (2516-3583)	−24.0	1.7e-18	44	39	67	16	52	40	26.07	40.9
HAP40-HTT	−37.1	5.7e-28	40	55	109	24	98	137	25.3	40.26
HAP40-HTT^{NTD} (407-1721)	−17.1	2.2e-13	10	13	34	10	45	49	23.76	42.88
HAP40-HTT^{CTD} (2095-3138)	−20.0	1.3e-15	21	33	60	12	46	68	23.92	45.13

Summary on interacting surfaces(IS) and solvent-exposed non-interacting surfaces (NIS):

Made with PRODIGY

Which HTT domain does HAP40 primarily interact with?

PRODIGY analysis shows HAP40 has a bias to bind to the C-terminal domain

HAP40 interacts with:	HTT/HAP40	dHTT/dHAP40
HTT's C-terminal domain	51.8%	62.6%
HTT's N-terminal domain	34.9%	23.5%
Other regions	13.3%	13.9%

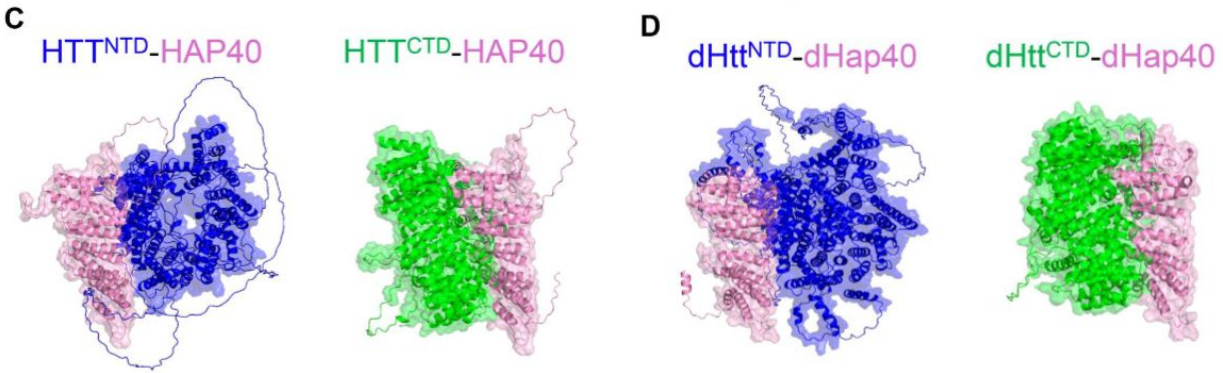
Why does the N-terminal and C-terminal interaction bias occur?

PRODIGY analysis of pseudo-complexes reveals C-terminal domain dominates complex stabilization

Table 1. Summary of protein-protein interfacial contacts and PRODIGY-predicted binding affinities.

Complex	*ΔG (kcal/mol)*	*Kd (M)*	*Charged-charged*	*Charged-polar*	*Charged-apolar*	*Polar-polar*	*Polar-apolar*	*Apolar-apolar*	*NIS- charged %*	*NIS-apolar %*
dHap40-dHtt	−34.1	9.8e-26	68	67	114	24	70	69	25.92	38.96
dHap40-dHtt^{NTD} (627-2236)	−10.1	1.6e-08	8	18	24	7	17	23	22.46	42.13
dHap40-dHtt^{CTD} (2516-3583)	−24.0	1.7e-18	44	39	67	16	52	40	26.07	40.9
HAP40-HTT	−37.1	5.7e-28	40	55	109	24	98	137	25.3	40.26
HAP40-HTT^{NTD} (407-1721)	−17.1	2.2e-13	10	13	34	10	45	49	23.76	42.88
HAP40-HTT^{CTD} (2095-3138)	−20.0	1.3e-15	21	33	60	12	46	68	23.92	45.13

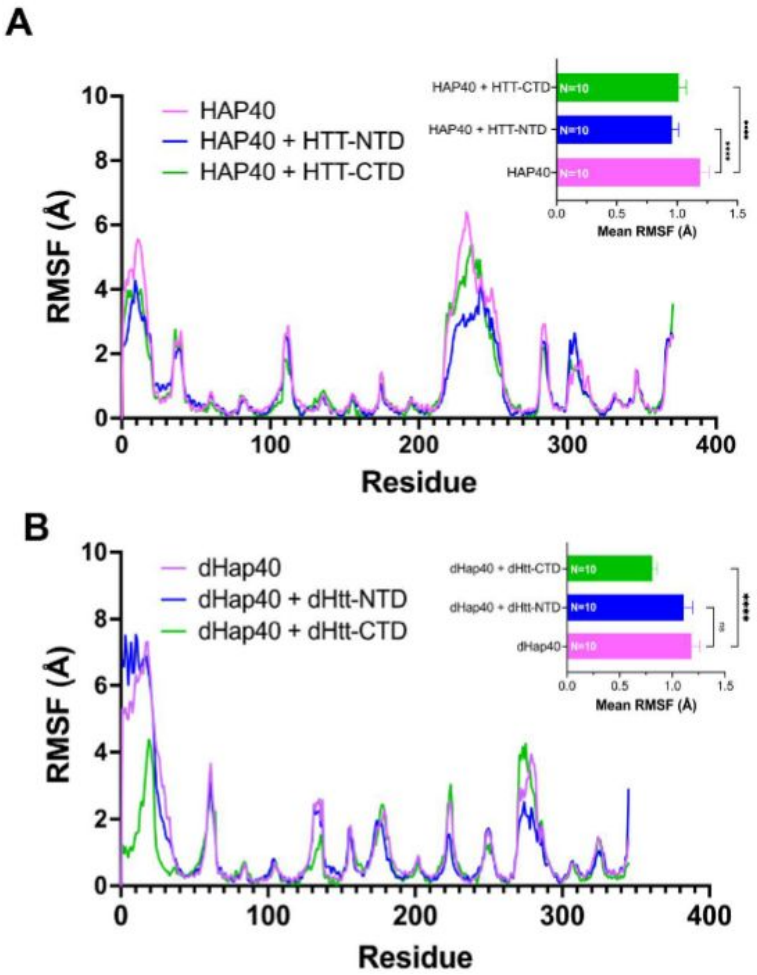
55% in human, 70% in flies.



Does the N-terminal and C-terminal interaction bias withstand instability?

CABSflex

Allows the measurement of a protein structures flexibility to measure dynamics in a value of root mean square fluctuation (RMSF). Simply put, this tool perturbs dihedral and torsional angles and calculates energetics to understand stability.

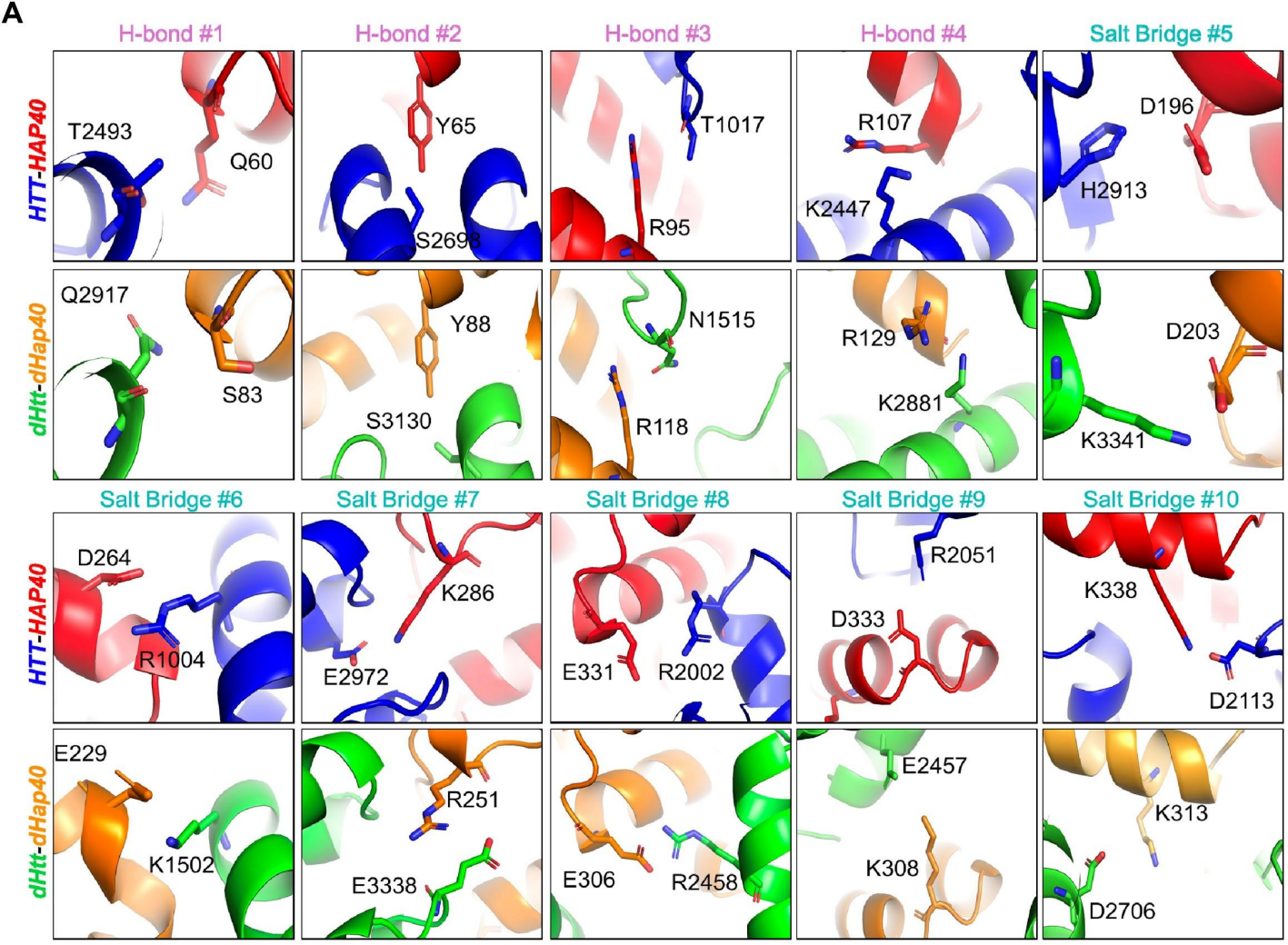


High RMSF at the interface suggests dynamic binding (e.g., weak or transient interactions).

If interface residues show low RMSF, the interaction is stable.

HTT-CTD & HTT-NTD stabilizes HAP40 in humans, but only HTT-CTD in flies does. Both implicate the dominant role of HTT-CTD.

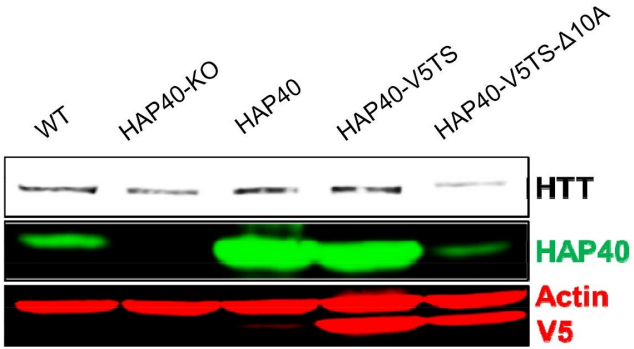
What specific conserved residues are at play?



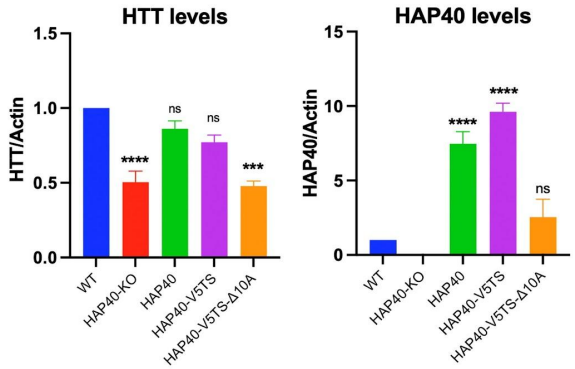
10 conserved interacting residues.4 H-bonding, 6 salt-bridges. In the core of the complex. According to PRODIGY, their deletion lowers their affinity between the complex.

In silico mutations using DDMut-PPI reveals a similar destabilizing effect as PRODIGY (Results not shown).

B



C

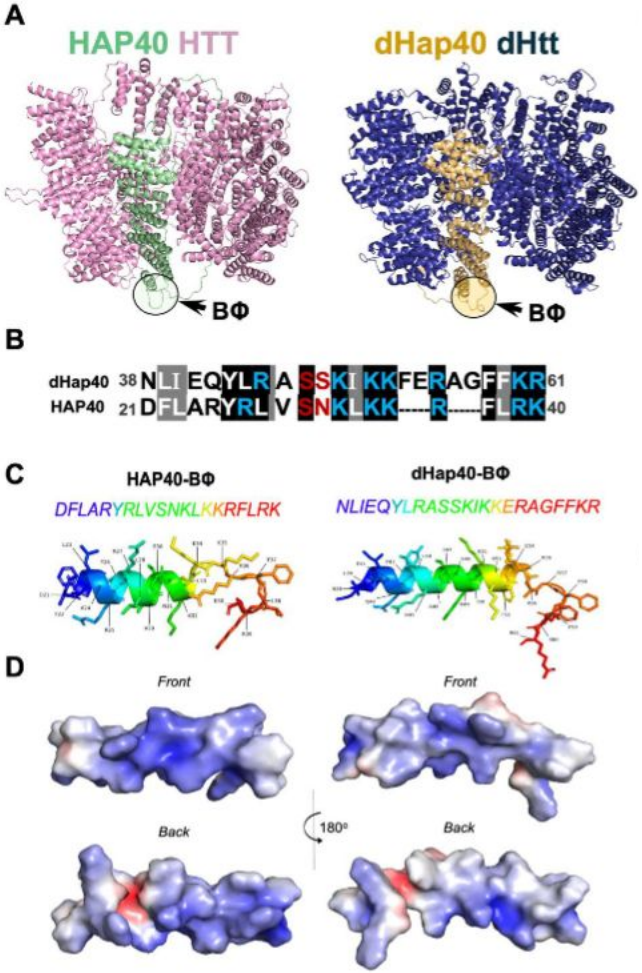


HAP40-V5TS = A HAP40 whose disordered region is replaced with a V5 an Twinstrep epitopes. Doesn't affect function.

HAP40-V5TS-delta10A = a HAP40 mutant whose 10 conserved interacting residues have been mutated with alanine as commonly used in scanning alanine mutagenesis

Transfected into HEK293T cells

Is that it?



A,C &D) AF3 predicted structures

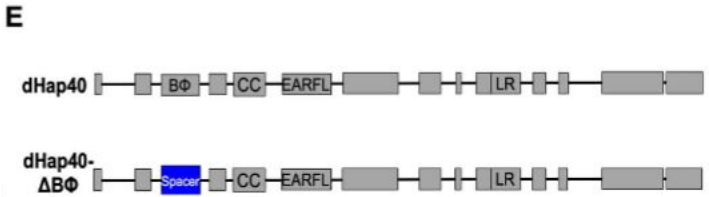
A = complex

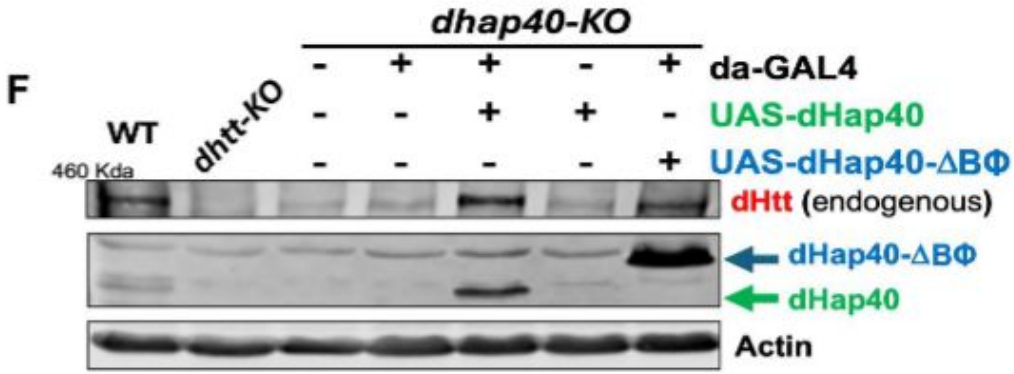
C = BΦ motif secondary structure

D = BΦ surface charges. White = hydrophobic, red = neg. charge, blue = pos. charge

B) Sequence alignment

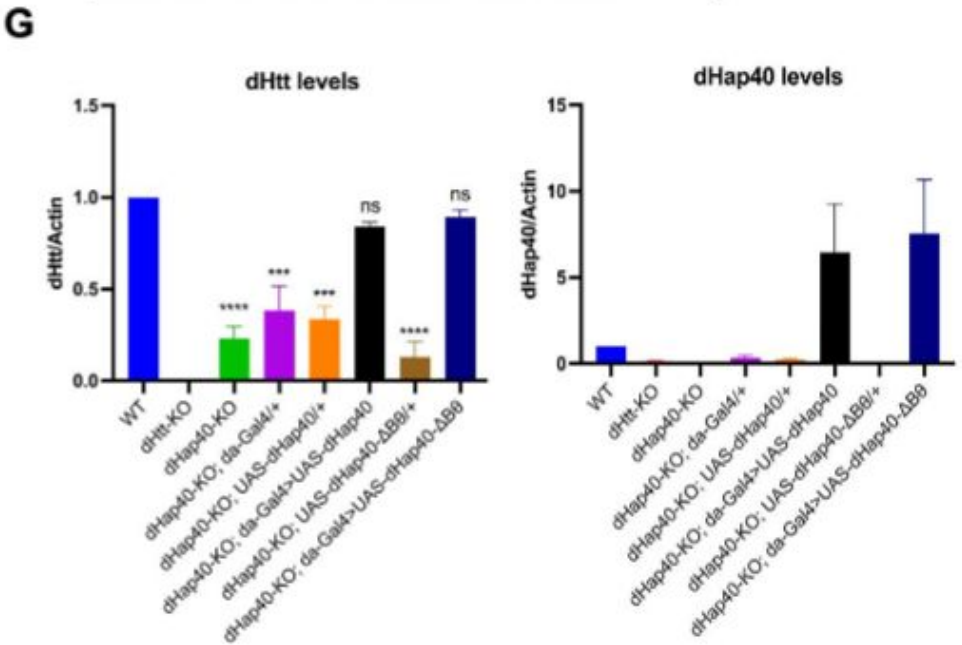
E) Domain alignment





F) Western blot on flies with HAP40 KO

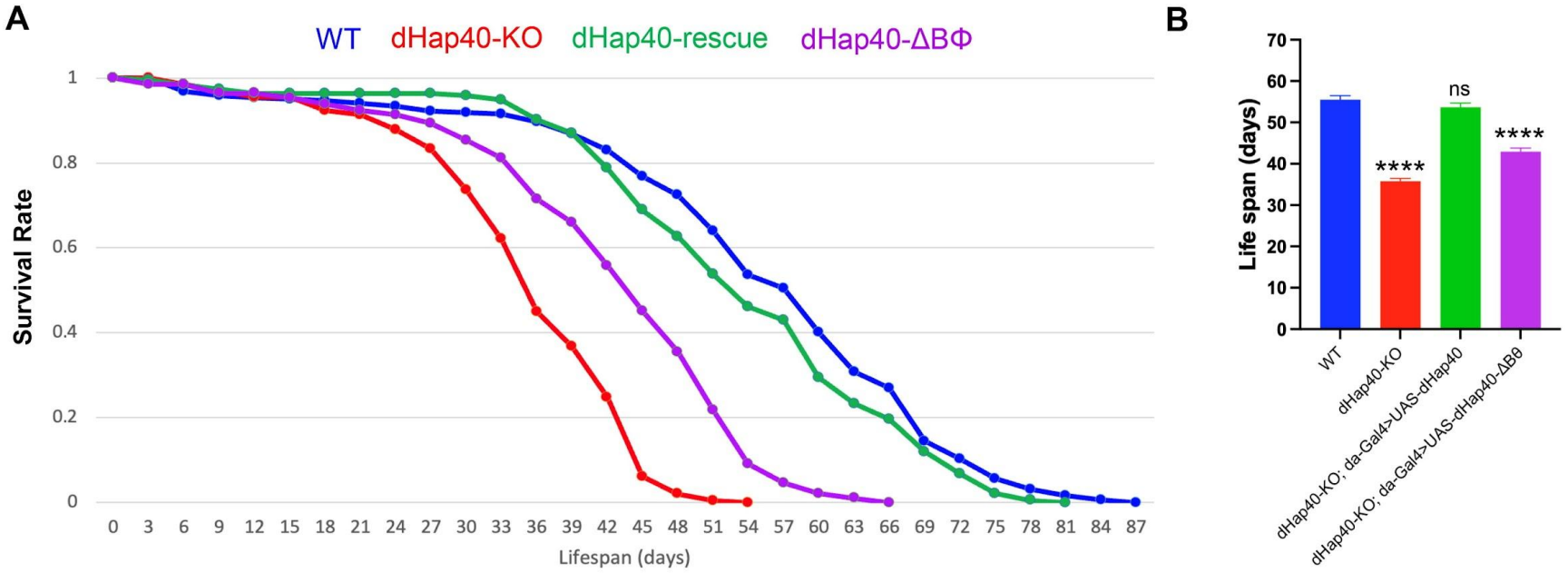
G) Graph representation of F

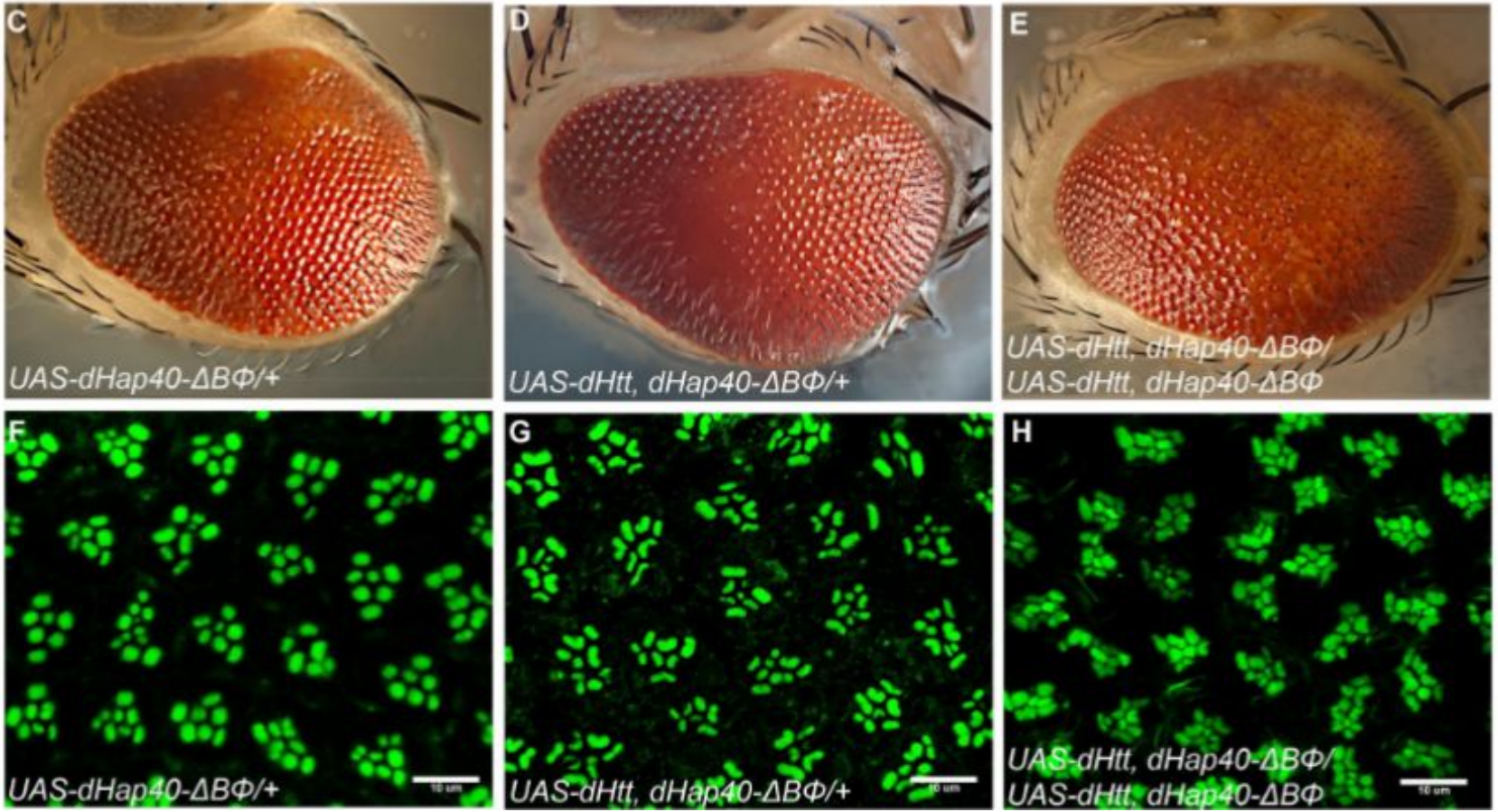


A BΦ motif is conserved in both HAP40's N-termini. It isn't important in the stability of the complex.

It is physiologically important though.

The physiological importance of BΦ motif in HAP40.


A
WT
dHap40-KO
dHap40-rescue
dHap40-ΔBΦ
Survival Rate
1
0.8
0.6
0.4
0.2
0
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60 63 66 69 72 75 78 81 84 87
Lifespan (days)
B
Life span (days)
70
60
50
40
30
20
10
0
ns
****
****
WT
dHap40-KO
dHap40-KO; da-Gal4>UAS-dHap40
dHap40-KO; da-Gal4>UAS-dHap40-ΔBθ


C
UAS-dHap40-ΔBΦ/+
D
UAS-dHtt, dHap40-ΔBΦ/+
E
UAS-dHtt, dHap40-ΔBΦ/
UAS-dHtt, dHap40-ΔBΦ
F
UAS-dHap40-ΔBΦ/+
G
UAS-dHtt, dHap40-ΔBΦ/+
H
UAS-dHtt, dHap40-ΔBΦ/
UAS-dHtt, dHap40-ΔBΦ

What could be done next?

HTT Facts:

Human aa length : 3142

Dros length: 3583

Sequence: 25.9% identity, 41.5% overall similarity

Secondary structure: HEAT repeats similarity using Cryo-EM, Multiple EM

40 in humans

38 in dros

HAP40 facts:

Human aa length: 371

Dros: 345

Sequence: 28.3% identity, 41.6% overall similarity

Secondary structure: TPR repeats