EXPERIMENTAL AND THEORETICAL STUDIES OF THE MITOCHONDRIAL REPLISOME

By

Gregory Alan Farnum

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Biochemistry and Molecular Biology - Master of Science

2013

ABSTRACT

EXPERIMENTAL AND THEORETICAL STUDIES OF THE MITOCHONDRIAL REPLISOME

By

Gregory Alan Farnum

Despite the ubiquitous trend of increasing complexity in eukaryotes, the vital process of mtDNA replication has retained a simplistic, three component replisome consisting of the mtDNA Pol γ , mtSSB and the mtDNA replicative helicase. This minimal mitochondrial replisome resembles the bacteriophage T7 replisome and may even be simpler due to the apparent absence of a primase function in the N-terminal domain of the mtDNA replicative helicase. Interestingly, insects have retained the four cysteines of motif I which have been shown to bind zinc in T7, so an N-terminal construct (NTD) of Drosophila melanogaster mtDNA replicative helicase was designed containing the conserved residues. Purified fractions of the NTD construct contain an iron sulfur cluster, as determined by UV-Visible spectroscopy, as well as iron and sulfide determination. This work describes for the first time an iron sulfur cluster site in dm-mtDNA replicative helicase and evidence of weak DNA binding by the NTD using a FRET assay. Mutations in Pol γ represent a major cause of human mitochondrial diseases, especially those affecting the nervous system in adults and in children. More than 160 POLG1 disease mutations have been identified, which are nearly uniformly distributed along the length of the POLG1 sequence. Comprehensive literature analysis of the biochemical properties and disease characteristics of Pol γ in light of the crystal structure have revealed genotype-phenotype correlations that support the clustering of mutations into five functional modules in the catalytic core of Pol γ . Our results suggest that cluster prediction can be used to evaluate both the likely biochemical defects and the relative pathogenicity of new POLG variants.

TABLE OF CONTENTS

LIST OF TABLES
LIST OF FIGURES vii
KEY TO SYMBOLS AND ABBREVIATIONS xiv
 CHAPTER 1: The N-terminal primase-like domain of the Drosophila replicative mitochondrial DNA helicase binds an iron sulfur cluster. INTRODUCTION RESULTS Design of N-terminal truncation variants guided by T7 gp4 homology Purification of the NTD Dm helicase construct. Dark brown fractions of NTD tend to aggregate without sufficient glycerol content and thiol reductant. Observation NTD forms SDS and DTT resistant dimers and higher multimers in SDS
 Observation-NTD forms SDS and DTT resistant dimers and nigher multimers in SDS-PAGE gels
 Optical features of as isolated NTD helicase nucleons inclusion to ZBD spectra
 Chemical analysis of colored fractions by Graphite Furnace Atomic Absorption Spectroscopy
 Surfice determination by methylene blue colorimetric method
 Proposal and design for mutagenesis of the NTD construct

	METHODS	33
	Bacterial cell growth and protein overexpression	33
	Cell lysis and preparation of soluble fraction (FrI)	33
	Ni-NTA metal affinity chromatography (FrII)	33
	Glycerol gradient sedimentation (FrIII)	34
	• Graphite furnace atomic absorption analysis of iron content of NTD (or ZBD) put	rified
	fractions	34
	• Quenching of fluorescence of fluorescein-labeled ssDNA (ssDNA-fl) by NTI) via
	FKE1	35
	 Calculate dissociation constant Kd for fl-dna binding by NTD 	36
	Methylene Blue Photometric Analysis of Sulfide Content in Protein Samples	36
	Inhibition of FRET quenching by salt	37
	Competitive FRET binding assay for polynucleic substrates	37
	• Calculate the inhibition coefficient K1 for each substrate	38
	• EPR Spectroscopy	39
	UV-Visible Spectroscopy	39
	Methods for mutagenesis	39
	• Growth and preparation of Ca ²⁺ competent cells	39
	\circ Transformation of Ca ²⁺ competent cells	40
	• Preparation of electrocompetent bacteria (XL-1 Blue)	40
	• Mini-preparation of plasmid DNA: pET-28a- Dm mt helicase N24-A333	from
	BL21(DE3) cells	41
	• Transformation of pET-28a-dm mt helicase N24-A333 plasmid into XL-1	Blue
	cells via electroporation	41
	• Mini-preparation of plasmid DNA: pET-28a- Dm mt helicase N24-A333	trom
	XL-1 Blue cells	42
\sim	• Quick change mutagenesis of <i>am</i> miDNA helicase N24-A555	42
	RIBLIOGRAPHV	45
		51
CF	IAPTER 2: Clustering of Alpers disease mutations and catalytic defects in biochemical	
var	riants reveal new features of molecular mechanism of the human mitochondrial replicase l	Pol
γ	1	55
≻	ABSTRACT	56
\succ	INTRODUCTION	57
	Approach to comparative structural analysis	59
	 Clustering of Alpers disease mutations within the catalytic core of Pol γ 	60
\triangleright	RESULTS	64
	• ALPERS CLUSTER 1: Residues affecting 5'-3' DNA polymerase activity	64
	• Catalytic residues	64
	• Architectural residues	67
	• Residues conterring a high attinity of the pol site for the primer-tem	iplate
	JUNCTION	08 10
	• CLOSTER 2. Recessive inutations affecting the upstream DNA-officing channel	12

	 CLUSTER 3: Mutations associated with a novel, Pol γ-specific functional module, conferring partitioning of DNA substrate between the pol and exo active sites
	APPFNDIX 91
>	BIBLIOGRAPHY 96
CH Po	APTER 3: Mapping 136 Pathogenic Mutations into Functional Modules in Human DNA ymerase γ Establishes Predictive Genotype-phenotype Correlations for the Complete
Sp	ectrum of POLG Syndromes
≻	ABSTRACT
\triangleright	INTRODUCTION
\triangleright	METHODS
	Computational analysis
	• Statistical analysis
\triangleright	RESULTS 107
	Cluster assignment of all reported POLG disease mutations
	• Cluster assignment can be used to predict the pathogenicity of a novel mutation 116
	• Cluster combinations of compound heterozygotes correlate with age of onset of POLG
	syndrome
	• Critical functions: clusters 1, 3, and 4
	• Moderately severe dysfunction: clusters 2 and 5
	• Severe combination: cluster 1 + 2
	• Severe combination: cluster 1+5
	• Other combinations
	• Expected incidence of specific symptoms caused by POLG syndrome can be predicted
~	by the age of onset
>	ADDENIDIX 124
>	AFFENDIA
\succ	

LIST OF TABLES

Table 1. A comparison of UV-visible electronic absorption spectra of 2Fe-2S and 4Fe-4Sclusters from published studies. The UV-visible spectra collected for the NTD construct (onpage 29) show a peak at 280 nm corresponding to aromatic amino acids in the protein, a shoulderat about 333 nm, and a peak at about 425 nm.17

Table 2. Calculated ratios of Fe: protein for the helicase and cyt c samples from graphite furnace atomic absorption data. Sample concentration was determined by the absorbance at 280 nm for purified helicase variants and by Bradford for the cyt *c* sample. All samples were measured in triplicate. Error bars were calculated by propagated error analysis and all measurements, dilutions, etc were assigned an uncertainty that was appropriate for the measurement, instrument, pipette, etc. (ex: my P20 pipette was calibrated and found to have a standard deviation of 0.6 μ L, so any measurement or transfer with this pipette, was given an uncertainty equivalent to the standard deviation (ex: 9.0 ± 0.6 μ L).

Table 3. Physical properties based on primary sequence of NTD helicase. Theoretical physical properties as determined from amino acid sequence of *Dm* mtDNA helicase N24-A333.

Table 4. Protein content and purity of purification steps. Total protein concentration wasdetermined by A280 analysis and by the colorimetric Bradford method throughout purification.Purity was estimated by SDS-PAGE analysis of target band versus contaminant bands.46

45

Table 5. Structural and functional features of the five proposed Alpers Clusters.	84
Table 6. Grouping of POLG syndrome symptoms	129

LIST OF FIGURES

Figure 1. Human minimal mitochondrial replisome. Orange arrow with dashed lines illustrates leading strand synthesis by human DNA polymerase γ . Three different colors represent domains of the mtDNA helicase that are discussed throughout the thesis: Purple, C-terminal helicase domain; Blue, RNA polymerase domain (RPD); Brown, Zinc binding domain (ZBD); Blue + Brown, equivalent to the N-terminal domain (NTD) construct. POLGA, human DNA polymerase γ catalytic subunit; POLGB, human DNA polymerase γ accessory subunit; mtSSB, human mitochondrial single-stranded DNA-binding protein; Twinkle, human mitochondrial replicative DNA helicase. "For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis."

Figure 2. Schematic representation of *Dm* helicase domain motifs NTD (36 kDa) and ZBD (12 kDa) constructs, and multiple sequence alignment of the zinc binding Motif I. Arrows indicate positions of four conserved cysteine residues that ligate zinc in T7 primase. 5

Figure 3. The NTD has a modular architecture consisting of a ZBD and RPD, similar to T7 gp4. SDS-PAGE (right) and anti-his immunoblot (left) analysis of 36 kDa N24-A333 construct after limited trypsin digestion. 9

Figure 4. UV-visible electronic absorption spectrum of NTD. Peaks at 422 nm, 464 nm, and 590 nm suggest the presence of an iron-sulfur cluster by comparison with published examples of UV-visible electronic absorption spectra of protein bound 2Fe-2S and 4Fe-4S clusters. 11

Figure 5 Comparison of UV-visible spectrum of NTD vs ZBD purified fractions. Pink trace represents NTD spectrum, both red and green are ZBD fractions after Ni-NTA fractionation. 11

Figure 6. UV-Visible spectrum of the NTD construct after treatment with a strong reducing agent. An immediate bleaching effect of dithionite on as isolated NTD Dm helicase sample (blue) after addition of 8 fold excess dithionite (red). Change in volume was negligible (2 µL in 500 µL) compared to the observed change in absorption at 420 nm.

Figure 7. Effect of DTT on the NTD *Dm* helicase sample after 36 hours on ice (red). As isolated NTD (blue) is less intense in 300-325nm range. Concentration of as isolated NTD *Dm* helicase is 28.5μ M. The oxidation of DTT to the cyclic configuration is accompanied by increased absorption below 270-340 nm. 14

Figure 8. UV-Visible analysis of *in vitro* **reconstitution of iron sulfur clusters.** The bottom most (pink) trace represents 1:1 ferrous iron and sulfide: helicase immediately after addition of NTD helicase (28.5 μ M). The second trace (green) is after 1 hour, the third trace (red) is after addition of 2:1 ferrous iron and sulfide: helicase, and the top trace (blue) is after 12 hours. 15

Figure 9. The Fe-S cluster is resistant to oxidizing conditions. Normalized absorbance plot for hydrogen peroxide titration of NTD *Dm* helicase. Blue: 7 μM as isolated NTD *Dm* helicase, Red: 10:1 H₂O₂: helicase, Green: 100:1 H₂O₂: helicase.

Figure 10. Low field EPR spectrum of as isolated NTD Dm helicase. Performed by a student prior to my time in the lab, sample contains EDTA. High field signals near g = 2 are obscured by a known copper contaminant in the instrument (see appendix on page 63). 22

Figure 11. Low field EPR spectrum of NTD Dm helicase after the addition of dithionite. Performed by a student prior to my time in the lab, sample contains EDTA. High field signals near g = 2 are obscured (see appendix on page 63). 22

Figure 12. Low field EPR spectrum of NTD Dm helicase after the addition of ferricyanide. Performed by a student prior to my time in the lab, sample contains EDTA. High field signals near g = 2 are obscured (see appendix on page 63). 23

Figure 13. Quenching of fluorescence in the presence of NTD construct. Excitation of Fluorescein (red)-labeled ssDNA (ssDNA-fl) in the absence of NTD exhibited normal fluorescence and addition of purified NTD construct (blue) resulted in quenching of fluorescence. Iron sulfur clusters (brown star) can act as acceptors in FRET and result in quenching.

Figure 14. Increasing ionic strength inhibits the quenching observed by the FRET assay. Salt ions can compete with the ionic interactions at the interface of DNA – protein complexes and increase the rate of dissociation and inhibit re-binding. Addition of salt (NaCl and

24

MgCl₂).to a solution containing 10 nM fl-ssDNA/ 5 uM NTD Dm helicase caused an immediate increase in fluorescent intensity. By inhibiting the binding of fl-ssDNA to the NTD the quenching of fluorescence by FRET is prevented because the Fe-S cluster is no longer in close proximity to the fluorescein. Complete inhibition of FRET quenching is observed at 150 mM 25

NaCl and 70 mM MgCl₂.

Figure 15. FRET assay shows that quenching can be inhibited competitively by unlabeled ssDNA. Semi-log plots of the FRET competitive binding assay comparing the polydT substrate with ssRNA substrate, polyrA. Concentrations of DNA and RNA substrates are by nucleotide concentration because the length of individual molecules are not constant. 26

Figure 16. Alignment of fly, frog and human mtDNA helicases for mutagenesis design and selection of conserved cysteine residues. Cysteines highlighted in green will be produced as single alanine substitutions in the NTD construct and cysteines highlighted in cyan will be produced as double alanine substitutions for each pair of CXXC motifs. Red blocks (α -helical) and blue arrows (β -sheet) below the sequence illustrates the predicted secondary structure. 28

Figure 17. SDS-PAGE analysis of NTD construct throughout the purification. We suspect an SDS-resistant disulfide-linked dimer is the band at 70 kDa. 46

Figure 18. Gel filtration analysis of NTD construct using a Superdex 75 column. The elution of a single peak corresponding to a monomer as determined by the A_{280} measurements shown as the blue trace. 47

Figure 19. Instrument interference at g = 2 prevented further characterization of Fe-S characteristics by EPR. This EPR spectrum was taken with an empty sample tube to show that the large signal at g = 2 is suspected to be a copper impurity in the instrument and not due to sample contamination. 47

Figure 20. Example of semi-log plot and fit used to calculate the IC50. (The x axis is logarithmic).

Figure 21. Example UV-Visible spectrum for standards: The reaction is consideredsuccessful if the ratio of absorbance between A670 and A750 is close to 2 as is observed in thesespectra. Blue- .5 nmol sulfide, Red- 1 nmol sulfide, Pink- 2 nmol sulfide, Dark green- 3 nmolsulfide, Brown- 4 nmol sulfide.49

48

Figure 22. Alpers disease mutations cluster within functional modules in the catalytic subunit of Pol γ . Upper panel: schematic diagram of the POLG1 gene showing the distribution of recessive Alpers disease mutations (Human DNA Polymerase gamma Database, http://tools.niehs.nih.gov/ polg/). AID and IP in the spacer domain refer to the accessory (subunit) interacting and intrinsic processivity subdomains, respectively, that are discussed in the text; NTD refers to the N-terminal domain. Lower panel: tertiary structural representation of the apoenzyme form of Pol γ [PDB code 3IKM, (7)] with modeled DNA, identifying the positions of five functional modules (shown in mesh) that are defined by clusters of amino acid residues (shown as spheres) affected by Alpers disease mutations as follows: Cluster 1, green; Cluster 2, yellow; Cluster 3, red; Cluster 4, blue; Cluster 5, cyan. The domains of Pol γ A are shown as surface representations, and in part as secondary structural elements (SSEs) that are colored as depicted in A. The proximal and distal accessory subunits are shown as surface representations in light and dark gray, respectively. Primer–template DNA was docked as described in the text and is displayed as orange ribbons.

Figure 23. Alpers Cluster 1 mutations affect the 5 DNA polymerase activity of Pol y. Amino acid residues affected by Alpers Cluster 1 mutations in POLG1 are shown as green spheres. Other Pol γ residues that are discussed in the text are shown in brown, and T7 residues are shown in red. Pol domain SSEs are shown in pink according to the schematic shown in Figure 22 on page 77; ptDNA is indicated by orange (template) and brown (primer) strands. The incoming dNTP is shown in blue. Mg^{2+} ions are shown as small gray spheres. A, upper panel, overview of the positions of Alpers Cluster 1 mutations with dashed black lines indicating the regions described in the text and depicted in B–D; A, lower panel, overview of the positions of Alpers Cluster 1 mutations relative to ptDNA and incoming dNTP; B–D, positioning of Pol γ residues in the apoenzyme form [PDB code 3IKM (7)] and T7 Pol residues in its closed ternary complex [PDB code 1T8E (25)] relative to ptDNA and incoming dNTP, with the dashed arrow showing the expected movements of the Pol γ residues upon formation of a closed complex: B, Alpers mutations surrounding the Mg^{2+} -binding residues; C, Alpers mutations affecting O-helix movement and dNTP binding, with the T7 Pol O-helix in its closed conformation superimposed in red; D, Alpers mutations affecting the RR loop and the surrounding residues. 70

Figure 24. Alpers Cluster 2 mutations affect the upstream DNA-binding channel of Pol γ . Amino acid residues affected by Alpers Cluster mutations in POLG1 are shown as yellow spheres. Other Pol residues that are discussed in the text are shown in brown. Spacer domain SSEs are shown in magenta and pol domain SSEs are in pink according to the schematic shown in Figure 22 on page 77; ptDNA is indicated by orange (template) and brown (primer) strands. The spacer domain is also shown as a transparent surface representation in pale gray and the exo domain is shown in purple. 74 Figure 25. Alpers Cluster 3 mutations are associated with a novel Pol γ -specific functional module proposed to be involved in primer strand partitioning between the pol and exo sites. Amino acid residues affected by Alpers Cluster 3 mutations in POLG1 are shown as red spheres/ mesh adjacent to a novel alpha helix with an associated loop–hairpin (the partitioning loop), also shown in red. The brown spheres and mesh represent the SYW (fly)/ SFW (human) and surrounding residues, respectively, that are described in the text. The pol domain is shown as a surface representation in pink and the exo domain is shown in purple, according to the schematic shown in Figure 22 on page 77; ptDNA is indicated by orange (template) and brown (primer) strands. (A) The predicted position of the partitioning loop relative to ptDNA in the pol mode, and (B) represents the exo mode. To dock the frayed ptDNA in the exo active site, the exo domain residues 324–518 of Klenow (PDB code 1KLN) were aligned with the exo domain residues 170–440 of Pol γ (PDB code 3IKM) (see appendix B for Figure 31 on page 105). 78

Figure 26. Alpers Cluster 4 mutations affect Pol γ A interactions with the distal Pol γ B upon DNA binding by Pol γ holoenzyme. Amino acid residues affected by Alpers Cluster 4 mutations in POLG1 are shown as blue spheres/ mesh and are located largely within the exo domain (shown in purple). The helix shown in magenta is the AID (accessory interacting domain) of the spacer region that interacts with the proximal accessory subunit. Other domains are colored according to the schematic shown in Figure 22 on page 77; duplex DNA is indicated by the orange strand. Orange spheres indicate the positions of the accessory subunit residues described in the text. The proximal and distal accessory subunits are shown as surface representations in light and dark gray, respectively.

Figure 27. Alpers Cluster 5 mutations are proposed to affect replisome interactions. Amino acid residues affected by Alpers Cluster 5 mutations in POLG1 are shown as cyan spheres. Spacer domain SSEs within the IP (intrinsic processivity subdomain) are shown in magenta; ptDNA is indicated by orange (template) and brown (primer) strands. The spacer domain is also shown as a transparent surface representation in pale gray and the exo domain is shown in purple. Brown spheres indicate the positions of the IP residues described in the text. 83

Figure 28. Combinations of mutations found in Alpers patients. Individual Alpers mutations are grouped by cluster as shown in Figure 22 on page 77, and black blocks represent Alpers manifesting combinations. The inset in the lower right presents a simplified version of the table by reducing the axes to the five clusters only, where gray blocks represent cluster combinations that have not been found in Alpers patients. The tabulated data suggest that two mutations from the same cluster do not typically manifest as early-onset Alpers disease. Furthermore, the data indicate that Cluster 4 (blue) mutations manifest as Alpers disease only when in combination with Cluster 2 (yellow) or Cluster 5 (cyan) mutations. These trends support the existence of unique functional relationships in Pol γ A that are inherent to each cluster. 88

Figure 29. Alignments used for docking DNA onto the *Hs* Pol γ holoenzyme structure (3IKM). Panels display the palm subdomain of the T7 Pol (1T8E) closed ternary complex (red) superimposed on the palm subdomain of Pol γ (pink). Top panel, alignment of the palm of T7 Pol (residues 409-487, 611-704) to the palm of Pol γ (815-910, 1095-1239); middle panel, alignment of T7 Pol residues 646-663 (green) to Pol γ residues 1127-1144; lower panel, alignment of T7 Pol residues 606-635 and 422-431 (green) to Pol γ residues 1093-1122 (Q-helix) and 846-855 (RR-loop). Alignments were performed in Pymol.

Figure 30. Comparative alignments of *Hs* Pol γ , T7 Pol and T7 RNA Pol illustrating the structural variations in the region between the P-helix and the Q-helix in the pol domain. Panels on the left display the complete pol domain in complex with DNA, and those at right show a close up view of the of the region between the Q-helix and P-helix. Top panel, the Pol γ (3IKM) pol domain is shown as pink cartoon and its partitioning loop in red, with its disordered region indicated as a dashed red line. Transparent surfaces of the exo (purple) and pol (pink) domains are also shown at left. DNA docked by alignment with T7 Pol (see top panel of Fig. 29 on page 103) is shown in orange; middle panel, the Pol domain of the T7 RNA Pol elongation complex (1H38) is shown as light blue cartoon and its specificity loop in red; lower panel, the T7 Pol (1T8E) pol domain is shown as blue cartoon, and transparent surfaces of its exo (purple) and pol (blue) domains are shown to highlight a similar architecture as in Pol γ . WXGG and WXAG are the amino acid sequence motifs in Pol γ and T7 Pol, respectively, which are discussed in the text.

Figure 31. Structural alignment of the exo domain of Klenow editing complex (PDB code 1KLN, residues 324-518, displayed as blue cartoon) with the exo domain of human Pol γ (PDB code 3IKM, residues 170-440, displayed as purple cartoon). This alignment was used to predict the editing complex of Pol γ by docking of the frayed primer template (primer strand in chocolate, template strand in orange) onto the apoholoenzyme structure, which we display in Figure 25B on page 91. 94

Figure 32. Clustering of 136 pathogenic mutations within five functional modules in the catalytic subunit of human Pol γ . Upper panel, schematic diagram of the human POLG1 gene illustrating the clustering of 136 pathogenic mutations into discrete blocks of amino acid residues, which we term subclusters. Mutant alleles and subclusters are colored according to the cluster to which they belong: cluster 1, green; cluster 2, yellow; cluster 3, red; cluster 4, blue; cluster 5, cyan (see the text for details). The palm (residues 815-910 and 1096-1239), fingers (residues 911-1048), and thumb (Th, residues 440-475 and 785-814) subdomains of the Pol domain are colored pink, and the partitioning loop (PL, residues 1049-1095) is red. The accessory (subunit) interacting domain (AID, residues 476-570) and the intrinsic processivity (IP, residues 571-784) subdomains of the spacer domain are colored in magenta. The N-terminal domain (NTD, residues 1-170) and the Exo domain (residues 171-439) are colored in purple. Lower panel, structural model of the human Pol y apo-holoenzyme (PDB code 3IKM) with docked primer template DNA shown as orange ribbon (see *Computational methods* for details). The catalytic subunit of Pol γ is shown as a cartoon representation of the secondary structural elements (SSEs), with regions defined by clusters illustrated as space-filled modules, colored according to the schematic. The proximal and distal accessory subunits are shown as surface representations in light and dark gray, respectively. 110

Figure 33. Architectural and functional subclusters of the Pol domain. Upper panel, schematic diagram of the *POLG1* gene as shown in Figure 1, with an additional upper section indicating the location of the three Exo motifs, labeled as I, II, and III, and the six Pol motifs, labeled as 1 (motif 1), 2 (motif 2), A (Pol A motif), B (Pol B motif), 6 (motif 6) and C (Pol C motif), which are conserved throughout family A polymerases (see text). Motifs and subclusters that are illustrated in the bottom panels are in shown bold. Bottom panels, SSEs and transparent surface representation of the palm and fingers subdomains of Pol γ (PDB code 3IKM) are shown (subdomains are colored in gray in the middle section of the schematic). Bottom-left panel, the five motifs of the Pol domain, which are colored in pink in the upper section of the schematic, are shown as pink SSEs and are labeled accordingly. Docked primer template DNA is shown as orange ribbon, Mg^{2+} ions are shown as orange spheres, and the incoming dNTP is shown as orange sticks (see Computational methods for details). Bottom-right panel, subclusters 1D, 1E and 1F encompass one or more of the five conserved motifs of the Pol domain and are colored in light green. Subclusters 1A, 1B, 1C and 1G, colored in dark green, are located further structurally from the pol active site, and are considered to play an architectural role. 112

Figure 34. SNPs reported in the *POLG1* **gene rarely map to clusters.** Of 87 reported SNPs in the dbSNP database, 75 (86%, black boxes) map outside the defined pathogenic clusters (shown as colored regions in the schematic). The remaining 12 SNPs (magenta boxes) that map within the clusters can be considered as non-deleterious changes within a high-risk region of the POLG tertiary structure. However, in our view, the possible pathogenicity of the SNPs reported in cluster 4 (dark blue) warrants careful experimental reevaluation. 118

Figure 35. Analysis of mutation combinations as cluster combinations reveals predictive genotype-phenotype correlations. The data used to compile the information in this figure was derived from the literature listed under supplemental references in appendix C. The number of mutation combinations manifesting POLG syndrome at each of the four age groups is shown in each panel for specific cluster combinations. Age of onset trends can be used to predict the severity of POLG syndrome for an individual with compound heterozygous mutations in *POLG1. Upper panel,* POLG syndrome age of onset trends for mutation combinations of two mutations from different clusters versus the same cluster. *Middle panel,* patient data show trends in which severe cluster combinations have an earlier age of onset. *Lower panel,* earlier age of onset trends for more severe cluster/ subcluster combinations. 120

Figure 36. Analysis of all compound heterozygous mutation combinations as cluster combinations. The data used to compile the information in this figure was derived from the literature listed under supplemental references in appendix C. The number of mutation combinations manifesting POLG syndrome at each of the four age groups is shown for all cluster combinations for which patient data have been reported; no patient data have been reported for the three cluster combinations 3+4, 4+4 and 4+5. Age of onset trends can be used to predict the severity of POLG syndrome for an individual with compound heterozygous mutations in *POLG1*. Patient data show trends in which severe cluster combinations have an earlier age of onset, whereas less severe or uncommon cluster combinations have a later age of onset. 122

Figure 37. Age of onset correlation for reported symptoms. Symptoms associated with POLG syndromes manifest typically at different ages. The symptoms reported in the patients have been grouped together based on affected tissue types, and ordered according to their ages of onset. The more severe symptoms and forms of POLG syndromes are at the left end of the figure, while the less severe symptoms are at the right end. Detailed information of the symptom grouping is shown in Table 6.

Figure 38. Symptoms associated with different POLG syndromes. Depending on the specific mutations and cluster combinations of the *POLG1* gene mutations in individual patients, POLG syndromes can become symptomatic at different ages and manifest as pathogenic conditions in different tissue types. The figure shows a continuing spectrum of decreasing symptom severity from top to bottom, as well as a delayed age of onset from left to right. MCHS, Childhood myocerebrohepatopathy spectrum; MEMSA, Myoclonic epilepsy myopathy sensory ataxia; ANS, Autonomic Nervous System Dysfunction; arPEO/ adPEO, Autosomal Recessive/ Dominant Progressive External Ophthalmoplegia. 128

KEY TO SYMBOLS AND ABBREVIATIONS

2Fe-2S	2 iron 2 sulfur cluster
3Fe-4S	3 iron 4 sulfur cluster
4Fe-4S	4 iron 4 sulfur cluster
6x His	6x histidine tag
Å	angstrom
aa	amino acid
Ab	antibody
adPEO	autosomal dominant progressive external ophthalmoplegia
anti-his	anti 6x histidine tag antibody
ATP	adenosine triphosphate
$cm^{-1} M^{-1}$	extinction coefficient
cyt c	horse heart cytochrome c
Dm	Drosophila melanogaster
Dm helicase	Drosophila melanogaster mitochondrial replicative DNA helicase
DMPD	N, N-dimethyl-p-phenylenediamine
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylene diamine tetra acetic acid

EPR	electron paramagnetic resonance	
Exo	$3' \rightarrow 5'$ exonuclease	
FA	fluorescence anisotropy	
Fe-S	iron sulfur cluster	
fl-ssDNA	fluorescein-labeled ssDNA	
FRET	Forster resonance energy transfer	
GFAA	Graphite furnace atomic absorption	
GHz	gigahertz	
H ₂ O ₂	hydrogen peroxide	
НОМО	highest occupied molecular orbital	
HsmtDNA helicase	Human mtDNA helicase	
Hspol γ	Human DNA polymerase γ	
kDa	kilodalton	
ki	coefficient of inhibition	
Klenow	bacterial DNA polymerase I large fragment	
Klentaq	Thermus aquaticus Klenow	
LUMO	lowest unoccupied molecular orbital	
mg/ ml	milligram per milliliter	
m _s	angular spin momentum	
mtDNA	mitochondrial DNA	
mtDNA helicase	mitochondrial replicative DNA helicase	
mtSSB	mitochondrial single-stranded DNA-binding protein	

Ni-NTA	nickel nitriloacetic acid
nm	nanometers
NTP	nucleoside triphosphate
NTPase	nucleoside triphosphatase
OXPHOS	oxidative phosphorylation system
Pol I	bacterial DNA polymerase I
Pol y	DNA polymerase gamma
Pol γA	DNA polymerase gamma catalytic subunit
Pol yB	DNA polymerase gamma accessory subunit
Pol	$5' \rightarrow 3'$ DNA polymerase
POLG1	Pol yA gene
ppb	parts per billion
ptDNA	primer-template DNA
RNA	ribonucleic acid
RNAP	RNA polymerase basic domain
ROS	reactive oxygen species
RPD	RNA polymerase domain
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSB	single-stranded DNA-binding protein
ssDNA	single-stranded DNA
ssDNA-fl	fluorescein labeled ssDNA oligo used in FRET assays
SSE	secondary structural element

T7 gp4	bacteriophage T7 gene 4 protein
T7 gp5	bacteriophage T7 DNA polymerase
T7 Pol	bacteriophage T7 DNA polymerase
TOPRIM	topoisomerase-primase
UV-visible	Ultraviolet-visible (electronic absorption)
ZBD	zinc binding domain
μM	micromolar

CHAPTER 1: The N-terminal primase-like domain of the Drosophila replicative mitochondrial DNA helicase binds an iron sulfur cluster.

INTRODUCTION

The mitochondrial replicative DNA helicase (mtDNA helicase), also known as Twinkle in the human system, is an essential component of the mitochondrial replisome (1). The mtDNA helicase catalyzes the NTP-dependent unwinding of DNA at the mitochondrial replication fork in the 5'-3' direction. *In vitro* studies have established a minimal mitochondrial replisome consisting of the mtDNA helicase and three additional purified protein components (Figure 1 on page 4): the DNA polymerase gamma catalytic subunit (Pol γ A), the DNA polymerase gamma accessory subunit (Pol γ B) and the mitochondrial single-stranded DNA-binding protein (mtSSB). Together, these four components can generate DNA products greater than 16 kb in length, which is similar in size to the human mitochondrial genome (1). Mutations affecting the PEO1 gene that encodes the human mtDNA helicase have been reported in patients with adult-onset autosomal dominant progressive external ophthalmoplegia (adPEO), as well as compound heterozygous patients with more severe mitochondriopathies such as infantile-onset spinocerebellar ataxia and hepatocerebral mtDNA depletion syndromes (1, 2).

The mtDNA helicase was identified originally for its sequence similarity to bacteriophage T7 gp4 primase-helicase (T7 gp4). Biochemical investigation of human the mtDNA helicase revealed a modular architecture that is similar to the T7 gp4 primase-helicase, as they both contain distinct N and C-terminal domains separated by a flexible linker domain (1). In both proteins the C-terminal domain harbors the active site responsible for NTPase activity, and can unwind duplex DNA substrates independently. The C-terminal domain is considered to be in the DnaB-like family of helicases, which form ring shaped hexamers that encircle the DNA substrate (3). Two motifs that are well known are the Walker A and Walker B motifs that are involved in nucleotide binding, and in all hexameric helicases these motifs are located at the subunit interface (3). Binding of ATP allows interaction with an arginine finger residue of an adjacent monomer and

causes conformational shifts in four of six subunits. This asymmetric movement provides the 5'-3' directional translocation along the ssDNA due to loops that extend into the center of the ring from each monomer. ATP hydrolysis allows propagation of these movements for continuous motion. Conserved motifs in the DnaB-like family are present in all known mtDNA helicase amino acid sequences and have been reviewed previously (4). Overall, the C-terminal domain has been well characterized and the focus of this thesis chapter is on the function of the N-terminal domain.

The N-terminal domain of T7 primase-helicase consists of two functional domains; a RNA polymerase domain (RPD) that synthesizes RNA primers to be extended by a DNA polymerase, and a zinc binding domain (ZBD) that is involved in priming site recognition and stabilization of the nascent primer-template complex to ensure efficient delivery to the replicative T7 DNA polymerase (2). Whereas the N-terminal domain of *Hs* mtDNA helicase shows significant sequence similarity to the N-terminal domain of T7 primase-helicase, no primase activity has been reported *in vitro*. Therefore, the role of the N-terminal domain remains unclear, but is of interest due to the evolutionary conservation of N-terminal sequence motifs in a wide range of eukaryotic organisms from algae to fly to humans (4). Most studies on the mtDNA replicative helicase have focused on the human enzyme, and biochemical studies have been limited due to solubility issues during purification (5). Our lab has faced the same solubility issues with human mtDNA helicase purifications and we have attempted to characterize the *Drosophila melanogaster* mtDNA helicase (*Dm* Helicase) to possibly avoid similar solubility issues. This first thesis chapter is a report of my initial attempts to study the mtDNA helicase from the *Drosophila melanogaster* system.



Figure 1. Human minimal mitochondrial replisome. Orange arrow with dashed lines illustrates leading strand synthesis by human DNA polymerase γ . Three different colors represent domains of the mtDNA helicase that are discussed throughout the thesis: Purple, C-terminal helicase domain; Blue, RNA polymerase domain (RPD); Brown, Zinc binding domain (ZBD); Blue + Brown, equivalent to the N-terminal domain (NTD) construct. POLGA, human DNA polymerase γ catalytic subunit; POLGB, human DNA polymerase γ accessory subunit; mtSSB, human mitochondrial single-stranded DNA-binding protein; Twinkle, human mitochondrial replicative DNA helicase. <u>"For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis."</u>

RESULTS

Design of N-terminal truncation variants guided by T7 gp4 homology

The N-terminal region of metazoan mtDNA helicase shares extensive sequence conservation with the primase domain of bacteriophage T7 primase-helicase (4). Although several Mg^{2+} binding residues of the primase active site in T7 primase are not conserved in metazoans, the conserved primase sequence motifs are well conserved (6). Interestingly, insects have retained the four cysteines of motif I that have been shown to bind zinc in T7 gp4 (Figure 2 below). Therefore, to study the function of the N-terminal domain of metazoan mtDNA helicase, an Nterminal construct of Drosophila melanogaster mtDNA helicase (Dm helicase) was designed to comprise residues N24-A333 NTD. Based on T7 gp4 homology, this region contains the conserved zinc binding domain (ZBD) and RNA-polymerase domain (RPD). In addition to the NTD construct, a ZBD construct of Dm helicase that comprises residues N24-P123 was also produced. An N-terminal 6x histidine tag was added to both constructs. ZBD LinkerHelicase ---RPD-----H1a Ι Π V VI H1 H2 H3 H4 IIIR IV *Dm* N24-A333 (*Dm*-NTD) *Dm* N24-P123 (*Dm*-ZBD) Ec 34 KNFHACCPFHNE-----KTPSFTVNGEKOFYHCFGCG FLYHIPCDNCGSSD-----GNSLF---SDGHTFCYVCE Т7 11 Hs 75 SCLRALSPFAESSOLK---GOTGVTTSFSLFIDKTTGHFL**C**MTSL X1 92 SCLHVPSPFTPEO-----DTCLFIDKTTGRFL**C**KATL Dm 61 TCLOLECRLCDRNROPVTNAOKGTDHGLLAYVNKRTGAFICPNCD ↑ ↑ ↑ ↑

Figure 2. Schematic representation of *Dm* helicase domain motifs NTD (36 kDa) and ZBD (12 kDa) constructs, and multiple sequence alignment of the zinc binding Motif I. Arrows indicate positions of four conserved cysteine residues that ligate zinc in T7 primase.

Purification of the NTD Dm helicase construct

E. coli cells containing IPTG-inducible plasmid with cDNA encoding *Dm* helicase ZBD and NTD constructs were grown, induced, and harvested via centrifugation to produce pellets that were frozen in liquid nitrogen and stored at -80° C until purification. Previous lab members had completed several purifications of the NTD mutant derived from bacterial overexpression, and they were able to obtain high yields only when the freeze-thaw lysis was performed in the presence of DoDM instead of sodium cholate. Despite their initial success, they reported several failed purifications due to spontaneous aggregation in performing chromatography steps at 4° C in the presence of protease inhibitors, just as was found for the full-length human mtDNA helicase does under similar conditions. Furthermore, the chromatographic columns that were used after the Ni-NTA fractionation step were unable to remove the apparent contaminants observed as larger molecular weight bands on SDS-PAGE gels. Additionally, the protein eluted as a yellow solution and would become progressively darker upon concentration. So the study of the NTD construct was set aside and became my project about a year later. I set out first to document the origin of the brown color.

Dark brown fractions of NTD tend to aggregate without sufficient glycerol content and thiol reductant.

In all subsequent purifications we avoided the use of EDTA and other chelating agents in purification buffers to avoid disrupting the native iron-sulfur cluster. Fresh β -mercaptoethanol was replenished each day in all NTD-containing solutions or the NTD was observed to aggregate irreversibly. These two small changes to the purification procedure improved noticeably the stability of the NTD throughout purification, preventing aggregation of samples stored on ice overnight. The prevention of aggregation showed the most severe dependence on a high glycerol content of 10-15% in all solutions; we observed on several occasions the signs of aggregation after only 15 min upon dilution.

Observation-NTD forms SDS and DTT resistant dimers and higher multimers in SDS-PAGE gels.

As a result of the stabilization of NTD samples, purification of the NTD was achieved in only two chromatographic steps (See appendix A on page 46). SDS-PAGE gels showed a nearhomogenous NTD species, but a 70 kDa species that had inconsistent intensity from gel to gel was still present and needed to be explained (see appendix A on page 46). The last step of the purification is a velocity sedimentation in a glycerol gradient, which can be used to estimate the apparent molecular weight in addition to separating protein contaminants. The NTD construct sediments at a position consistent with a monomer of 35 kDa. Consistent with the sedimentation results, the NTD eluted as a sharp peak at a position corresponding to 35 kDa on a Superdex 75 gel filtration column (See appendix on page 47). Samples were golden-brown and became very dark when the protein was concentrated. Analysis of the Ni-NTA fractions by SDS-PAGE revealed the presence of SDS-resistant multimers of the target protein in the early eluting fractions of the protein peak. To confirm that these bands were indeed multimers of the target protein and not bacterial protein contaminants, an immunoblot of the fractions was performed using a primary antibody that recognizes the 6x histidine tag. As suspected, the multimer bands were recognized and stained by the anti-his antibody. We speculate that the high cysteine content of the construct facilitates non-native, intermolecular disulfide bond formation during the SDS-denaturation process. Analysis of the velocity sedimentation fractions by SDS-PAGE showed that fractions with a greater rate of sedimentation exhibited SDS-resistant multimers. In addition, we observed that the fractions that produced SDS-resistant multimers had a lesser intensity of brown coloration as compared to fractions that did not produce multimers. We found subsequently that the ratio of iron and sulfide to protein was significantly lower in the fractions that produced multimers, and we speculate that the apo-protein is more likely to produce SDS-resistant disulfide-linked multimers. Iron-sulfur clusters are ligated by four cysteine residues and in the apo form, these cysteine residues would be free to form disulfide bonds adventitiously.

The ZBD construct co-purifies with a similar color that is significantly lighter than the NTD fractions.

Purification of the ZBD construct was performed under the same conditions as the NTD construct using a Ni-NTA agarose column. Similar to the NTD construct, a brown band was observed on the Ni-NTA resin when the protein was loaded onto the column and eluted as goldenbrown fractions. A similar pattern of SDS-resistant multimers was observed for the early eluting fractions, and they were determined to have a lower ratio of iron and sulfide to protein than fractions which did not produce SDS-resistant multimers. The peak ZBD fractions exhibited similar UV-visible spectra to the purified NTD. However, the highest ratio of iron and sulfide atoms to protein observed was 0.9. In addition, the purified ZBD fractions were unstable at 4° C, as evidenced by the appearance of significant aggregation after only 4 hours. Many fractions of near-homogenous ZBD were observed to yield a ladder of bands as for the NTD, and were confirmed to contain a 6x His tag by immunoblot. We conclude that the ZBD is able to bind the iron sulfur cluster, but the binding is less stable than in the NTD construct. Therefore, it is likely that all of the residues required for iron sulfur cluster ligation are present in the ZBD. We speculate that the reduced stability may be due to increased solvent exposure of the iron sulfur cluster in the absence of the larger RPD domain, which may indicate the iron sulfur cluster is located at the interface of the ZBD and RPD.



Figure 3. The NTD has a modular architecture consisting of a ZBD and RPD, similar to T7 gp4. SDS-PAGE (right) and anti-his immunoblot (left) analysis of 36 kDa N24-A333 construct after limited trypsin digestion.

Partial digestion patterns suggest the NTD has a modular architecture consisting of a ZBD and RPD, similar to T7 gp4.

Limited trypsin proteolysis of the NTD (36 kDa) construct produced two stable fragments of 10 kDa and 26 kDa (Figure 3 above). Immunoblot analysis with the anti-his antibody (Ab) stained the 10 kDa species but not the 26 kDa species, indicating the 10 kDa fragment contains the N-terminal His-tag. This domain architecture is consistent with an N-terminal ZBD tethered flexibly to a larger RPD (26 kDa), as is the case in the N-terminal half of T7 primase-helicase. We note that the 10 kDa proteolytic fragment is smaller than the 12 kDa ZBD construct. We also performed the limited trypsin proteolysis in the presence of DNA and RNA substrates specify which but no change in the digestion profile was observed, indicating no conformational changes that would protect or influence cleavage occur in the presence of these substrates.

Electronic absorption spectra of purified Dm helicase fractions

Electronic absorption spectrometry measures the absorption of light in the visible to ultraviolet frequency spectrum of a sample. The absorption peaks pertain to allowed electronic transitions between energy states, and can give useful information about the electronic structure of the molecule, particularly conjugated π systems and transition metal orbitals. Proteins absorb radiation at about 280 nm, caused by π to π^* transitions of aromatic amino acids, and this method has been used to quantify the amount of protein in a given sample. UV-visible spectroscopy is not a conclusive technique, as a single peak could be attributed to many different compounds with the same energy difference between their HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital). Nevertheless, this technique can be useful as supporting evidence for a particular hypothesis.

The UV-visible spectra collected for the NTD construct show a peak at 280 nm corresponding to aromatic amino acids in the protein, a shoulder at about 333 nm, and a peak at about 425 nm (Figure 4 on page 11). For the purposes of this thesis, I have analyzed these UV-visible spectra extensively and have made specific comparisons to previous literature. However, I have recognized that for the study of proteins with iron sulfur clusters in the biochemical literature that this type of analysis is not generally used and does not yield conclusions to distinguish 2Fe-2S, 3Fe-4S, and 4Fe-4S cluster types. Typically this method is used for initial identification of iron sulfur clusters, and more specialized spectroscopic methods such as electronic paramagnetic resonance are used for further characterization.



Figure 4. UV-visible electronic absorption spectrum of NTD. Peaks at 422 nm, 464 nm, and 590 nm suggest the presence of an iron-sulfur cluster by comparison with published examples of UV-visible electronic absorption spectra of protein bound 2Fe-2S and 4Fe-4S clusters.



Figure 5 Comparison of UV-visible spectrum of NTD vs ZBD purified fractions. Pink trace represents NTD spectrum, both red and green are ZBD fractions after Ni-NTA fractionation.

Optical features of as isolated NTD helicase samples and comparison to ZBD spectra

Figure 4 above shows the UV-visible spectrum of an NTD sample (1 mg/ml, 28.5 μ M). Experimental molar extinction coefficients ε (cm⁻¹ M⁻¹) can be calculated for each peak by the equation $A = \varepsilon \times b \times c$, where A is the absorbance (y axis), b is the cuvette path length (0.5 cm in this study), and c is the molarity (mol/ L) of the analyte (7). The 280 nm peak arising from $\pi \rightarrow \pi^*$ transitions of aromatic residues has an experimental extinction coefficient ($\epsilon = 30000 \text{ cm}^{-1}\text{M}^{-1}$) that is in excellent agreement with the theoretical value (29910 cm⁻¹ M^{-1}) as predicted from amino acid sequence (see appendix on page 45). The shoulder centered on 330 nm ($\epsilon = 10500 \text{ cm}^{-1}\text{M}^{-1}$) and the broad peak at 420 nm ($\varepsilon = 7000 \text{ cm}^{-1} \text{ M}^{-1}$) correspond to ligand-to-metal charge transfer bands (opposite cysteine $S \rightarrow Fe$) characteristic of iron sulfur clusters (8). Less noticeable but apparent from the first derivative plot is a broad shoulder around 455 nm (5600 cm⁻¹ M⁻¹) that also arises from ligand-to-metal charge transfer transitions, and is most commonly observed in 2Fe-2S clusters (9). Spectra obtained for the ZBD were not as reliable as the NTD spectra because of low concentration and scattering due to aggregation in the samples (Figure 5 on page 11). In these spectra the peak at 420 nm was less intense, but the peak at 330 nm was still apparent. Extinction coefficients for these peaks were always very low ($\epsilon = 500 - 1000 \text{ cm}^{-1} \text{ M}^{-1}$) as compared to those for the NTD.

Spectral features under reducing conditions

UV-visible electronic absorption spectra of peak fractions of the NTD construct from a preparation in which the last step used was velocity sedimentation had a spectrum characteristic of iron-sulfur cluster binding proteins with peaks at 330 and 420 nm (Figure 4 on page 11).

Bleaching of the 420 nm peak was observed immediately upon reduction with dithionite (Figure 6 below), and slowly over time in the presence of DTT (Figure 7 on page 14). Both reactions exhibit unchanged first derivative spectra at wavelengths that are not also absorbed by the reducing agents (>350 nm for DTT and >400 nm for dithionite) (10). Decreased absorbance associated with no change in spectral slope characteristics can be interpreted for reduction reactions as either cluster destruction/ loss or change in electronic structure that influences the capability of this electronic transition. This reaction to reducing agents is typical of iron sulfur proteins, and could be caused by destabilization of the cluster leading to cluster loss, especially during chromatographic methods used in purification (11, 12). The single electron reduction of the cluster could also disrupt significantly or change its electronic structure, which will impact the energy and intensity of observed transitions (9, 13, 14, 15, 16). Reduction of the cluster will decrease the electropositive character of iron atom(s) that could disfavor the ligand-to-metal charge transfer responsible for the 420 nm absorption peak.



Figure 6. UV-Visible spectrum of the NTD construct after treatment with a strong reducing agent. An immediate bleaching effect of dithionite on as isolated NTD Dm helicase sample (blue) after addition of 8 fold excess dithionite (red). Change in volume was negligible (2 µL in 500 µL) compared to the observed change in absorption at 420 nm.



Figure 7. Effect of DTT on the NTD *Dm* helicase sample after 36 hours on ice (red). As isolated NTD (blue) is less intense in 300-325nm range. Concentration of as isolated NTD *Dm* helicase is 28.5 μ M. The oxidation of DTT to the cyclic configuration is accompanied by increased absorption below 270-340 nm.

Spectral features of in vitro reconstituted samples

A slow increase in the absorbance at 420 nm was observed after the addition of stoichiometric quantities of ferrous iron (Fe²⁺) and sodium sulfide (S²⁻) to a NTD sample in the presence of 10 mM DTT (Figure 8 on page 15) (17, 18). The observed increase in absorbance at 420 nm can be due to the assembly of protein bound iron sulfur clusters present in the sample. These results suggest that *in vitro* reconstitution has occurred and are supported by the observed 2-fold increase of absorbance at 420 nm. Additionally, the derivative plots are essentially the same showing that the same type of cluster is being formed during reconstitution, and other types of clusters are not forming as they would most likely cause noticeable differences in the derivative spectra (19). The intense absorbance in the UV range is due to DTT; when DTT performs reduction it will cyclize and have increased absorbance (20, 21).



Figure 8. UV-Visible analysis of *in vitro* **reconstitution of iron sulfur clusters**. The bottom most (pink) trace represents 1:1 ferrous iron and sulfide: helicase immediately after addition of NTD helicase (28.5 μ M). The second trace (green) is after 1 hour, the third trace (red) is after addition of 2:1 ferrous iron and sulfide: helicase, and the top trace (blue) is after 12 hours.

Stability of iron-sulfur cluster in the presence of hydrogen peroxide

Addition of hydrogen peroxide up to 100 fold in excess of helicase resulted in no change in the 280 nm/ 420 nm normalized absorbance ratio (Figure 9 on page 16). These results indicate that the NTD helicase can retain its iron sulfur cofactor in the presence of mild and high concentrations of reactive oxygen species (ROS), and this property could have important consequences *in vivo* (21, 22).



Figure 9. The Fe-S cluster is resistant to oxidizing conditions. Normalized absorbance plot for hydrogen peroxide titration of NTD Dm helicase. Blue: 7 μ M as isolated NTD Dm helicase, Red: 10:1 H₂O₂: helicase, Green: 100:1 H₂O₂: helicase.

Comparison to literature examples of iron sulfur cluster proteins

A major goal of this spectroscopic investigation was to gain insight into the type of iron sulfur cluster present in the NTD helicase. Table 1 on page 17 presents a UV-visible spectral comparison of published compounds containing either 2Fe-2S or 4Fe-4S clusters, or both. Although the UV-visible spectra of both of these cluster types are very similar, some subtle differences can be discerned. In the 400-500 nm region, 4Fe-4S clusters tend to have a single maximum at a slightly lower wavelength than 2Fe-2S clusters, which commonly have additional (lower intensity) absorption peaks around 450 and 550 nm. Reducing agents are known to destabilize and bleach both types of clusters (23, 24, 25). However, 4Fe-4S clusters are commonly very unstable toward reactive oxygen species such as H₂O₂ and O₂, often degrading completely

or into a stable 2Fe-2S cluster (21, 22, 23, 26, 27). Although these trends have exceptions, they

contribute evidence towards the NTD helicase containing a 2Fe-2S cluster (8, 28, 29, 30, 31).

Table 1. A comparison of UV-visible electronic absorption spectra of 2Fe-2S and 4Fe-4S clusters from published studies. The UV-visible spectra collected for the NTD construct (on page 11) show a peak at 280 nm corresponding to aromatic amino acids in the protein, a shoulder at about 333 nm, and a peak at about 425 nm.

PUBLISHED FE-S PROTEIN	2FE-2S SPECTRAL PEAKS (NM)	4FE-4S SPECTRAL PEAKS (NM)
SYNTHETIC (32)	325, 420, 450	
RUMA (33)	444 (oxidized)	390 (native)
FNR (34)	330, 423, 550	315, 409
FERREDOXIN (9)	414, 455, 592	
SOXR (12)	330, 414, 462, 548	
XPD (35)		405
ADDAB (36)		325, 410
DING (37)		403
RAD3 (38)		400
PRIMASE (39)		315, 400

Chemical analysis of colored fractions by Graphite Furnace Atomic Absorption Spectroscopy

Graphite furnace atomic absorption (GFAA) spectroscopy is a commonly-used technique used to detect very small concentrations (ppb) of transition metals in solution. This method is known for its extreme sensitivity, and therefore smaller amounts of sample can be used for analysis. Although the instrument can be time-consuming to troubleshoot to obtain correct atomization conditions, the amount of precious protein that can be saved by choosing this method mitigates the difficulties associated with the procedure. The instrument operates by injecting a small (10-40 μ L) volume of sample into a graphite coated cuvette that can be heated to high temperatures (3000° C). Once the element of interest reaches its atomization temperature, it will achieve swiftly the gas phase, where its characteristic wavelength can be measured for absorption using monochromatic light and a detector. A linear absorption versus concentration plot is obtained by use of standards, and the extrapolated best fit function can be used for sample analysis. GFAA spectroscopy was applied to a purified, concentrated sample of the NTD and ZBD constructs to determine the iron content per protein molecule, as described below.

Iron content

The iron content was determined for numerous NTD and ZBD purifications by GFAA spectroscopy and horse heart cytochrome *c* was used as a standard (cyt *c* contain 1Fe atom per protein molecule). For NTD purifications, the iron content varied from 1.3-2.0 Fe atoms per protein molecule, and from 0.4-0.9 Fe atoms per protein molecule for ZBD purifications. An example of such an analysis of an NTD and ZBD preparations is shown in Table 2 on page 19, with uncertainty values that were typical for this analysis.

Zinc content

The NTD and ZBD helicase samples were also tested for zinc content, and the determined zinc content was not statistically different from the background (data not shown). The standard cytochrome *c* sample was analyzed by the same method and the result was the same (as expected), supporting the reproducibility of the method used to obtain this negative result. The lack of zinc suggests that the ZBD does not bind zinc like the ZBD of T7 gp4.

Table 2. Calculated ratios of Fe: protein for the helicase and cyt c samples from graphite furnace atomic absorption data. Sample concentration was determined by the absorbance at 280 nm for purified helicase variants and by Bradford for the cyt *c* sample. All samples were measured in triplicate. Error bars were calculated by propagated error analysis and all measurements, dilutions, etc were assigned an uncertainty that was appropriate for the measurement, instrument, pipette, etc. (ex: my P20 pipette was calibrated and found to have a standard deviation of 0.6 μ L, so any measurement or transfer with this pipette, was given an uncertainty equivalent to the standard deviation (ex: 9.0 ± 0.6 μ L).

Sample	Sample concentration (µM)	Calculated Fe concentration (µM)	mols of Fe per mol of protein
NTD Dm helicase	3.6 ± 0.1	4.7 ± 0.2	1.3 ± 0.1
Horse heart cytochrome <i>c</i>	6.9 ± 0.4	5.8 ± 0.2	0.85 ± 0.06
ZBD Dm helicase	9.6 ± 0.8	9.2 ± 0.3	0.95 ± 0.23

Sulfide determination by methylene blue colorimetric method

Sulfide content was determined by a colorimetric method (see example in appendix on page 49) that monitored methylene blue formation via the reaction of sulfide with N, N-dimethylp-phenylenediamine (DMPD) (40). Purified fractions of the NTD were determined to contain 2.2 \pm 0.2 sulfide atoms per protein. Together, the iron and sulfide content of the NTD samples support the presence of an iron-sulfur cluster. If the iron sulfur cluster is of the 2Fe-2S type, the iron content indicates that approximately 25% of the protein is without a 2Fe-2S cluster and may be a result of degradation over the course of the purification, which is common for iron sulfur cluster proteins. We cannot rule out the possibility that the iron sulfur cluster is the 4Fe-4S type and >65% of the total protein is apo-protein.

Electron paramagnetic resonance studies

EPR, is the most common method reported in the literature to identify the iron-sulfur cluster type. The spectral characteristics of 2Fe-2S, 3Fe-4S, and 4Fe-4S have been well
documented in each possible redox state and can be used for comparison. Below is a general introduction to the method and background of EPR.

An electron has an intrinsic spin that imparts an angular spin momentum (ms) of $\pm 1/2$ h, where h is Planck's constant ($6.626 \times 10^{-34} \text{ J} \times \text{s}$). According to Pauli's exclusion property, two electrons cannot have the same spin quantum number (ms) if they inhabit the same orbital. This property gives rise to paramagnetic and diamagnetic molecules, where the latter signifies that all electrons are paired, whereas the former defines a molecule with unpaired electrons. A molecule with unpaired electron(s) imparts an angular momentum vector to the molecule, and this property can be influenced by an applied magnetic field. Therefore, when paramagnetic molecules are subjected to a magnetic field, their electrons spin momentum can either align with the field or oppose it, creating a lower and higher energy state, respectively. The difference in energy of the two states can be probed by EPR. This is done by applying a fixed microwave radiation frequency to the sample (9.46 GHz in our trials) and varying the magnetic field from roughly 400 to 4000 gauss. The equation that describes this technique is $\Delta E = g\beta H$, where E is the energy difference between states, H is the applied magnetic field, B is the bohr magneton (9.274 \times 10⁻²¹ erg), and g is a variable pertaining to the environment of the unpaired electron (similar to the shielding constant in NMR). Absorption peaks on EPR spectra can be used to calculate the g value that varies greater or less than 2.0023193 (value for a free electron), and is caused by the interaction of the spin momentum of the electron with the orbital momentum of the molecule. Therefore, the g value for a given peak will be characteristic for that molecule based on its orbital structure and symmetry, and can provide insight to its electronic structure.

Instrumental interference obscured high field signals

A copper impurity in the low temperature device prevented analysis of the high field g = 2 region of the NTD samples (see appendix on page 47). Signals in this region have been well characterized for iron-sulfur clusters and can be an essential tool in determining the type(s) of clusters present in a sample. Repeating EPR analysis correctly should be a priority, as the g = 2 region can reliably identify different cluster types and even variable oxidation states (28, 41, 42, 43, 31, 44, 45).

Low field signals may indicate cluster degradation upon reduction

The spectra shown in Figures 10-12 below were performed by a student prior to my time in the lab, but the analyses I present are my own. A major difference in the purification of NTD construct by the previous student was that all buffers included EDTA. Figure 10 on page 22 presents the low field EPR spectra of the NTD, which consists of two sharp peaks at g = 4.3 and g = 3.5. The g = 4.3 peak is characteristic of rhombic high spin ferric iron present in solution (possibly chelated by EDTA) or bound at the iron sulfur site by itself due to loss of the other iron from the cluster (46, 47, 48, 49, 41, 50, 51). Upon addition of excess dithionite, a broad low field feature appears from $g \sim 7$ to $g \sim 4$, with the sharp g = 4.3 signal still evident (Figure 11 on page 22). The new broad feature can be attributed to cluster degradation caused by reduction leading to the release of ferric iron which can be chelated by EDTA and sulfhydryl ligands (DTT or dithionite) (50, 52). A similar but less intense spectrum is observed upon addition of the oxidizing agent potassium ferricyanide (Figure 12 on page 23). In this case it is either the oxidation of aqueous ferrous iron or the abstraction of iron from iron sulfur clusters that allows for formation of the EDTA/ DTT iron complexes.



Figure 10. Low field EPR spectrum of as isolated NTD *Dm* helicase. Performed by a student prior to my time in the lab, sample contains EDTA. High field signals near g = 2 are obscured by a known copper contaminant in the instrument (see appendix on page 47).



Figure 11. Low field EPR spectrum of NTD *Dm* helicase after the addition of dithionite. Performed by a student prior to my time in the lab, sample contains EDTA. High field signals near g = 2 are obscured (see appendix on page 47).



Figure 12. Low field EPR spectrum of NTD *Dm* helicase after the addition of ferricyanide. Performed by a student prior to my time in the lab, sample contains EDTA. High field signals near g = 2 are obscured (see appendix on page 47).

Forster resonance energy transfer (FRET) experiments

To examine a possible function of the NTD in DNA binding, I attempted to perform fluorescence anisotropy (FA) binding assays with a single-stranded oligonucleotide fluoresceinlabeled (ssDNA-fl) by the NTD. When FA was attempted, quenching of 80% of the fluorescence intensity occurred, which is the opposite effect of what is expected in an FA analysis. I then investigated if the Fe-S cluster was within range to mediate Forster resonance energy transfer by acting as the acceptor to quench the incident fluorescence. Indeed, as shown in Figure 13 below, titration of increasing amounts of NTD construct increased the quenching of fluorescence. This quenching was absent when the oligonucleotide was not attached to the fluorescein (by using free fluorescein).



Figure 13. Quenching of fluorescence in the presence of NTD construct. Excitation of Fluorescein (red)-labeled ssDNA (ssDNA-fl) in the absence of NTD exhibited normal fluorescence and addition of purified NTD construct (blue) resulted in quenching of fluorescence. Iron sulfur clusters (brown star) can act as acceptors in FRET and result in quenching.

Adding increasing concentrations of salt to preformed NTD - ssDNA-fl complexes was observed to inhibit the quenching, and restore fluorescence intensity (Figure 14 on page 25). Independent experiments were performed using NaCl and MgCl₂ to show that salt inhibition of quenching was not salt-specific and depended only on total ionic strength (Figure 14); addition of 150 mM NaCl or 70 mM MgCl₂ resulted in complete interruption of the fluorescence quenching. Quenching could also be eliminated or reduced by addition of nucleic acids when added to the preformed NTD - ssDNA-fl as competitive substrates (Figure 15 on page 26) A polydT substrate (single stranded DNA, never more than 1-2 kDa) was observed to have a coefficient of inhibition of $k_i = 100$ nM. This k_i observed is similar to the binding affinity determined for the ssDNA-fl substrate, and suggests that the NTD may bind all ssDNA substrates with the weak affinity of approximately 100 nM. Equilibrium of fluorescence intensity was achieved instantaneously after aliquots of salt were added, whereas it required up to 10 min for complete equilibration after DNA solutions were added. Overall, the FRET analyses show that the NTD construct has a weak affinity for DNA substrate and may contribute a functional role.



Figure 14. Increasing ionic strength inhibits the quenching observed by the FRET assay. Salt ions can compete with the ionic interactions at the interface of DNA – protein complexes and increase the rate of dissociation and inhibit re-binding. Addition of salt (NaCl and MgCl₂).to a solution containing 10 nM fl-ssDNA/ 5 uM NTD *Dm* helicase caused an immediate increase in fluorescent intensity. By inhibiting the binding of fl-ssDNA to the NTD the quenching of fluorescence by FRET is prevented because the Fe-S cluster is no longer in close proximity to the fluorescein. Complete inhibition of FRET quenching is observed at 150 mM NaCl and 70 mM MgCl₂.



Figure 15. FRET assay shows that quenching can be inhibited competitively by unlabeled ssDNA. Semi-log plots of the FRET competitive binding assay comparing the polydT substrate with ssRNA substrate, polyrA. Concentrations of DNA and RNA substrates are by nucleotide concentration because the length of individual molecules are not constant.

Proposal and design for mutagenesis of the NTD construct

The spectroscopic and chemical analyses of the purified NTD construct support the presence of an iron sulfur cluster but the residues responsible for ligation of the cluster remain to be determined. The majority iron sulfur clusters are coordinated by 4 cysteine residues and applies to both the 2Fe-2S and 4Fe-4S types. Histidine can substitute for one or more of the cysteine ligands but examples of alternative residues participating in iron sulfur coordination are much less

common. Of the 10 cysteines present in the NTD construct, 9 are conserved among insects and are highlighted in Figure 16 on page 28. Due to the diversity of iron sulfur cluster motifs that have been reported in the literature, all 9 cysteine residues were selected for mutagenesis and will be substituted with an alanine residue. 6 of the 9 selected cysteine residues exist as CXXC motifs (where X represents any amino acid) so each of the 3 cysteine pairs (C68/C71, C102/C105, and C245/C248) will be produced as double alanine substitution variants (CXXC \rightarrow AXXA) because this CXXC motif is a very common iron sulfur motif in which both cysteines participate in cluster coordination. The 3 remaining cysteine residues (C63, C260, and C297) will be produced as single alanine substitution variants for a grand total of 6 alanine substitution variants of the NTD construct.

The variants most likely to bind the iron sulfur cluster are C68A/C71A and C102A/C105A because these are the cysteines that align with the 4 cysteine ligands of the ZBD subdomain that coordinate a zinc ion in T7 gp4 (see figure 2 on page 5). These 4 cysteines of T7 gp4 are also CXXC motifs and it will be very interesting to examine how four strictly conserved cysteines that have been preserved in identical motifs are able to bind zinc vs Fe-S clusters. Perhaps the analysis will show that the cysteines haven't changed and it's the surrounding residues that are important for binding specific metal ions. These residues are present in the ZBD construct of *Dm* helicase and although the construct showed limited stability during purification, the ZBD polypeptide coeluted with a light-brown color and suggests the possibility that the ZBD construct can bind the iron sulfur cluster because all the cysteine residues involved in ligation are located in the ZBD.

C63A C68A/C71A

Dm	41	KEYVDFKRQLRQLNLPHKDGHT <mark>C</mark> LQLE <mark>C</mark> RL C DRNRQPVTN	80
Ds	41	KEYVDFKRQLRQLNLPHKDGHT <mark>C</mark> LQLE <mark>C</mark> RL <mark>C</mark> DRNRQPVTN	80
Xl	72	VTMTEIRQYLRSKNVSFHDGYS <mark>C</mark> LHVPSPFTPEQ	105
Xt	70	VTVTEIRQYLRSKNVSFRDGYS <mark>C</mark> LHVPSPFTPER	103
Mm	55	VTTTEIRQYLRAHGIPFQDGHS <mark>C</mark> LRAPSPFVVSSDIKN	92
Hs	55	VTATEIROYLRGHGIPFODGHS <mark>C</mark> LRALSPFAESSOLKG	92

C102A/C105A

Dm	81	AQKGTDHGLLAYVNKRTGAFI <mark>C</mark> PN <mark>C</mark> DVKTSLTSALLSYQL	120
Ds	81	AQKGTDHGLLAYVNKRTGAFI <mark>C</mark> PN <mark>C</mark> DVKTSLTSALLSYQL	120
Xl	106	DTCLFIDKTTGRFL <mark>C</mark> KATLLEGSWEDFQDSMEL	138
Xt	104	DTCLFIDKTTGRFL <mark>C</mark> KATLLEGSWEDFQDSMEL	136
Mm	93	EKKDAPTSFCLFIDKTTGHFL <mark>C</mark> MTSLAEGSWEDLQASVEG	132
Hs	93	QT-GVTTSFSLFIDKTTGHFL <mark>C</mark> MTSLAEGSWEDFQASVEG	131

C245A/C248A C260A

Dm Ds Xl Xt Mm Hs	245 245 297 295 280 279	CVVCLPYELKTLPQECLPALERFKELIFWLHYDASHSWDA CVVCLPYELKTLPQECLPALERFKELIFWLHYDASHSWDA TTIALPRGASCLPPVLLPYLEQFKRIVLWLGDDL-RSWEA TTLALPRGASCLPPVLLPYLEQFKRIILWLGDDL-RSWEA PTLSLPRGTVCLPPALLPYLEQFRRIVFWLGDDL-RSWEA	284 284 335 333 318 317
115	215		917
		C297A	
Dm	286	RAFALKLDERR <mark>C</mark> LLIRPTETEPAPHLALRRRLNLRHILAK	325
Ds	286	RAFALKLDERR <mark>C</mark> LLIRPTETEPAPHLALRRRLNLRHILAK	325
Xl	337	KLFARKLNMKR <mark>C</mark> SLIRPGDKQPSPLHAFNDGLNIGKILKA	376
Xt	335	KLFARKLNMKR <mark>C</mark> SLIRPGDKQPSPLQALNDGLNITKILKA	374
Mm	320	KLFARKLNPKR <mark>C</mark> SLVRPGNQQPRPLEALNQGLSLPRILRT	359
Hs	319	KLFARKLNPKR <mark>C</mark> FLVRPGDQQPRPLEALNGGFNLSRILRT	358

Figure 16. Alignment of fly, frog and human mtDNA helicases for mutagenesis design and selection of conserved cysteine residues. Cysteines highlighted in green will be produced as single alanine substitutions in the NTD construct and cysteines highlighted in cyan will be produced as double alanine substitutions for each pair of CXXC motifs. Red blocks (α -helical) and blue arrows (β -sheet) below the sequence illustrates the predicted secondary structure.

Experimental approach and possible outcomes

Purification of each variant can be performed via the same steps and conditions used to purify the wild type NTD construct. Comparing the elution patterns, SDS-PAGE gel mobility, etc of the variants with wild type NTD purifications can serve as a method to evaluate stability and physical properties of the variants relative to wild type NTD. In addition, UV-Visible spectra should be acquired for all variants after each fractionation step in purification to evaluate the electronic state of the iron sulfur cluster. Variants that exhibit altered spectral characteristics from the wild type NTD spectra will provide evidence for which cysteines are involved in cluster coordination. The optimal outcome for this mutagenesis analysis would be for a variant to purify as a colorless solution and maintain stability. The absence of an iron sulfur cluster eliminates the possibility for analysis by the FRET binding assay but would allow the variant to be analyzed by fluorescent anisotropy for DNA binding. The same fluorescently-labeled oligo can be used for both FRET and fluorescent anisotropy allowing the two methods to be compared accurately to evaluate the effect of the iron sulfur cluster on DNA binding affinity.

DISCUSSION

Previous work by our lab and others has focused mainly on the C-terminal helicase domain and the linker region of the mtDNA helicase, and has shown that the DNA helicase, NTPase, and oligomerization characteristics of the mtDNA helicase are very similar to those of the T7 primasehelicase. The N-terminal domain of metazoan mtDNA helicase has lost critical residues required for the primase activity, which has been studied extensively in T7 primase-helicase. Despite losing a crucial function, the N-terminal domain retains strong conservation throughout metazoa, and several disease mutations have been identified that affect the NTD, arguing that the NTD has retained an old or gained a new role, that is vital for mtDNA metabolism.

Our analysis of the NTD of *Drosophila* MtDNA helicase presents the first evidence that a metazoan has retained the modular architecture inherent to bacterial DnaG-like and T-odd phage primases. In addition, the highly conserved Motif I that binds zinc in DnaG-like primases has been transformed into an Fe-S cluster-binding motif, presenting the first example of such a change, to our knowledge. Furthermore, all four cysteines of Motif I, which have been shown to ligate zinc, are strictly conserved in fly, suggesting the same side chains act as the ligands for Fe-S cluster binding. The NTD construct bound Fe specifically and showed no presence of Zn, highlighting the powerful influence that the local environment of the binding site has in discrimination of ligand binding.

The characterization of the iron-sulfur is only preliminary despite our attempts to perform EPR; although we obtained electron paramagnetic resonance spectra of the purified NTD construct in the presence and absence of dithiothreitol, a strong reducing agent. In the absence of reducing agent we observed a sharp signal at g = 4.3, which is characteristic of ferric iron (Fe⁺³), and we observed no change in this signal upon the addition of reducing agent or oxidizing agent

(ferricyanide). EPR can be a powerful tool for the identification of iron-sulfur clusters, but only if one is able to reduce the cluster quickly and stably for the duration of the EPR scan. Iron-sulfur clusters are typically EPR-silent in the oxidized state, but exhibit signal motifs in the g = 2 region that are unique to the specific type of iron-sulfur moiety that is present, for example, a 2Fe-2S, 3Fe-4S, or 4Fe-4S. Unfortunately, the only peak we observed in the g = 2 region (see Figure 19 on page 47) was due to a copper contaminant in the instrument itself, and we observed no change in this background signal throughout the analysis. The g = 4.3 signal indicates the presence of free or partially-bound ferric iron in the NTD samples, and could be caused by Fe-S cluster degradation during purification or sample preparation.

A ratio of 1.3-2.0 iron atoms per protein determined by GFAA is consistent with a 2Fe-2S cluster but does not exclude the possibility of 3Fe-4S or 4Fe-4S clusters. This low amount could be caused by substantial cluster loss during purification due to the use of reducing agents, which have been shown to cause cluster degradation by EPR and UV-visible methods. The UV-visible spectra are typical of 2Fe-2S and 4Fe-4S proteins, as is the observed cluster stability in the presence of reactive oxygen species, and in the presence of various reducing and oxidizing agents (21, 22, 23, 26, 27).

Conclusions

Overall, the studies presented here for the ZBD and NTD constructs of the *Drosophila melanogaster* mtDNA helicase only yield limited conclusions. The UV-visible spectra, iron content, and sulfide analyses indicate that the NTD construct contains an iron-sulfur cluster, but the cluster type remains inconclusive. The EPR spectra do not contribute any significant information other than that the iron sulfur cluster may be unstable in our purified fractions. The FRET binding experiments can be used to measure a weak affinity of the NTD construct for DNA substrates, and this interaction may be functional. Further studies must be performed to add upon this preliminary data.

METHODS

Bacterial cell growth and protein overexpression

E. coli cells containing IPTG-inducible plasmid with cDNA encoding Drosophila melanogaster mitochondrial helicase residues N24-A333 with N-terminal 6x His tag were grown, induced, and harvested via centrifugation to produce pellets that were frozen in liquid nitrogen and stored at -80° C until purification.

Cell lysis and preparation of soluble fraction (FrI)

Pellets were thawed on ice and resuspended in Tris-sucrose buffer (50 mM Tris-HCl pH 7.5, 10% (w/v) sucrose, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium metabisulfite, 2 µg/mL leupeptin). 5x Lysis buffer (1.25M NaCl, 7.5% n-dodecyl- β -D-maltoside, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium metabisulfite, 2 µg/mL leupeptin) was added in a volume 1/5 of the resuspended solution (Final concentration ~0.25 M NaCl, 1.5% n-dodecyl- β -D-maltoside). Solution was frozen with liquid nitrogen and thawed on ice. Solution was centrifuged at 12 krpm for 50 minutes at 4° C in a SS-34 rotor. Supernatant was collected and pellet discarded. Golden-brown supernatant (FrI) is soluble fraction and total protein was determined by colorimetric Bradford method.

Ni-NTA metal affinity chromatography (FrII)

A column of Ni-NTA agarose resin was prepared in a ratio of 1 mL resin/ 50 mg total protein. The column was equilibrated by flowing 10 column volumes (CV) of equilibration buffer (35 mM Tris-HCl pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 15 mM imidazole, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium metabisulfite, 2 µg/mL leupeptin) through the resin. FrI was loaded onto the column at 2 CV/hr. Column was washed with 10 CV of wash buffer (35 mM Tris-HCl pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 20 mM imidazole, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM KCl, 10% (v/v) glycerol, 20 mM

metabisulfite, 2 µg/mL leupeptin) and 1/4 CV fractions were collected. Column was washed with 4 CV elution buffer 1 (35 mM Tris-HCl pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 200 mM imidazole, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium metabisulfite, 2 µg/mL leupeptin) to elute target protein and 1/8 CV fractions were collected. Column was washed with 3 CV elution buffer 2 (35 mM Tris-HCl pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 500 mM imidazole, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium metabisulfite, 2 µg/mL leupeptin) to elute target protein and 1/8 CV fractions were collected. Column was washed with 3 CV elution buffer 2 (35 mM Tris-HCl pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 500 mM imidazole, 5 mM β -mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium metabisulfite, 2 µg/mL leupeptin) to elute any residual protein bound to column and 1/4 CV fractions were collected. Collected fractions were analyzed via SDS-PAGE (12% polyacrylamide gels) and fractions were pooled according to target protein content and purity (FrII). Total protein content of FrII was determined to be 10 mg/ mL by colorimetric Bradford method and SDS-PAGE analysis estimated a purity of 85%.

Glycerol gradient sedimentation (FrIII)

12-30% glycerol gradients were prepared (35 mM Tris-HCl pH 7.5, 250 mM NaCl, 5 mM β -mercaptoethanol). FrII was loaded onto gradients and centrifuged at 38 krpm for 72 hrs at 4° C in a SW-40 rotor. Gradients were fractionated, analyzed by SDS-PAGE (12% polyacrylamide gels), and fractions were pooled according to purity (FrIII). Total protein content of FrIII was determined by colorimetric Bradford method and SDS-PAGE analysis revealed no contaminant bands suggesting a purity of >95%.

Graphite furnace atomic absorption analysis of iron content of NTD (or ZBD) purified fractions

A Hitachi Z-9000 GFAA spectrometer was used for iron determination courtesy of Dr Kathryn Severin. Iron standard solutions were prepared from a 1 mM iron (III) nitrate nonahydrate (crystal, Sigma) solution containing 2% HNO₃ (70%, EMD) for each concentration listed in Table 1. Standard solutions additionally contained 3.33% glycerol (100%, J.T. Baker), 0.167 mM EDTA disodium dehydrate (crystal, 100.8%, J.T. Baker), and 11.67 mM Tris-HCl pH 7.5 (Ultrapure, Research Organics). Digestion of protein samples was achieved by adding 76.0 \pm 0.5 µL of the helicase construct (or 8.1 \pm 0.5 µL of cyt c) to 0.5mL HNO₃ and 0.5mL H₂O₂ in a 5mL glass vial. This solution was heated carefully until boiling with occasional manual stirring to avoid bumping. Heating was continued until complete evaporation of solution, and the resulting ash was redissolved by adding 200.0 \pm 0.5 µL HNO₃, and then subsequently diluted to 3.00 \pm 0.01 mL with doubly distilled H₂O. Digested samples were sealed with parafilm to avoid iron contamination from screw-caps during storage. Standard solutions and digested samples were ran in triplicate on the GFAA spectrometer using a drying temperature of 80-120° C for 30 seconds, ashing temperature of 630° C for 30 seconds, and an atomization temperature of 2700° C for 10 seconds. Absorbance values were recorded at a wavelength of 248.3 nm.

Quenching of fluorescence of fluorescein-labeled ssDNA (ssDNA-fl) by NTD via FRET

Fluorescence intensity was measured from 500-550 nm using a Jobin Yvon Spex Fluorolog-3 spectrofluorimeter with an excitation wavelength of 480 nm. In a 1 mL cuvette, 500 uL of fl-dna solution (10 nM fl-dna oligo/ 50 mM tris-HCl pH 7.5) was added and the fluorescence spectrum was recorded. A small aliquot (1-2 uL) of concentrated NTD solution was added to this solution for a total concentration of 10 nM NTD and mixed. Record fluorescence spectrum for this solution and continue for variable NTD concentrations from 10 nM to 5000 nM by successive addition of 2 uL aliquots of purified NTD. It was often necessary to wait 2-10 min after mixing until equilibrium was reached.

Calculate dissociation constant Kd for fl-dna binding by NTD

Convert data and plot bound protein (BP) versus free protein (FP). Conversion is done by assuming 1 binding site per fl-dna for NTD, so $BP_{max} = 10$ nM. Fit plot using the following equation to determine K_d.

$$[BP] = \frac{BP_{max} [FP]}{K_d + [FP]}$$

Methylene Blue Photometric Analysis of Sulfide Content in Protein Samples

This general method was used to determine the sulfide content of protein samples suspected of containing iron sulfur clusters. Add protein solution to a microcentrifuge tube in an amount equivalent to 1 nmol protein. For standard solutions add an amount of 1 mM Na₂S•9H₂O in 0.1 M NaOH equivalent to 0.5, 1, 2, 3, and 4 nmol S²⁻. Dilute to 100 μ L with ddH2O. Add 300 μ L 1% aqueous solution of zinc (II) acetate dihydrate then add immediately 15 μ L 12% aqueous NaOH. Perform additions to one tube at a time. Let tubes sit for 30 min with inversions. Underlay suspension with 75 μ L DMPD solution (0.1% N,N-dimethyl-p-phenylenediamine (DMPD)•HCl in 5 M HCl), then add 30 μ L FeCl₃ solution (23 mM FeCl₃ in 1.2 M HCl) to the bottom layer and invert the tube immediately to make the solution homogenous. Centrifuge tubes at 2 krpm for 20 minutes. Transfer to cuvette and record visible spectrum from 600-800 nm after at least 30 minutes have elapsed since the addition of FeCl₃. Sulfide reacts with DMPD in the presence of Fe³⁺ at acidic conditions (optimal pH = 0.4-0.7) to form methylene blue chromophore which can be detected quantitatively by absorbance at 670 nm ($\epsilon = 34500 \text{ cm}^{-1} \text{ M}^{-1}$). See appendix on page 49 for example spectrum of standards.



Inhibition of FRET quenching by salt

Fluorescence intensity was measured from 500-550 nm using a Jobin Yvon Spex Fluorolog-3 spectrofluorimeter with an excitation wavelength of 480 nm. In a 1 mL cuvette, 500 uL of fl-dna solution (10 nM fl-dna oligo/ 50 mM tris-HCl pH 7.5) was added and the fluorescence spectrum was recorded. Concentrated (10 mg/mL) NTD solution was added for a final concentration of 5 uM to achieve max quenching and fluorescence spectrum was recorded after mixing and waiting 10 min. A small aliquot (1-2 uL) of concentrated salt solution (NaCl or MgCl₂) was added to this solution for a total concentration of 5-10 mM salt and mixed. Record fluorescence spectrum for this solution and continue for variable NaCl or MgCl₂ concentrations (10-1000 mM) by successive addition of aliquots of 3M NaCl or 1M MgCl₂.

Competitive FRET binding assay for polynucleic substrates

Fluorescence intensity was measured from 500-550 nm using a Jobin Yvon Spex Fluorolog-3 spectrofluorimeter with an excitation wavelength of 480 nm. In a 1 mL cuvette, 500 uL of fl-dna solution (10 nM fl-dna oligo/ 50 mM tris-HCl pH 7.5) was added and the fluorescence spectrum was recorded. Concentrated (10 mg/mL) NTD solution was added for a final concentration of 5 uM to achieve max quenching and fluorescence spectrum was recorded after mixing and waiting 10 min (this spectrum is used for F₀ in calculations below). A small aliquot

(1-2 uL) of concentrated substrate solution (polydT, polyrA, polyrA: dT) was added to this solution for a total concentration of 5 nM-1 uM substrate (determined as nucleotide concentration) and mixed. Record fluorescence spectrum for this solution and continue for variable substrate concentrations (10 nM-100 uM determined as nucleotide concentration) by successive addition of aliquots of concentrated substrate (polydT, polyrA, polyrA:dT). After the max substrate concentration is reached, NaCl was added to a final concentration of 150 mM and spectrum was recorded. Addition of salt is used to determine F_{max} , defined as the fluorescent intensity when NTD- fl-dna binding is completely inhibited.

Calculate the inhibition coefficient K_i for each substrate

Determine IC₅₀ (inhibitor concentration at 50% inhibition) by fitting a function: y = a + b*log(x), where x = [I] (concentration of inhibitor), y = % inhibition. Once a and b are determined, use the function to determine x when y = 50. The % inhibition is defined as $(F/F*_{max})*100$, where F = the fluorescent intensity at 522 nm for a given [I], $F*_{max}$ is $F_{max} - F_0$, F_0 is the fluorescent intensity at 522 nm when no inhibitor is present and F_{max} , is the fluorescent intensity at 522 nm when no inhibitor is present and F_{max} , is the fluorescent intensity at 522 nm when no inhibitor. Determine K_i via the following equation, where $K_d = 122$ nM and [S] = [NTD] = 5 uM. See appendix on page 48 for example of plot and fit used to calculate the IC₅₀.

$$K_i = \frac{IC_{50}K_d}{[S] + K_d}$$

EPR Spectroscopy

100-150 uL samples of concentrated NTD solutions (10 mg/ mL) were frozen in liquid nitrogen prior to analysis. For oxidized samples, 10-fold excess potassium ferricyanide was added 10 min prior to freezing. For reduced samples, 10-fold dithionite was added 10 min prior to freezing. All spectra were recorded on instruments in the McCracken lab. Baselines were recorded using sample buffer (35 mM tris-HCl/ 50 mM NaCl/ 15% glycerol). All experiments were performed at 4 K.

UV-Visible Spectroscopy

All spectra were recorded on an Agilent 8453 UV-Visible spectrophotometer. For all analyses, a quartz cuvette with a pathlength of 0.5 cm was used and required a minimal volume of 0.5 mL.

Methods for mutagenesis

Growth and preparation of Ca^{2+} competent cells

Add 1 mL of an overnight culture of bacteria to 20 mL of LB brother in a 125 mL flask. Grow the cells at 37° C in a water bath with shaking. Monitor OD595 every 0.5 hr and record/plot growth. When the culture has $OD_{595} = 0.2$ place the flask on ice for 3-5 min to chill the bacteria. Transfer to a chilled Oakridge tube and centrifuge in the SS-34 rotor at 10 krpm for 2 min. Decant supernatant and resuspend cells in 2.5 mL of ice-cold 0.05 M CaCl₂. Leave cells on ice for >15 min. Centrifuge cells in the SS-34 rotor at 10krpm for 2 min and remove supernatant with a pipette instead of decanting. Resuspend cells in 2.5 mL of ice-cold 0.05 M CaCl₂.

Transformation of Ca²⁺ competent cells

To 5 micro-centrifuge tubes on ice, add 50 μ L of TCM. Add DNA- (Plasmid from miniprep) to tubes 1, 2, and 3 in the amount of 1 μ L, 2 μ L, and 3 μ L, respectively, and leave 2 tubes without DNA for controls by later plating them +/- 25 μ g/ mL kanamycin. To each tube add 100 μ L of cells. Flick the tube gently with your finger and incubate on ice > 20 min. Incubate tubes for 2 min at 42° C. Add 1 mL LB broth to each tube. If you plan to select for drug resistance incubate 30 min in 37° C (WITH NO DRUG ADDED TO BROTH). Otherwise plate on LB plates. Plate the transformed cells on LB plates containing drug by spreading or by addition of 2.5 mL of top agar to the transformed cells after putting your transformed cells into 13 x 100 tubes. Incubate plates for 1-2 days. If plate for plaques with M13 or derivatives of M13, add indicator bacteria (0.2-0.3 mL of an overnight culture to the sample and use top-agar for plating). If you are using one of the M13 mp phages add 10 μ L of 0.1 M IPTG, 50 μ L 2% X-Gal to top agar.

Preparation of electrocompetent bacteria (XL-1 Blue)

Inoculate a day culture in LB medium with appropriate antibiotics (15 µg/mL tetracycline) by diluting an overnight culture 200 times (500-1000 mL of culture, 250 mL per 1L flask). Grow bacteria with shaking at appropriate temperature until OD₅₉₅ is about 0.35-0.40. Take measurement every 30 min for growth curve. At an OD₅₉₅ of 0.35-0.40 place bacteria on ice for at least 30 min. Harvest bacteria by spinning in 250 mL bottles in Sorval GSA rotor for 10 min at 8,000 rpm. Resuspend bacterial pellet in original volume of ice-cold sterile ddH₂O. Keep on ice for at least 10 min. Harvest bacteria by spinning as above. Gently remove supernatant. Bacterial pellet is very fragile at this point. Resuspend the pellet in 1/2 original volume of ice-cold sterile ddH₂O. Spin in 250 mL bottles in Sorval GSA rotor for 10 min at 8,000 rpm. Resuspend the pellet

in 5 mL ice-cold sterile 10% (v/v) glycerol and transfer into oackridge tubes (equivalent of 500 mL of original culture per tube). Spin at 8000 rpm for 10 min. Remove supernatant, resuspend the pellet in 0.6 mL of sterile 10% (v/v) glycerol. Make aliquots of 130 μ L and flash freeze in liquid N₂. Store at -70° C.

Mini-preparation of plasmid DNA: pET-28a- Dm mt helicase N24-A333 from BL21(DE3) cells

To obtain pET-28a-dm mt helicase N24-A333 DNA from BL21(DE3) bacterial cells, Inoculate two colonies from pET-28a-dm mt helicase N24-A333 BL21(DE3) plate in 2 separate tubes of 3 mL liquid L-broth w/ 15 μ g/ mL kanamycin. Grow cultures overnight at 37° C w/ shaker. Centrifuge cultures in micro-centrifuge tubes for 1 min at 13000 rpm. Discard supernatant. Resuspend cells in 300 μ L of P1 (50 mM Tris-HC1 pH 7.5/ 10 mM EDTA pH 8.0/ 0.1 mg/mL RNase) by vortex. Make sure all cells are resuspended. Add 300 μ L of P2 (200 mM NaOH/ 1% SDS) and invert tubes ~5 times. Incubate at room temperature for no longer than 5 min. Add 300 μ L of P3 (3M potassium acetate pH 5.5), mix by inversion, and incubate 2 min at room temperature. Centrifuge at 13000 rpm for 10 min. Collect supernatant and transfer to new microcentrifuge tubes. Add 400 μ L 100% isopropanol, incubate 5 min, and centrifuge at 13000 rpm for 10 min. Discard supernatant. Add 600 μ L 70% ethanol and resuspend pellet. Centrifuge again at 13000 rpm for 5min. Discard supernatant. Let the pellet sit out for 20 min to dry and resuspend in 20 μ L of ddH₂O.

Transformation of pET-28a-dm mt helicase N24-A333 plasmid into XL-1 Blue cells via electroporation

Thaw one tube of XL-1 Blue competent cells on ice. Distribute 40 μ L into one new tube. For controls, distribute 40 μ L into a separate tube and do not add any plasmid. Add 1 μ L of miniprep DNA and mix gently for 2 min. Transfer cells/plasmid to the E. Coli Pulser cuvette, set the voltage of the electroporator to 2.3 V and hit on. Transfer electroporated cells immediately to 1 mL L-broth (WITHOUT DRUGS) and grow them at 37° C for ~1 hr w/ shaker. Plate 20, 100, and 500 μ L on separate agar plates containing drug. Plate 500 μ L of the control on a drug containing plate and 500 μ L on a plate without drugs. Incubate plates overnight at 37° C.

Mini-preparation of plasmid DNA: pET-28a-Dm mt helicase N24-A333 from XL-1 Blue cells

To obtain methylated pET-28a-dm mt helicase N24-A333 DNA from XL-1 Blue bacterial cells, Inoculate two colonies from pET-28a-dm mt helicase N24-A333 XL-1 Blue plate in 2 separate tubes of 3 mL liquid L-broth w/ 15 μ g/ mL kanamycin. Grow cultures overnight at 37° C w/ shaker. For both overnight cultures, transfer 1 mL into each of 3 micro-centrifuge tubes and centrifuge for 1 min at 13000 rpm. Discard supernatant. Resuspend cells in 100 μ L of P1 (50 mM Tris-HCl pH 7.5/ 10 mM EDTA pH 8.0/ 0.1 mg/mL RNase) by vortex. Make sure all cells are resuspended. Combine the three 100 μ L suspensions into 1 tube for both cultures. Add 300 μ L of P2 (200 mM NaOH/ 1% SDS) to both tubes and invert tubes ~5 times. Incubate at room temperature for no longer than 5 min. Add 300 μ L of P3 (3M potassium acetate pH 5.5) to both tubes, mix by inversion, and incubate 2 min at room temperature. Centrifuge at 13000 rpm for 10 min. Collect supernatant and transfer to new micro-centrifuge tubes. Add 400 μ L 100% isopropanol, incubate 5 min, and centrifuge at 13000 rpm for 10 min. Discard supernatant. Add 600 μ L 70% ethanol and resuspend pellet. Centrifuge again at 13000 rpm for 5min. Discard supernatant. Let the pellet sit out for 20 min to dry and resuspend in 20 μ L of ddH₂O.

Quick change mutagenesis of dm mtDNA helicase N24-A333

Dilute mini-prep to 10 ng/ μ L. Prepare master mix (0.2 mM dNTP's, 1X pfu stratagene turbotag buffer, 2.5 units/ reaction pfu turbotag pol stratagene). Add 1 μ L of each primer and 5 μ L plasmid. PCR program consisting of 25 cycles of 95° C for 30 sec, 55° C for 60 sec and 68° C

for 12 min. Remove tubes from thermocycler and add 1 μ L of DpnI to all tubes except tube 8. Incubate at 37° C for 1hr. Add 1 μ L of each reaction in 40 μ L of XL-1 Blue cells. Mix gently for 1-2 min. Transfer cells/reaction to E. coli pulser cuvette, set the voltage of the electroporator to 2.3 V, and hit on. Transfer electroporated cells immediately to 500 μ L L-broth (WITHOUT DRUG) and grow cells at 37° C for ~1 hr w/ shaker. Spin cells down for 1 min at 13000 rpm. Resuspend them in 200 μ L L-broth and plate on agar plates (WITH DRUG). Incubate plates overnight at 37° C.

APPENDIX

APPENDIX

Amino acid sequence of the N-terminally 6x His-tagged NTD (N24-P333) construct of *Dm* mtDNA helicase. The sequence of the ZBD (N24-P123) construct is underlined and the 10 additional residues of the 6x His tag are in bold format. The ZBD construct also contains the N-terminal 6x His tag.

MGHHHHHHATN₂₄YATQVVSGLEECSLDPKEYVDFKRQLRQLNLPHKDGHTCLQLEC RLCDRNRQPVTNAQKGTDHGLLAYVNKRTGAFICPNCDVKTSLTSALLSYQLPKP₁₂₃ VGYKQPLQRQPVYESRFPHLAVVTPEACAALGIKGLKEDQLNAIGAQWEPQQQLLHFK LRNAAQVEVGEKVLYLGDRREEIFQSSSSSGLLIHGAMNKTKAVLVSNLIDFIVLATQNI ETHCVVCLPYELKTLPQECLPALERFKELIFWLHYDASHSWDAARAFALKLDERRCLLI RPTETEPAPHLALRRRLNLRHILAKATPVQHKA333

Table 3. Physical properties based on primary sequence of NTD helicase. Theoretical physical properties as determined from amino acid sequence of *Dm* mtDNA helicase N24-A333.

Molecular Weight (Daltons)	Isoelectric Point (pI)	Extinction Coefficient ($cm^{-1}M^{-1}$)
36283.8	8.81	29910 (280 nm)

Table 4. Protein content and purity of purification steps. Total protein concentration was determined by A₂₈₀ analysis and by the colorimetric Bradford method throughout purification. Purity was estimated by SDS-PAGE analysis of target band versus contaminant bands.

Pool Name	Purification Step	Total Protein Concentration by A ₂₈₀ (mg/ mL)	Total Protein Concentration by Bradford (mg/ mL)	Predicted Purity (%)
FrI	Cell Lysis	ND	7	20
FrII	Ni-NTA metal affinity chromatography	ND	10	85
FrIII	Glycerol Gradient Sedimentation	1.1	1.4	>95

Fr I Fr IIa Fr IIb Fr IIIa Fr IIIb



Figure 17. SDS-PAGE analysis of NTD construct throughout the purification. We suspect an SDS-resistant disulfide-linked dimer is the band at 70 kDa.



Figure 18. Gel filtration analysis of NTD construct using a Superdex 75 column. The elution of a single peak corresponding to a monomer as determined by the A_{280} measurements shown as the blue trace.



Figure 19. Instrument interference at g = 2 prevented further characterization of Fe-S characteristics by EPR. This EPR spectrum was taken with an empty sample tube to show that the large signal at g = 2 is suspected to be a copper impurity in the instrument and not due to sample contamination.



Figure 20. Example of semi-log plot and fit used to calculate the IC50. (The x axis is logarithmic).



Figure 21. Example UV-Visible spectrum for standards: The reaction is considered successful if the ratio of absorbance between A_{670} and A_{750} is close to 2 as is observed in these spectra. Blue- .5 nmol sulfide, Red- 1 nmol sulfide, Pink- 2 nmol sulfide, Dark green- 3 nmol sulfide, Brown- 4 nmol sulfide.

BIBLIOGRAPHY

BIBLIOGRAPHY

- 1. Ziebarth, T., Farr C., Kaguni, L. (2007). Modular Architecture of the Hexameric Human Mitochondrial DNA Helicase. *J. Mol. Biol.*, 367, 1382-1391.
- 2. Matsushima, Y., Kaguni, L. (2009). Functional importance of the conserved N-terminal domain of the mitochondrial replicative DNA helicase. *Biochimica et Biophysica Acta*, 1787, 290–295.
- 3. S. S. Patel, K. M. Picha. STRUCTURE AND FUNCTION OF HEXAMERIC HELICASES1. Annual Review of Biochemistry, 69, 651-697
- 4. Shutt TE, Gray MW. (2006). Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *Journal of Molecular Evolution*, 62, 588-599.
- Longley MJ, Humble MM, Sharief FS, Copeland WC. (2010) Disease variants of the human mitochondrial DNA helicase encoded by C10orf2 differentially alter protein stability, nucleotide hydrolysis, and helicase activity. *Journal of Biological Chemistry*, 285, 29690-702
- 6. Ilyina TV, Koonin EV. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acid Research*, 20, 3279-85.
- Schmid, F. (2001). Biological Macromolecules: UV-Visible Spectrophotometry. *ENCYCLOPEDIA OF LIFE SCIENCES*, Macmillan Publishers Ltd, Nature Publishing Group / <u>www.els.net</u>.
- 8. Tsibri, J., Woody, R. (1970). Structural Studies of Iron-Sulfur Proteins. *Coordin. Chem. Rev.*, 5417-458.
- 9. Noodle, L., Baerends, E. (1984). Electronic Structure, Magnetic Properties, ESR, and Optical Spectra for 2-Fe Ferredoxin Models by LCAO-Xα Valence Bond Theory. *J. Am. Chem. Soc.*, 106, 2316-2327.
- 10. Iorio, E. (1981). Preparation of Derivatives of Ferrous and Ferric Hemoglobin. *METHODS IN ENZYMOLOGY*, VOL. 76.
- Hidalgo, E., Bollinger, J., Bradley, Terence., Walsh, C., Demple, B. (1995). Binuclear [2Fe-2S] Clusters in the Escherichia coli SoxR Protein and Role of the Metal Centers in Transcription. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, Vol. 270, No. 36, Issue of September 8, pp. 20908–20914.
- 12. Hidalgo, E., Demple, B. (1994). An iron-sulfur center essential for transcriptional activation by the redox-sensing SoxR protein. *The EMBO Journal*, 13, 138-146.

- Torres, R., Lovell, T., Noodleman, L., Case, D. (2003). Density Functional and Reduction Potential Calculations of Fe4S4 Clusters. *J. AM. CHEM. SOC.*, 125, 1923-1936.
- Mouesca, J., Chen, Jun., Bashford, D., Case, D. (1994). Density Functional Poisson-Boltzmann Calculations of Redox Potentials for Iron-Sulfur Clusters. J. Am. Chem. SOC. 116, 11898-11914.
- 15. Stephens, P. J., Jollie, D. R., Warshel, A. (1996). Protein Control of Redox Potentials of Iron-Sulfur Proteins. *Chem. Rev.* 96, 2491-2513.
- Noodleman, L., Norman, J., Osborne, J., Aizman, A., Case, D(1985). Models for Ferredoxins: Electronic Structures of Iron-Sulfur Clusters with One, Two, and Four Iron Atoms. J. Am, Chem. SOC. 107, 3418-3426.
- 17. Malkin, R., Rabinowitz, J. (1966). The Reconstitution of Clostridial Ferredoxin. *Biochemical and Biophysical Research Communications*, 23, No. 6.
- Zheng, C., Zhang, Y., Liu, Y., Wu, A., Xia, L., Zeng, J., Liu, J., Qiu, G. (2009). Characterization and Reconstitute of a [Fe4S4] Adenosine 5'-Phosphosulfate Reductase from Acidithiobacillus ferrooxidans. *Curr. Microbiol.*, 58, 586-592.
- 19. Rickard, D., Luther III, G. (2007). Chemistry of Iron Sulfides. *Chem. Rev.*, 107, 514-562.
- Johnson, C., Asher, S. (1987). UV Resonance Raman Excitation Profiles of 1-Cystine. Journal of Raman Spectroscopy, 18, 345-349.
- Ding, H., Demple, B. (1998). Thiol-Mediated Disassembly and Reassembly of [2Fe-2S] Clusters in the Redox-Regulated Transcription Factor SoxR. *Biochemistry*, 37, 17280-17286.
- Reents, H., Gruner, I., Harmening, U., Bottger, Lars., Layer, G., Heathcote, P., Trautwein, A., Jahn, D., Hartig, E. (2006). Bacillus subtilis FNR senses oxygen via a [4Fe-4S] cluster coordinated by three cysteine residues without change in the oligomeric state. *Molecular Microbiology*, 60, 1432-1445.
- Lillig, C., Berndt, C., Vergnolle, O., Lonn, M., Hudemann, C., Bill, E., Holmgren, A. (2005). Characterization of human glutaredoxin 2 as iron-sulfur protein: A possible role as redox sensor. *PNAS*, 102, 8168-8173.
- Bych, K., Kersercher, S., Netz, D., Pierik, A., Zwicker, K., Huynen, M., Lill, R., Brandt, U., Balk, J. (2008). The iron-sulphur protein Ind1 is required for effective complex 1 assembly. *The EMBO Journal*, 27, 1736-1746.
- Dailey, H., Finnegan, M., Johnson, M. (1994). Human Ferrochelatase is an Iron-Sulfur Protein. *Biochemistry*, 33, 403-407.
- 26. Singh, A., Guidry, L., Narasimhulu, K., Mai, D., Trombley, J., Redding, K., Giles, G., Lancaster, J., Steyn, A. (2007). Mycobacterium tuberculosis WhiB3 responds to O₂ and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient starvation survival. *PNAS*, 104, 11562-11567.

- 27. Jervis, A., Crack, J., White, G., Artymiuk, P., Cheesman, M., Thomson, A., Brun, N., Green, J. (2009). The O₂ sensitivity of the transcription factor FNR is controlled by Ser24 modulating the kinetics of [4Fe-4S] to [2Fe-2S] conversion. *PNAS*, 106, 4659-4664.
- 28. Sweeney, W., Rabinowitz, J. (1980). Proteins Containing 4Fe-4S Clusters: An Overview. *Ann. Rev. Biochem.*, 49, 139-161.
- 29. Orme-Johnson, W. H. (1973) Iron-Sulfur Proteins: Structure and Function. *Annu. Rev. Biochem*, 42, 159-204.
- 30. Koay, M., Antonine, M., Girtner, W., Lubitz, W. (2008). Modeling Low-Potential [Fe4S4] Clusters in Proteins. *Chemistry and Biodiversity*, 5, 1571.
- McLean, K., Warman, A., Seward, H., Marshall, K., Girvan, H., Cheesman, M., Waterman, M., Munro, A. (2006). Biophysical Characterization of the Sterol Demethylase P450 from Mycobacterium tuberculosis. Its Cognate Ferredoxin, and their Interactions. *Biochemistry*, 45, 8427-8443.
- 32. Kennedy, M., Kent, T., Emptage, M., Merkle, H., Beinert, H., Munck, E. (1984). Evidence for the Formation of a Linear [3Fe-4S] Cluster in Partially Unfolded Aconitase. *The Journal of Biological Chemistry*, 259, 14463-14471.
- Agarwalla, S., Stroud, R., Gaffney, B. (2004). Redox Reactions of the Iron-Sulfur Cluster in a Ribosomal RNA Methyltransferase, RumA. *The Journal of Biological Chemistry*, 279, 34123-34129.
- 34. Khoroshilova, N., Popescu, C., Munck, E., Beinert, H., Kiley, P. (1997). Iron-Sulfur cluster disassembly in the FNR protein of Escherichia coli by O₂: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity. *Proc. Natl. Acad. Sci.*,94, 6087-6092.
- Rudolf, J., Makrantoni, V., Ingledew, J., Stark, M., White, M. (2006). The DNA Repair Helicases XPD and FancJ Have Essential Iron-Sulfur Domains. *Molecular Cell*, 23, 801-808.
- Yeeles, J., Cammack, R., Dillingham, M. (2009). An Iron-Sulfur Cluster is Essential for the Binding of Broken DNA by AddAB-type Helicase-Nucleases. *The Journal of Biological Chemistry*, 284, 7746-7755.
- Ren, B., Duan, X., Ding, H. (2009). Redox Control of the DNA Damage-inducible Protein DinG Helicase Activity via its Iron-Sulfur Cluster. *The Journal of Biological Chemistry*, 284, 4829-4835.
- 38. Pugh, R., Honda, M., Leesley, H., Thomas, A., Lin, Y., Nilges, M., Cann, I., Spies, M. (2008). The Iron-containing Domain is Essential in Rad3 Helicases for Coupling of ATP Hydrolysis to DNA Translocation and for Targeting the Helicase to the Single-stranded DNA-Double-Stranded DNA Junction. *The Journal of Biological Chemistry*, 283, 1732-1743.
- Weiner, B., Huang, H., Dattilo, B., Nilges, M., Flanning, E., Chazin, W. (2007). An Iron-Sulfur Cluster in the C-terminal Domain of the p58 Subunit of Human DNA Primase. *The Journal of Biological Chemistry*, 282, 33444-33451.

- 40. Rabinowitz, JC. (1978). Analysis of acid-labile sulfide and sulfhydryl groups. *Methods Enzymol.*, 53, 275–277.
- 41. Hagen, Wilfred R.(2007) 'Wide zero field interaction distributions in the high-spin EPR of metalloproteins', *Molecular Physics*, 105: 15, 2031 2039
- 42. Blumberg, W. E., Peisach, J. (1974). On the interpretation of Electron Paramagnetic Resonance Spectra of Binuclear Iron-Sulfur Proteins. *Achives of Biochemistry and Biophysics*, 162, 502-512.
- 43. Guigliarelli, B., Bertrand, P., Gayda, J. (1986). Contribution of the fine structure terms to the g values of the biological Fe III Fe II clusters. *J. Chem. Phys.*, 85, 1689-1675.
- 44. Mulder, D. (2005). Fe4S4 clusters: A review of the electronic and geometric properties of synthetic analogues of Fe4S4 active sites in iron-sulfur proteins. (4Fe-4SclusterselectronicandgeometricREVIEW.pdf)
- 45. Noodleman, L. (1991). Exchange Coupling and Resonance Delocalization in Reduced [Fe4S4]⁺ and [Fe4Se4]⁺ clusters. 1. Basic Theory of Spin-State Energies and EPR and Hyperfine Properties. *Inorganic Chemistry*, 30, 246-256.
- 46. Abdallah, F., Chasteen, N. (2008). Spin concentration measurements of high-spin (g=4.3) rhombic iron(III) ions in biological samples: theory and application. *J. Biol. Inorg. Chem.*, 13, 15-24.
- 47. Marton, A., Sukosd-Rozlosnik, N., Vertes, A., Horvath, I. (1987). The effect of EDTA-Fe(III) complexes with different chemical structure on the lipid peroxidation in brain microsomes. *Biochemical and Biophysical Research Communications*, 145, 211-217.
- 48. Aisen, P., Aasa, R., Malmstrom, B., Vanngard, T. (1967). Bicarbonate and the Binding of Iron to Transferrin. *The Journal of Biological Chemistry*, 242, 2484-2490.
- Wegner, P., Bever, M., Schunemann, V., Trautwein, A., Schmidt, C., Bonisch, H., Gnida, M., Meyer-Klaucke, W. (2004). Iron-Sulfur Proteins Investigated by EPR, Mossbauer, and EXAFS Spectroscopy. *Hyperfine Interactions*, 156/157, 293-298.
- 50. Philip, C. V., Brooks, D. (1974). Iron(III) Chelate Complexes of Hydrogen Sulfide and Mercaptans in Aqueous Solution. *Inorganic Chemistry*, 13, 384-387.
- 51. Cooper, C., Salerno, J. (1992). Characterization of a Novel g=2.95 EPR Signal from the Binuclear Center of Mitochondrial Cytochrome c Oxidase. *The Journal of Biological Chemistry*, 267, 260-285.
- 52. Werth, M., Kurtz, Jr, D., Howes, B., Huynh, B. (1989). Observation of S=2 EPR Signals from Ferrous Iron-Thiolate Complexes. Relevance to Rubredoxin-Type Sites in Proteins. *Inorganic Chemistry*, 28, 1357-1361.

CHAPTER 2: Clustering of Alpers disease mutations and catalytic defects in biochemical variants reveal new features of molecular mechanism of the human mitochondrial replicase Pol γ.

Liliya Euro¹, Gregory A. Farnum², Eino Palin¹, Anu Suomalainen^{1,3} and Laurie S. Kaguni²

¹Research Program of Molecular Neurology, Biomedicum-Helsinki, University of Helsinki, Haartmaninkatu 8, 00290 Helsinki, Finland, ²Department of Biochemistry and Molecular Biology and Center for Mitochondrial Science and Medicine, Michigan State University, East Lansing, MI 48824-1319, USA, and ³Department of Neurology, Helsinki University Hospital, Helsinki, Finland
ABSTRACT

Mutations in Pol γ represent a major cause of human mitochondrial diseases, especially those affecting the nervous system in adults and in children. Recessive mutations in Pol γ represent nearly half of those reported to date, and they are nearly uniformly distributed along the length of the POLG1 gene (Human DNA Polymerase gamma Mutation Database); the majority of them are linked to the most severe form of POLG syndrome, Alpers–Huttenlocher syndrome. In this report, we assess the structure–function relationships for recessive disease mutations by reviewing existing biochemical data on site-directed mutagenesis of the human, Drosophila and yeast Pol γ s, and their homologs from the family A DNA polymerase group. We do so in the context of a molecular model of Pol γ in complex with primer–template DNA, which we have developed based upon the recently solved crystal structure of the apoenzyme form. We present evidence that recessive mutations cluster within five distinct functional modules in the catalytic core of Pol γ . Our results suggest that cluster prediction can be used as a diagnosis-supporting tool to evaluate the pathogenic role of new Pol γ variants.

INTRODUCTION

Mitochondrial DNA polymerase, Pol γ , is the sole known DNA polymerase in animal mitochondria and is responsible for mitochondrial DNA (mtDNA) replication and repair (1, 2). The human enzyme is a heterotrimer consisting of a catalytic subunit, Pol γ A, and dimer of an accessory subunit, Pol γ B (3). Pol γ A is a member of the family A DNA polymerase group to which bacterial DNA Pol I belongs (1). It comprises three domains: an N-terminal domain containing $3' \rightarrow 5'$ exonuclease (exo) activity, a spacer domain and a C-terminal domain, containing $5' \rightarrow 3'$ DNA polymerase (pol) activity in three subdomains termed the palm, fingers and thumb. Pol γ A also bears a 5'-deoxyribose phosphate lyase activity (4), but the location of its active site is unknown. The accessory subunit, Pol γ B, serves as a processivity factor, enhancing the DNA-binding affinity and catalytic activities of Pol γ A (5,6). The crystal structure of human Pol γ in its apoenzyme form was solved recently (PDB code 3IKM) (7).

Mutations in Pol γ represent a major cause of human mitochondrial diseases, especially those affecting the nervous system in adults and in children. Dominant mutations typically cause adult-onset myopathies and encephalopathies (8, 9), whereas recessive mutations result in severe adult or juvenile onset ataxia-epilepsy syndromes (MIRAS, SCA-E, SANDO) (10–12), or devastating early-childhood Alpers syndrome (Alpers–Huttenlocher syndrome) characterized by intractable epilepsy, psychomotor retardation and liver failure that leads to early death (13, 14). On the molecular level, POLG syndromes are accompanied by either tissue-specific mtDNA depletion, deletions or combination of both (15, 16). To date, more than 145 POLG1 disease mutations have been identified (Human DNA Polymerase gamma Mutation Database, http://tools.niehs.nih.gov/polg/), of which more than half are recessive. The functional consequences of dominant mutations, which affect primarily catalytic residues in the pol domain, have been explained by effective competition for the DNA substrate by the mutant Pol γ with the wild-type enzyme (7,17,18), and by site-specific stalling of mutant DNA polymerase (18). Studies that explain effects of individual recessive mutations on protein function are limited (19–22). The majority of the reported recessive mutations are linked to the most severe form of POLG syndrome, Alpers–Huttenlocher syndrome. Manifestation of Alpers disease typically requires the presence of at least two recessive, compound heterozygous mutations, and is accompanied primarily with mtDNA depletion (16, 23). The detailed mechanisms of pathogenesis and mtDNA depletion are unknown.

The assignment of a newly identified amino acid substitution as a pathogenic mutation is generally based on the absence of the variant in the normal population, conservation of the site among species and segregation in families. These assignments are particularly challenging because of the extensive variation in the amino acid sequences of Pol γ A in the normal population. To date, data accumulated on recessive disease mutations show that they are almost uniformly distributed among the three structural domains of the catalytic subunit and in most cases, it is unclear which property of the enzyme is affected, and how it contributes to mtDNA deletion or depletion.

The present study aims to assess the structure–function relationships for recessive disease mutations, by reviewing existing biochemical data on site-directed mutagenesis of the human, Drosophila and yeast Pol γ s, and their homologs from the family A DNA polymerase group; we do so in the context of the recently solved crystal structure of human Pol γ (7), onto which we have modeled primer–template DNA (ptDNA). We present evidence that recessive mutations cluster within five distinct functional modules in the catalytic core of Pol γ that we designate as 'Alpers Clusters 1–5'. We note that our analysis of the recessive mutations found in compound heterozygous form in Alpers patients reveals that a severe disease manifestation is typically caused by a combination of at least two mutations from different clusters. Our results suggest that cluster

prediction can be used as a diagnosis-supporting tool to evaluate the pathogenic role of new Pol γ variants.

Approach to comparative structural analysis

We docked ptDNA into the putative DNA-binding channel of the apoenzyme form of the human Pol γ structure (PDB code 3IKM) (7) by superposition of the closed ternary complex of T7 Pol bound to ptDNA and dNTP (PDB code 1T8E) (24). The coordinates of the resulting Pol γ : DNA model are provided in Supplementary Data of the online article. To ensure reliable positioning for subsequent structural analysis of Alpers recessive mutations, we evaluated three different alignments of the Pol γ and T7 palm domains that are presented in Figure 29 on page 91. First, we aligned the palm domain of Pol yA (PDB code 3IKM, chain A, residues 815–910 and 1095–1239) with that of T7 Pol (PDB code 1T8E, chain A, residues 409–487 and 611–704). Additionally, we aligned the two structures by superposition of their 12 β -loop-13 β motifs from the palm subdomain, encompassing residues G1127-E1144 in Pol yA and F646-E663 in the T7 catalytic core [for secondary structure element assignment see (25)]. This part of the central β sheet in the palm subdomain is the most structurally conserved element between DNA polymerases in the family A DNA polymerase group (25), and it superposed with an RMSD of 1.906 Å for Pol γA and T7. Alignment showed that the functionally significant secondary structure elements (SSEs) of the pol domain of Pol γA , including helixes αO , αL and αQ , as well as the loop corresponding to the 7β-loop-8β motif in T7 Pol, are tilted away from the active site, as compared to the analogous SSEs from T7 Pol and Klenow. This suggests conformational rearrangements in the Pol γ catalytic core upon DNA binding. To model the ptDNA in the putative DNA-binding channel of Pol γA , and to identify residues that interact with the DNA duplex, we performed a comparative analysis of the T7 Pol (24) and Klentaq (26) DNA complexes in both the open and closed states. We sought to identify SSEs in the pol domain that fulfill the requirements of interaction with DNA and preservation of conserved positions between the enzymes. As a result, the Q-helix and the 7 β -loop-8 β motif were identified, and a third alignment of Pol γ A and T7 Pol was based on superposition of their α Q helices (residues M1093-E1122 in Pol γ and P606-E635 in T7 Pol), and the hairpins formed by the 7 β -loop-8 β (residues T846-V855 in Pol γ and P422-T431 in T7 Pol). The RMSD for the superposed hairpins was 1.397 Å.

Clustering of Alpers disease mutations within the catalytic core of Pol γ

Recessive mutations are distributed along the length of the POLG1 gene sequence (http://tools.niehs.nih.gov/polg/), but they cluster within the tertiary structure of Pol γ A to distinct regions within the catalytic core (Figure 22 on page 62). Five functional modules, termed clusters 1–5, were assigned within Pol γ A as follows: Cluster 1 (in green) locates within the pol domain and comprises largely residues affecting DNA polymerase activity; Cluster 2 (in yellow) represents residues lining the upstream DNA-binding channel; Cluster 3 (in red) is associated with a novel structural motif in the fingers subdomain, which we propose confers partitioning of the DNA substrate between the pol and exo sites; Cluster 4 (in blue) lies on the periphery of the exo domain and mediates interactions with the distal Pol γ B, stabilizing the Pol γ : DNA complex; and Cluster 5 (in cyan) locates within the spacer domain and represents a region that we propose is involved in replisome interactions.



Figure 22. Alpers disease mutations cluster within functional modules in the catalytic subunit of Pol γ .





Figure 22 (cont'd) Upper panel: schematic diagram of the POLG1 gene showing the distribution of recessive Alpers disease mutations (Human DNA Polymerase gamma Database, http://tools.niehs.nih.gov/ polg/).AID and IP in the spacer domain refer to the accessory (subunit) interacting and intrinsic processivity subdomains, respectively, that are discussed in the text; NTD refers to the N-terminal domain. Lower panel: tertiary structural representation of the apoenzyme form of Pol γ [PDB code 3IKM, (7)] with modeled DNA, identifying the positions of five functional modules (shown in mesh) that are defined by clusters of amino acid residues (shown as spheres) affected by Alpers disease mutations as follows: Cluster 1, green; Cluster 2, yellow; Cluster 3, red; Cluster 4, blue; Cluster 5, cyan. The domains of Pol γ A are shown as surface representations, and in part as secondary structural elements (SSEs) that are colored as depicted in A. The proximal and distal accessory subunits are shown as surface representations in light and dark gray, respectively. Primer–template DNA was docked as described in the text and is displayed as orange ribbons.

RESULTS

ALPERS CLUSTER 1: Residues affecting 5'-3' DNA polymerase activity

Catalysis of template-directed nucleotide polymerization in DNA polymerases proceeds via a phosphoryl transfer reaction assisted by two Mg^{2+} ions (27, 28). Incorporation of the correct dNMP is mediated by a nucleotide-induced conformational shift of the fingers subdomain between an open and closed state; the closed state results in phosphoryl transfer when correct Watson–Crick base pairs between the incoming nucleotide and a templating base can fit stably in the active site (29, 30). Similarities in the overall mechanism of DNA polymerases suggest a conserved structure of their active sites and provide a basis for comparative structural analysis and extrapolation of site-directed mutagenesis data from one DNA polymerase to its close homologs. The amino acid residues that we attribute to Alpers Cluster 1 contribute to nucleotide polymerization per se, to shaping the overall architecture of the pol active site, or to positioning the primer template within the active site (Figure 23A on page 72).

Catalytic residues

The largest group of recessive Alpers mutations map to the pol domain in POLG1 (Figure 22 on page 62), and a number of the affected amino acids are likely involved in catalysis by Pol γ . Residues E1136 and K1191 surround the highly conserved residues D890 and D1135 (Figure 23B on page 72), which locate at the base of the pol active site and have been shown to coordinate the two magnesium ions required for catalysis (7). *In vitro* mutagenesis of residue E883 in Klenow (E1136 in Pol γ) resulted in decreased pol rate, k_{cat} and affinity for dNTPs and DNA (31, 32), which fits with the recessive nature of this mutation and its effect on mtDNA copy number in vivo (16, 23). Comparative structural analysis of Pol γ and its close homologs revealed that E895 likely confers discrimination against ribonucleotides, and it is involved in aligning the incoming dNTP

within the catalytic site, together with the catalytic residues Y955 and Y951 [Figure 23C on page 72 (33)]. Interestingly, E895 also occupies part of the dNTP-binding pocket, and whereas E895G is a recessive mutation, Y955C is dominant. In vitro mutagenesis of Pol γ showed that the Y955C variant exhibited dramatically reduced catalytic activity, but had the same affinity for DNA as the wild-type enzyme (33). In a disease situation, these effects cannot be compensated for by the wildtype allele and as a result, Y955C causes a dominantly inherited adult-onset mitochondrial myopathy (autosomal dominant progressive external ophthalmoplegia) associated with mtDNA deletions (34). E1136 is part of one of three conserved sequence motifs in the palm subdomain of family A DNA polymerases (the PolC motif, Figure 23B on page 72). Its counterparts E655 in the T7 Pol (25) (Figure 23B on page 72) and E786 in the Klentaq (26) structures are not in contact with primer or incoming dNTP, but mutagenesis of Klenow showed that this residue is critical for catalysis (31). The ternary complex of T7 Pol reveals an electrostatic interaction between E655 and H704, and suggests that the role of E655 is to position H704 in close proximity to the phosphate moiety of the 3' nt in the primer strand, and it was shown to be essential for catalysis (25). K1191 of Pol γ occupies the equivalent position of H704 in T7 Pol, suggesting that it would interact with E1136 in a similar manner upon binding of the primer template to facilitate catalysis.

In contrast to the rigid pocket formed by the palm active site residues, the opposite half of the active site is highly flexible to allow a fast transition between open and closed conformations of the fingers subdomain. This conformational asymmetry provides the structural basis for the high affinity of DNA polymerases for the primer–template junction: the template must be bound firmly and consistently each catalytic cycle to assure the fingers sufficient freedom to examine each incoming nucleotide. Several Alpers mutations have been reported in this region of the fingers, scattered mainly along the O-helix (Figure 23C on page 72). Biochemical studies on the A957S mutant showed mild defects in k_{cat} and dNTP binding, but a modest increase in DNA-binding affinity (33). A957 locates in the loop between helices O and O1, within one of the three hinge regions, the 'GAG' hinge (residues G956/A957/G958), which in related DNA polymerases is known to confer the flexibility necessary for closing and opening of helixes O and O1 (27,35). Glycine to alanine substitution in the GAG hinge in T7 Pol resulted in complete loss of catalytic activity, which can be attributed to complete or diminished ability to shift between the open and closed states (36). Mutations affecting the hinge region likely introduce steric clashes that would limit the extent of conformational shifts. Interestingly, different amino acid replacements for A957 in Pol γ do not have similar consequences: A957S causes dominantly inherited adult-onset myopathy with multiple mtDNA deletions, whereas an A957P substitution has been reported only as a recessive, compound heterozygous mutation in patients with Alpers syndrome and mtDNA depletion. A957S showed an increased affinity for DNA (33), a consequence that is consistent with a slower open-close rate, and could potentially cause pol stalling or dissociation. In contrast, the severe defect caused by A957P suggests that the presence of proline in the hinge abolishes the flexibility between helixes O and O1, a defect that we predict will result in decreased affinity for both dNTP and DNA and may prohibit the formation of a productive ternary complex. The loop between helixes O and O1 has been shown to form the pre-insertion site in Geobacillus stearothermophilus Klenow to accommodate the ssDNA template and prevent its premature entry into the active site (35). However, the available library of DNA polymerase structures reveals variability in the pre-insertion site, making it difficult to assign a critical role. Residue K947, which is associated with Alpers syndrome, is located on the O-helix adjacent to R943. The corresponding residues in T7 Pol (K522 and R518, respectively) contact the triphosphate moiety of the incoming nucleotide in the closed ternary complex (PDB code 1T8E). These electrostatic interactions

provide most of the driving force for O-helix movement, and mutations in these positions will weaken interactions.

Architectural residues

Optimal catalysis requires not only the residues interacting directly with dNTP and primer template, but also those that maintain the architecture of the pol active site. We suggest that the pathogenic role of the Alpers mutations Q879H, T885S, L886P, G888S, E1143 and D1184N can be attributed to modifications in the local β -sheet architecture surrounding the catalytic residues in the palm (Figure 23A, upper panel on page 72). Defects are predicted to reduce pol rate, and it appears that the closer the mutated architectural residue is to the catalytic site, the more pronounced its effect. In support of this, in vitro mutagenesis of Q879 and T885, located a distance of ~15 Å from the catalytic residues, caused only a 2-fold decrease in pol rate (20). In contrast, mutagenesis of E883 located at the beginning of the strand β 13 within the catalytic site in Klenow (E1136 in Pol γ A) reduced k_{cat} 26-fold (31).

The Alpers mutation D930N (Figure 23C on page 72) was studied in vivo in yeast, and led to complete loss of mtDNA (37). The corresponding residue in T7 Pol (D504) interacts with triphosphate-binding residue R518 (R943 in Pol γ) in the closed conformation (PDB code 1T8E), and therefore contributes indirectly to correct dNTP binding. Residues T914 and L966 located in the fingers subdomain (Figure 23C on page 72) are conserved only among Pol γ s, and are linked to Alpers syndrome. Our structural analysis predicts that mutations in these residues will likely compromise the stability and conformation, and/ or the transition rate between the open and closed states of the fingers. In sum, we propose that the primary effect of mutations surrounding the O-helix residues is altered affinity for dNTPs, which decreases pol rate.

Residues conferring a high affinity of the pol site for the primer-template junction.

Docking of DNA onto the Pol γ structure shows that the DNA-binding channel of the enzyme encompasses ~20 bp of DNA duplex, whereas those in T7 Pol and Klentag are relatively short and interact with ~8 bp of DNA duplex (24–26). In Pol γ , similar to other family A DNA polymerases, the DNA-binding channel starts with residues located in the β 7-loop- β 8 motif in the palm, which form extensive contacts with the first four base pairs in the ptDNA. In Pol γA , this fragment encompasses residues 842–856 and appears as a loop in the crystal structure with the current resolution of 3.24 Å (7). Because the β 7-loop- β 8 motif has two arginines in its tip, R852 and R853, the former of which is conserved only among Pol γ s, we designated it as the 'RR loop' (Figure 23D on page 72). Several recessive mutations affect the RR loop: G848S, T849H, T851A, R852C and R853Q, as well as the surrounding residues H1134R and H1110Y. In vitro mutagenesis of G848, T851, R852 and R853 in human Pol γ (20), and R668 in Klenow ((31, 32), corresponding to R853 in Pol γ), caused a dramatic decrease in catalytic activity and DNA-binding affinity. Polesky et al. (32) highlighted the complex effect of these mutations, noting that a specific amino acid can bind a critical part of the substrate, contributing to the overall affinity of the enzyme for ptDNA, and to catalysis by positioning the substrate in the active site. Comparative structural analysis of Pol γ homologs solved in ternary complexes with bound DNA prompts us to propose that the primary function of the RR loop is in binding and alignment of ptDNA within the DNAbinding channel. H1134 locates adjacent to the RR loop in the tertiary structure (Figure 23D on page 72). The corresponding histidines in Klenow (31, 32), T7 Pol (25) and Klentaq (26) were shown to be essential for DNA binding, and for coordination of the first nucleotide in the primer strand. Analysis of Pol γ suggests that Q1102 from the Q-helix (Q615 in T7 Pol) and R853 (R429) in T7 Pol) from the RR loop coordinate the first base pair in the ptDNA, and therefore may be crucial in sensing a mispair. T851 (T427 in T7 Pol) likely participates indirectly by positioning

Q1102. T849 is predicted to coordinate the DNA duplex by interaction with the phosphate moiety of the third nucleotide from the 3'-end of the primer. In the Pol γ apoenzyme structure, the side chains of the residues from the RR loop are oriented differently than the orthologous residues in the T7 Pol ternary complex, which suggests that DNA and dNTP binding affect the positions of the residues that coordinate them. For example, R853 in the Pol γ apoenzyme interacts with the metal-binding residue D1135 as was noted by Stumpf and Copeland (17) and with E895, but in complex with DNA, R853 would shift to interact with the base of the 3' nt of the primer strand, as in T7 Pol (24–26). A distinguishing feature of the RR loop, in comparison with the orthologous β 7-loop- β 8 motif in other family A DNA polymerase group members, is the electrostatic interaction between R852 and D1107, which may confer additional stability to the RR loop in Pol γ (Figure 23D on page 72). Disease mutations in this domain affect primarily DNA binding and positioning in the active site channel, manifesting defects in both DNA-binding affinity and in catalytic efficiency.



Figure 23. Alpers Cluster 1 mutations affect the 5 DNA polymerase activity of Pol γ.

Figure 23 (cont'd)





Amino acid residues affected by Alpers Cluster 1 mutations in POLG1 are shown as green spheres. Other Pol γ residues that are discussed in the text are shown in brown, and T7 residues are shown in red. Pol domain SSEs are shown in pink according to the schematic shown in Figure 22 on page 62; ptDNA is indicated by orange (template) and brown (primer) strands. The incoming dNTP is shown in blue. Mg²⁺ ions are shown as small gray spheres. A, upper panel, overview of the positions of Alpers Cluster 1 mutations with dashed black lines indicating the regions described in the text and depicted in B–D; A, lower panel, overview of the positions of Alpers Cluster 1 mutations relative to ptDNA and incoming dNTP; B–D, positioning of Pol γ residues in the apoenzyme form [PDB code 3IKM (7)] and T7 Pol residues in its closed ternary complex [PDB code 1T8E (25)] relative to ptDNA and incoming dNTP, with the dashed arrow showing the expected movements of the Pol γ residues upon formation of a closed complex: B, Alpers mutations surrounding the Mg²⁺ -binding residues; C, Alpers mutations affecting O-helix movement and dNTP binding, with the T7 Pol O-helix in its closed conformation superimposed in red; D, Alpers mutations affecting the RR loop and the surrounding residues.

CLUSTER 2: Recessive mutations affecting the upstream DNA-binding channel

Processivity of a DNA polymerase depends both on DNA-binding affinity and on the rate of DNA polymerization. Structural analysis shows that the mutations affecting interaction of Pol γ with the upstream DNA duplex are located within the spacer domain (Figure 22 on page 62). This would argue that the AID [accessory (subunit)-interacting domain] and IP (intrinsic processivity) subdomains adopt a different conformation in Pol y: DNA complexes. To evaluate the potential impact of recessive spacer region mutations on DNA binding, we considered results of in vitro mutagenesis of spacer domain in Drosophila Pol γ (38) and in Saccharomyces cerevisiae Mip1 (22) (Figure 24 on page 74). Analysis of the Pol γ structure with modeled ptDNA reveals that residues K755 from the IP subdomain and Q497 in the K-tract of the AID subdomain interact with upstream DNA. In the human Pol γ : DNA model, the L752P and A767D substitutions would affect the structure in the tip of IP subdomain, which contacts the minor groove of the DNA duplex in a similar way as reported for T7 Pol [residues T354-V363 (25)]. In the Pol γ : DNA complex, L752 and A767 neighbor the K768/D769/F770 triplet from fly Pol γ ; triple alanine substitution of KDF in fly Pol γ resulted in a 1.4-fold decrease in DNA-binding affinity (38). We predict that the Alpers A767D mutation likely causes similar consequences for DNA binding. Mutations of P587 and P589 in the spacer domain are also linked to Alpers disease. These residues constrain the β hairpin loop between the IP and AID subdomains and are positioned close to the Y452/E453/E454 triplet that lies within the thumb [corresponding to the YED triplet in Drosophila Pol γ (38)]. We postulate that the P587L and P589L mutations may suffer the same consequences as for the triple alanine substitution of YED in the fly enzyme, which caused reduced processivity, most likely as a result of misalignment of the ptDNA, with respect to the pol catalytic site (38). Furthermore, a triple alanine mutant of P556A/K557A/L558A in the fly was shown to be completely deficient in DNA binding and pol activity (38); this defect is likely due to disruption of the hydrophobic structure of the spacer domain, which shapes the DNA-binding channel wall. Notably, the residue affected in the common A467T human disease allele is also part of this PKL hydrophobic center, and its biochemical properties most likely derive from similar structural alterations that interrupt hydrophobicity with the introduction of a hydroxyl group. Whereas we and others have found that

the A467T catalytic core alone exhibits highly reduced DNA-binding affinity and pol activity (19, 21), we showed that these defects are mitigated partially by its association with the accessory subunit (21). This likely results from partial stabilization of the overall conformation of mutant catalytic core within the reconstituted Pol γ holoenzyme.



Figure 24. Alpers Cluster 2 mutations affect the upstream DNA-binding channel of Pol γ . Amino acid residues affected by Alpers Cluster mutations in POLG1 are shown as yellow spheres. Other Pol residues that are discussed in the text are shown in brown. Spacer domain SSEs are shown in magenta and pol domain SSEs are in pink according to the schematic shown in Figure 22 on page 62; ptDNA is indicated by orange (template) and brown (primer) strands. The spacer domain is also shown as a transparent surface representation in pale gray and the exo domain is shown in purple.

CLUSTER 3: Mutations associated with a novel, Pol γ-specific functional module, conferring partitioning of DNA substrate between the pol and exo active sites

Comparative analysis of the pol domains from human Pol γ , T7 Pol and T7 RNA Pol reveals a remarkable structural similarity in their overall folds. In fact, the only significant difference between the analogous pol domains lies in the region that connects the P-helix of the

fingers subdomain and Q-helix of the palm (see Figure 30 in appendix B on page 92), which contains a novel module (Figure 25 on page 78) whose amino acid sequence is highly conserved in Pol γ s from yeast to man. This, and the finding that several recessive Alpers mutations cluster in this region (Figure 22 on page 62), argue that it serves an important role. Indeed, in our Pol γ : DNA model, this module extends into the DNA-binding channel in a position very near the pol active site (<10 Å from the Mg²⁺-binding catalytic residues). In bacteriophage N4 RNA polymerase (PDB code 2PO4) (39) and in T7 RNA polymerase [PDB code 1ARO (40); Figure 30 on page 92], the fragment connecting the fingers with the helix corresponding to the Q-helix in Pol γ participates in specific recognition of the transcriptional promoter; it is aptly termed a specificity loop (39–41). In Pol γ , we have adopted the term 'partitioning loop' based upon the functional role we propose it serves. We discuss our rationale below, and use it to establish structure–function justifications of disease manifestation for the Alpers mutations we assign to Cluster 3.

We propose that the partitioning loop modulates the partitioning of the primer strand between the pol and exo active sites by forming stable contacts with correctly base-paired ptDNA, and destabilizing ptDNA that contains mispairs or lesions. Comprising residues 1050–1095, the atomic structure reveals that the first 10 residues of the partitioning loop adopt an α -helical fold that extends from the fingers subdomain directly into the DNA-binding channel. The beginning of this helix contains a W1049XGG1052 motif, which is strictly conserved in Pol γ , and a similar motif is also present in T7 Pol (W579XAG582). Our Pol γ : DNA model shows that the WXXG motif maps to the same region of the fingers domain in both Pols. In T7 Pol, it is responsible for binding downstream template ssDNA via base-stacking interactions between W579 and the nucleotide base [PDB code 1T8E (25)], and likely performs an equivalent function in Pol γ . The novel functionality that the partitioning loop contributes to Pol γ is achieved by its unusual loophairpin component spanning residues 1060–1074 that appears to grip the modeled ptDNA along the major groove as shown in Figure 25A on page 78. We propose that it uses a steric exclusion mechanism that confers exquisite specificity by associating closely with the major groove of primer template, such that correctly base-paired substrates are bound stably. In contrast, DNA lesions and mispairs that adopt an altered helical structure clash sterically with the partitioning loop, resulting in exclusion of the primer template from the pol active site. Excluding primer template may facilitate fraying of the duplex, generating ssDNA primer ends that can be bound by and hydrolyzed in the exo active site (Figure 25B on page 78).

Mutations in the residues composing the partitioning loop, such as the Alpers mutations R1047W and P1073L, as well as the progressive external ophthalmoplegia mutations G1051R and G1076V, have been studied in vivo. Yeast strains homozygous for G1051R Pol γ exhibited a point mutational frequency >10-fold higher than the wild-type strain, and heterozygous strains showed frequencies >2-fold higher relative to homozygous wild-type strains (42). Increased point mutation frequencies were also reported for yeast strains heterozygous and homozygous for the Alpers mutation P1073L and the G1076V mutation associated with progressive external ophthalmoplegia (22, 43). Both the P1073L and G1076V mutations cause substitutions in strictly conserved residues located in the loop region that is expected to contact DNA, and the G1051R mutation affects a strictly conserved glycine in the WXXG motif. Interestingly, yeast strains heterozygous and homozygous for L304R, S305R, Q308H, R309L and R309H showed similar increases in point mutation frequency as well as mtDNA depletion (43). These exo domain residues are located in a loop-helix motif adjacent to the partitioning loop, and have been described previously as the orienter module (22) (Figure 25 on page 78). When S305R and P1073L were present as compound

heterozygous, the point mutation frequency increased drastically to >70-fold that of the wild-type strain (44). Non-complementation of these two mutations suggests that both are involved in the same function, and we propose that the role of the orienter module is to position correctly the partitioning loop. L304R, R309H, R309L and W312R were analyzed biochemically by producing and purifying homologous yeast Pol γ variants (22). All variants exhibited reduced DNA-binding affinity and reduced pol activity and in addition, the L304R variant showed a significant increase in exo activity (22). An identical biochemical phenotype was reported previously in the fly Pol γ SYW triple alanine variant, which correspond to thumb residues S799/F800/W801 in human Pol γ (38). As illustrated in Figure 25 on page 78, the SYW residues are located on the face of the thumb subdomain directly across the DNA-binding channel from the orienter module, and we predict that this region of the thumb plays an equivalent role to the orienter module. Therefore, we define Cluster 3 as mutations affecting residues of the partitioning loop, as well as nearby residues that govern the position and conformation of the partitioning loop. Mutations in Cluster 3 will alter the balance between pol and exo activity and diminish the fidelity of the polymerase. We note the possibility that enhanced fidelity may also be experimentally feasible, introducing the potential to engineer fine-tuned anti-mutator replicase.



Figure 25. Alpers Cluster 3 mutations are associated with a novel Pol γ -specific functional module proposed to be involved in primer strand partitioning between the pol and exo sites. Amino acid residues affected by Alpers Cluster 3 mutations in POLG1 are shown as red spheres/ mesh adjacent to a novel alpha helix with an associated loop–hairpin (the partitioning loop), also shown in red. The brown spheres and mesh represent the SYW (fly)/ SFW (human) and surrounding residues, respectively, that are described in the text. The pol domain is shown as a surface representation in pink and the exo domain is shown in purple, according to the schematic shown in Figure 22 on page 62; ptDNA is indicated by orange (template) and brown

Figure 25 (cont'd) (primer) strands. (A) The predicted position of the partitioning loop relative to ptDNA in the pol mode, and (B) represents the exo mode. To dock the frayed ptDNA in the exo active site, the exo domain residues 324-518 of Klenow (PDB code 1KLN) were aligned with the exo domain residues 170-440 of Pol γ (PDB code 3IKM) (see appendix B for Figure 31 on page 94).

In sum, our justification for the proposed role of the partitioning loop is based on its sequence conservation among Pol γ s from yeast to man, its positioning relative to the specificity loop found in T7 RNA Pol, the clustering of several recessive Alpers mutations in this region, its absence in any other Pol family and the documented high fidelity of Pol γ . We argue that an alternate hypothesis that this structural element would simply rotate away upon DNA binding (7) seems unlikely based on the above considerations and, in particular, on the evolutionary conservation of a novel structural module that is positioned strategically in the DNA-binding channel very near the pol active site. Clearly, validation of either hypothesis warrants future experimentation.

CLUSTER 4: Mutations affecting Pol γA interactions with the distal Pol γB upon DNA binding by Pol γ holoenzyme

Pol γ A contains independent 5'-3' DNA polymerase and 3'-5' exonuclease active sites whose functions are coordinated in proofreading DNA synthesis. The human Pol γ crystal structure showed that its homodimeric Pol γ B interacts asymmetrically with Pol γ A, such that one Pol γ B protomer forms the dominant subunit interface and is designated as the proximal accessory subunit, while the other protomer makes very limited contact with Pol γ A and is designated as the distal accessory subunit (7). In fact in the apoezyme structure, the Pol γ A interaction with the distal Pol γ B is mediated through one ion bond between Pol γ A R232 and Pol γ B E394 [(7) and Figure 26 on page 81]. A subsequent biochemical study on the interaction between the catalytic and accessory subunits showed that the proximal Pol γ B increases DNA-binding affinity of the holoenzyme, while the distal protomer enhances the polymerization rate of the holoenzyme (45). Again as suggested in the section on Cluster 2, we posit that the holoenzyme likely undergoes substantial conformational rearrangements in the IP and AID subdomains of the spacer upon formation of an active complex with DNA.

Mapping of Alpers mutations within the crystal structure revealed that the majority of the mutations in the exo domain cluster on the protein surface within 13–17 Å from Pol yA residue R232 (Cluster 4 shown in blue in Figure 22 on page 62 and Figure 26 on page 81). The functional role of this region can be deduced from biochemical studies on mutations of the neighboring R232 (46) and L244 residues (22, 43). Mutations of R232 were shown to decrease pol activity, DNAbinding affinity and processivity of the holoenzyme yet at the same time, to enhance its exonuclease activity, which was also rendered less selective for mismatches (46). These data argue that in the Pol γ : DNA complex, and/or in its ternary complex with dNTP, the distal accessory subunit associates more tightly with the catalytic subunit, thereby enhancing binding of the upstream DNA in the pol mode, and limiting the rate of translocation of the frayed 3'-end of the primer to the exo site (46). Accordingly, an orthologous variant of the L244P human mutation in yeast caused increased mutation frequency (22, 43). Previous site-directed mutagenesis of human Pol γB also suggested extensive interaction between the catalytic and distal accessory subunits within the ternary complex (5). Analysis of the Pol γ appendix structure shows that E445 and T447 of Pol yB are located at the tip of the its C-terminal region, and that only in the distal subunit are these residues oriented toward Cluster 4 on the edge of exo domain. A double mutation of these residues (E445A/ T447A) led to a dramatic decrease of the stimulatory effect of Pol γB on the wild-type Pol γA , and a decrease in DNA binding (5). We propose that residues in Cluster 4 of Pol γ A contribute to tight interaction between exo domain of catalytic subunit and the distal accessory subunit, resulting in more extensive enclosure of the primer template in the pol mode. We also

predict the participation of the AID of the spacer domain in these conformational changes, because it is associated intimately with the accessory subunit. In support of this, biochemical studies performed on the yeast Pol γ core variant of Alpers mutation R574, which is located in the AID subdomain, showed not only reduced DNA-binding affinity, but also reduced exo activity and a severe processivity defect (22, 43). To some extent, these interactions would provide structural constraints on the upstream DNA in the direction toward the partitioning loop. Therefore, the partitioning loop, orienter module and thumb in Pol γ A, and the interface between its exo domain and the distal accessory subunit might function in a concerted manner to affect switching between the pol and exo modes in Pol γ function.



Figure 26. Alpers Cluster 4 mutations affect Pol γ A interactions with the distal Pol γ B upon DNA binding by Pol γ holoenzyme. Amino acid residues affected by Alpers Cluster 4 mutations in POLG1 are shown as blue spheres/ mesh and are located largely within the exo

Figure 26 (cont'd) domain (shown in purple). The helix shown in magenta is the AID (accessory interacting domain) of the spacer region that interacts with the proximal accessory subunit. Other domains are colored according to the schematic shown in Figure 22 on page 62; duplex DNA is indicated by the orange strand. Orange spheres indicate the positions of the accessory subunit residues described in the text. The proximal and distal accessory subunits are shown as surface representations in light and dark gray, respectively.

CLUSTER 5: Mutations affecting a region of the IP subdomain that is likely involved in replisome contacts.

Many reported Alpers mutations (L623W, R627W/Q, P648R, G737R, G746S, W748S and F749S) map to the distal surface of the IP subdomain (Figure 27 on page 83). We considered these as distinct from the Cluster 2 residues because they are located much further from the DNAbinding channel. In addition, in vitro studies of W748S and R627W/Q variants showed no defects in pol activity, processivity or DNA-binding affinity (47, 48). In contrast, biochemical defects were demonstrated earlier in single alanine variants of fly Pol γ within this region of the spacer: a W576A variant was nearly inactive, F578A retained half of wild-type activity, G575A displayed wild-type activity and all three variants showed substantially reduced stimulation by mtSSB (38). The fly G575/W576/F578 residues correspond to residues G619/W620/Y622 in human Pol γ. In the human Pol γ apoenzyme structure, W620 is buried in the hydrophobic core of the IP subdomain, whereas Y622 and G619 are located closer to the surface (7). The combined biochemical data suggest that residues closer to the distal surface of the IP domain are not critical for catalytic function. Cluster 5 mutations map to this surface, and therefore a catalytic defect per se is not expected to be the cause of Alpers manifestation. Rather, we propose that this region is involved in protein-protein contacts, a likely partner being mtSSB as suggested by the GWF fly Pol γ variants (38). Another candidate may be the mitochondrial replicative DNA helicase mtDNA helicase, but such an interaction has not been investigated to date.



Figure 27. Alpers Cluster 5 mutations are proposed to affect replisome interactions. Amino acid residues affected by Alpers Cluster 5 mutations in POLG1 are shown as cyan spheres. Spacer domain SSEs within the IP (intrinsic processivity subdomain) are shown in magenta; ptDNA is indicated by orange (template) and brown (primer) strands. The spacer domain is also shown as a transparent surface representation in pale gray and the exo domain is shown in purple. Brown spheres indicate the positions of the IP residues described in the text.

Cluster	Structural location	Predicted Primary Biochemical Defect	Predicted Phenotype (primary, secondary)	Causes Alpers when found in <i>trans</i> with cluster
1	Pol active site and environs	Decreased pol rate	Reduced pol rate Reduced DNA binding Reduced processivity	2, 3, 5
2	Upstream DNA-binding channel	Decreased DNA binding affinity	Reduced DNA-binding Reduced pol rate Reduced processivity	1, 3, 4, or 5
3	Partitioning loop	Partitioning of primer strand between pol and exo active sites	Altered exo/ pol ratio Altered error rate Altered exo rate Reduced DNA-binding Reduced processivity	1, 2, 5
4	Exo-IP interface with POLGB	Stabilization of ternary complex	Reduced pol rate enhancement by Pol γB Increased error rate Reduced processivity	2 or 5
5	Periphery of IP domain	Protein–protein interactions?	Reduced SSB stimulation Deficient protein-protein interactions?	1, 2, 3, 4,

 Table 5. Structural and functional features of the five proposed Alpers Clusters.

PROSPECTS

Toward a diagnostic tool to assess new human mitochondrial disease mutations

We show that although the reported Alpers mutations scatter across the entire POLG1 gene, they cluster to distinct functional regions in the Pol γ tertiary structure. Table 5 on page 84 summarizes the structure-function predictions for each cluster and also displays the cluster combinations that have been reported in compound heterozygous patients with Alpers syndrome. These combinations do not occur randomly; rather, multiple occurrences of specific cluster combinations is evident whereas others are absent, suggesting that the latter combinations are not tolerated. Cluster 1 mutations locate in the pol active site region and will invariably cause reduced pol activity, potentially combined with reduced DNA binding. Cluster 2 mutations locate in the upstream DNA-binding channel, resulting in reduced DNA-binding affinity, but are too distant from the catalytic site to affect directly pol activity to the extent of Cluster 1 mutations. Therefore, Cluster 2 mutations will be recessive, whereas Cluster 1 mutations may be either dominant or recessive, depending on the severity of the biochemical defect. Cluster 3 mutations locate to the partitioning loop or its environs, such as the orienter module, and we predict that they will alter the balance of polymerase and exonuclease activities of mutant Pol γ . Cluster 4 mutations locate on the surface of the exo domain along the interface of the distal accessory subunit. These mutations are predicted to reduce the stimulation affected by the distal accessory subunit, which has been documented to enhance pol rate and to reduce exo activity (45). Both Clusters 3 and 4 mutations will likely cause increased mutagenesis in vivo, a phenotype that has been observed in yeast models to be associated with amino acid alterations within these clusters (43). Cluster 5 mutations locate on the distal surface of the intrinsic processivity (IP) subdomain, removed from the pol active site, DNA-binding channel and accessory subunit interface. Their distant location and lack of biochemical defects prompt us to speculate that this region is involved in replisome

contacts. This hypothesis, in particular, highlights the need for future experimentation on both the physical and functional interactions among the key proteins at the mtDNA replication fork.

POLG1 shows significant polymorphic variation in the human population, and a constant challenge in DNA diagnosis is to distinguish pathogenic mutations from neutral variants. We suggest that the remarkable clustering of Alpers mutations into specific functional regions (Figure 22 on page 62) enables the use of our Pol γ : DNA model as a diagnosis-supporting tool for evaluating pathogenic potential of new sequence variants. For example, Q1102 is located within the Cluster 1 region, and mutations affecting this amino acid can be predicted to cause catalytic defects. This prediction is consistent with biochemical studies on the Q1102 counterpart in bacteriophage T7 Pol (Q615), which was shown to be essential for the fidelity of nucleotide incorporation and for ptDNA binding (25). Functional predictions can be made in the absence of biochemical data as well; for example, the tam⁹ mutant allele, carrying an E595V mutation in the catalytic subunit of Pol γ , causes mtDNA depletion in the fly (49, 50). E595V corresponds to E641V in human Pol γ , which maps to Cluster 5 (Figure 27 on page 83). Our model predicts that the tam⁹-associated mtDNA depletion is caused by deficient replisome interactions. Consequently, we would predict that both Q1102 and E641 are potential candidates for causing recessive POLG syndrome with mtDNA depletion.

Predicting the consequences of compound heterozygosity via Alpers cluster analysis

Alpers syndrome is typically caused by compound heterozygosity of two mutations occurring along the entire length of the POLG1 gene. In contrast, homozygosity for a single mutation often is associated with a somewhat milder disorder (2, 51). Utilizing the Human DNA Polymerase γ Database, we compiled data to show all combinations of mutations that have been reported to trigger Alpers disease (Figure 28 on page 88). Alpers patients typically do not show a

combination of two mutations from the same cluster. In addition, Cluster 4 mutations only manifest as Alpers disease when combined with Clusters 2 or 5 mutations. These trends provide merit for our structure-guided assignment of the mutations into the five proposed clusters, and suggest unique functional roles inherent to each. With the rapid development of massive parallel sequencing methodologies, the number of new Pol γ A variants is likely to increase significantly, emphasizing the importance of bioinformatic tools to evaluate the pathogenic role of identified variants. Furthermore, the identification of synergistic functional relationships between the characterized clusters provides novel insight into the mechanism of Alpers disease manifestation, and will serve as a framework for future structure–function studies of the mitochondrial DNA replicase.



Figure 28. Combinations of mutations found in Alpers patients. Individual Alpers mutations are grouped by cluster as shown in Figure 22 on page 62, and black blocks represent Alpers

Figure 28 (cont'd) manifesting combinations. The inset in the lower right presents a simplified version of the table by reducing the axes to the five clusters only, where gray blocks represent cluster combinations that have not been found in Alpers patients. The tabulated data suggest that two mutations from the same cluster do not typically manifest as early-onset Alpers disease. Furthermore, the data indicate that Cluster 4 (blue) mutations manifest as Alpers disease only when in combination with Cluster 2 (yellow) or Cluster 5 (cyan) mutations. These trends support the existence of unique functional relationships in Pol γ A that are inherent to each cluster.

APPENDIX

APPENDIX



Figure 29. Alignments used for docking DNA onto the *Hs* Pol γ holoenzyme structure (**3IKM**). Panels display the palm subdomain of the T7 Pol (1T8E) closed ternary complex (red) superimposed on the palm subdomain of Pol γ (pink). Top panel, alignment of the palm of T7 Pol (residues 409-487, 611-704) to the palm of Pol γ (815-910, 1095-1239); middle panel, alignment of T7 Pol residues 646-663 (green) to Pol γ residues 1127-1144; lower panel, alignment of T7 Pol residues 606-635 and 422-431 (green) to Pol γ residues 1093-1122 (Q-helix) and 846-855 (RR-loop). Alignments were performed in Pymol.


Figure 30. Comparative alignments of *Hs* Pol γ , T7 Pol and T7 RNA Pol illustrating the structural variations in the region between the P-helix and the Q-helix in the pol domain. Panels on the left display the complete pol domain in complex with DNA, and those at right

Figure 30 (cont'd) show a close up view of the of the region between the Q-helix and P-helix. Top panel, the Pol γ (3IKM) pol domain is shown as pink cartoon and its partitioning loop in red, with its disordered region indicated as a dashed red line. Transparent surfaces of the exo (purple) and pol (pink) domains are also shown at left. DNA docked by alignment with T7 Pol (see top panel of Fig. 29 on page 91) is shown in orange; middle panel, the Pol domain of the T7 RNA Pol elongation complex (1H38) is shown as light blue cartoon and its specificity loop in red; lower panel, the T7 Pol (1T8E) pol domain is shown as blue cartoon, and transparent surfaces of its exo (purple) and pol (blue) domains are shown to highlight a similar architecture as in Pol γ . WXGG and WXAG are the amino acid sequence motifs in Pol γ and T7 Pol, respectively, which are discussed in the text.



Figure 31. Structural alignment of the exo domain of Klenow editing complex (PDB code 1KLN, residues 324-518, displayed as blue cartoon) with the exo domain of human Pol γ (PDB code 3IKM, residues 170-440, displayed as purple cartoon). This alignment was used to predict the editing complex of Pol γ by docking of the frayed primer template (primer strand in chocolate, template strand in orange) onto the apoholoenzyme structure, which we display in Figure 25B on page 78.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Kaguni, L.S. (2004) DNA polymerase gamma, the mitochondrial replicase. *Ann Rev Biochem*, 73, 293-320.

2. Graziewicz, M.A., Longley, M.J. and Copeland, W.C. (2006) DNA polymerase gamma in mitochondrial DNA replication and repair. *Chem Rev*, 106, 383-405.

3. Yakubovskaya, E., Chen, Z., Carrodeguas, J.A., Kisker, C. and Bogenhagen, D.F. (2006) Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *J Biol Chem*, 281, 374-382.

4. Longley, M.J., Prasad, R., Srivastava, D.K., Wilson, S.H. and Copeland, W.C. (1998) Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro. *Proc Natl Acad Sci U S A*, 95, 12244-12248.

5. Fan, L., Kim, S., Farr, C.L., Schaefer, K.T., Randolph, K.M., Tainer, J.A. and Kaguni, L.S. (2006) A novel processive mechanism for DNA synthesis revealed by structure, modeling and mutagenesis of the accessory subunit of human mitochondrial DNA polymerase. *J Mol Biol*, 358, 1229-1243.

6. Lim, S.E., Longley, M.J. and Copeland, W.C. (1999) The mitochondrial p55 accessory subunit of human DNA polymerase gamma enhances DNA binding, promotes processive DNA synthesis, and confers N-ethylmaleimide resistance. *J Biol Chem*, 274, 38197-38203.

7. Lee, Y.S., Kennedy, W.D. and Yin, Y.W. (2009) Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell*, 139, 312-324.

8. Suomalainen, A., Majander, A., Haltia, M., Somer, H., Lonnqvist, J., Savontaus, M.L. and Peltonen, L. (1992) Multiple deletions of mitochondrial DNA in several tissues of a patient with severe retarded depression and familial progressive external ophthalmoplegia. *J Clin Invest*, 90, 61-66.

9. Zeviani, M., Servidei, S., Gellera, C., Bertini, E., DiMauro, S. and DiDonato, S. (1989) An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the Dloop region. *Nature*, 339, 309-311.

10. Hakonen, A.H., Heiskanen, S., Juvonen, V., Lappalainen, I., Luoma, P.T., Rantamaki, M., Goethem, G.V., Lofgren, A., Hackman, P., Paetau, A. *et al.* (2005) Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin. *Am J Hum Genet*, 77, 430-441.

11. Van Goethem, G., Luoma, P., Rantamaki, M., Al Memar, A., Kaakkola, S., Hackman, P., Krahe, R., Lofgren, A., Martin, J.J., De Jonghe, P. *et al.* (2004) POLG mutations in neurodegenerative disorders with ataxia but no muscle involvement. *Neurology*, 63, 1251-1257.

12. Winterthun, S., Ferrari, G., He, L., Taylor, R.W., Zeviani, M., Turnbull, D.M., Engelsen, B.A., Moen, G. and Bindoff, L.A. (2005) Autosomal recessive mitochondrial ataxic syndrome due to mitochondrial polymerase gamma mutations. *Neurology*, 64, 1204-1208.

13. Naviaux, R.K. and Nguyen, K.V. (2004) POLG mutations associated with Alpers' syndrome and mitochondrial DNA depletion. *Ann Neurol*, 55, 706-712.

14. Nguyen, K.V., Ostergaard, E., Ravn, S.H., Balslev, T., Danielsen, E.R., Vardag, A., McKiernan, P.J., Gray, G. and Naviaux, R.K. (2005) POLG mutations in Alpers syndrome. *Neurology*, 65, 1493-1495.

15. Suomalainen, A. and Isohanni, P. (2010) Mitochondrial DNA depletion syndromes-many genes, common mechanisms. *Neuromuscul Disord*, 20, 429-437.

16. Wong, L.J., Naviaux, R.K., Brunetti-Pierri, N., Zhang, Q., Schmitt, E.S., Truong, C., Milone, M., Cohen, B.H., Wical, B., Ganesh, J. *et al.* (2008) Molecular and clinical genetics of mitochondrial diseases due to POLG mutations. *Hum Mutat*, 29, E150-172.

17. Stumpf, J.D. and Copeland, W.C. (2010) Mitochondrial DNA replication and disease: insights from DNA polymerase gamma mutations. *Cell Mol Life Sci*.

18. Atanassova, N., Fuste, J.M., Wanrooij, S., Macao, B., Goffart, S., Backstrom, S., Farge, G., Khvorostov, I., Larsson, N.G., Spelbrink, J.N. *et al.* (2011) Sequence-specific stalling of DNA polymerase {gamma} and the effects of mutations causing progressive ophthalmoplegia. *Hum Mol Genet*, 20, 1212-1223.

19. Chan, S.S., Longley, M.J. and Copeland, W.C. (2005) The common A467T mutation in the human mitochondrial DNA polymerase (POLG) compromises catalytic efficiency and interaction with the accessory subunit. *J Biol Chem*, 280, 31341-31346.

20. Kasiviswanathan, R., Longley, M.J., Chan, S.S. and Copeland, W.C. (2009) Disease mutations in the human mitochondrial DNA polymerase thumb subdomain impart severe defects in mitochondrial DNA replication. *J Biol Chem*, 284, 19501-19510.

21. Luoma, P.T., Luo, N., Loscher, W.N., Farr, C.L., Horvath, R., Wanschitz, J., Kiechl, S., Kaguni, L.S. and Suomalainen, A. (2005) Functional defects due to spacer-region mutations of human mitochondrial DNA polymerase in a family with an ataxia-myopathy syndrome. *Hum Mol Genet*, 14, 1907-1920.

22. Szczepanowska, K. and Foury, F. (2010) A cluster of pathogenic mutations in the 3'-5' exonuclease domain of DNA polymerase gamma defines a novel module coupling DNA synthesis and degradation. *Hum Mol Genet*, 19, 3516-3529.

23. Chan, S.S. and Copeland, W.C. (2009) DNA polymerase gamma and mitochondrial disease: understanding the consequence of POLG mutations. *Biochim Biophys Acta*, 1787, 312-319

24. Brieba, L.G., Eichman, B.F., Kokoska, R.J., Doublie, S., Kunkel, T.A. and Ellenberger, T. (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J*, 23, 3452-3461

25. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature*, 391, 251-258.

26. Li, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J*, 17, 7514-7525.

27. Berdis, A.J. (2009) Mechanisms of DNA polymerases. *Chem Rev*, 109, 2862-2879.

28. Johnson, K.A. (2010) The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim Biophys Acta*, 1804, 1041-1048.

29. Lee, H.R., Helquist, S.A., Kool, E.T. and Johnson, K.A. (2008) Base pair hydrogen bonds are essential for proofreading selectivity by the human mitochondrial DNA polymerase. *J Biol Chem*, 283, 14411-14416.

30. Lee, H.R., Helquist, S.A., Kool, E.T. and Johnson, K.A. (2008) Importance of hydrogen bonding for efficiency and specificity of the human mitochondrial DNA polymerase. *J Biol Chem*, 283, 14402-14410.

31. Polesky, A.H., Dahlberg, M.E., Benkovic, S.J., Grindley, N.D. and Joyce, C.M. (1992) Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from Escherichia coli. *J Biol Chem*, 267, 8417-8428.

32. Polesky, A.H., Steitz, T.A., Grindley, N.D. and Joyce, C.M. (1990) Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from Escherichia coli. *J Biol Chem*, 265, 14579-14591.

33. Graziewicz, M.A., Longley, M.J., Bienstock, R.J., Zeviani, M. and Copeland, W.C. (2004) Structure-function defects of human mitochondrial DNA polymerase in autosomal dominant progressive external ophthalmoplegia. *Nat Struct Mol Biol*, 11, 770-776.

34. Van Goethem, G., Dermaut, B., Lofgren, A., Martin, J.J. and Van Broeckhoven, C. (2001) Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nat Genet*, 28, 211-212.

35. Johnson, S.J., Taylor, J.S. and Beese, L.S. (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci U S A*, 100, 3895-3900.

36. Jin, Z. and Johnson, K.A. Role of a GAG hinge in the nucleotide-induced conformational change governing nucleotide specificity by T7 DNA polymerase. *J Biol Chem*, 286, 1312-1322.

37. Baruffini, E., Horvath, R., Dallabona, C., Czermin, B., Lamantea, E., Bindoff, L., Invernizzi, F., Ferrero, I., Zeviani, M. and Lodi, T. (2011) Predicting the contribution of novel POLG mutations to human disease through analysis in yeast model. *Mitochondrion*, 11, 182-190.

38. Luo, N. and Kaguni, L.S. (2005) Mutations in the spacer region of Drosophila mitochondrial DNA polymerase affect DNA binding, processivity, and the balance between Pol and Exo function. *J Biol Chem*, 280, 2491-2497.

39. Murakami, K.S., Davydova, E.K. and Rothman-Denes, L.B. (2008) X-ray crystal structure of the polymerase domain of the bacteriophage N4 virion RNA polymerase. *Proc Natl Acad Sci U S A*, 105, 5046-5051.

40. Jeruzalmi, D. and Steitz, T.A. (1998) Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J*, 17, 4101-4113.

41. Gleghorn, M.L., Davydova, E.K., Rothman-Denes, L.B. and Murakami, K.S. (2008) Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell*, 32, 707-717.

42. Baruffini, E., Ferrero, I. and Foury, F. (2007) Mitochondrial DNA defects in Saccharomyces cerevisiae caused by functional interactions between DNA polymerase gamma mutations associated with disease in human. *Biochim Biophys Acta*, 1772, 1225-1235.

43. Stumpf, J.D., Bailey, C.M., Spell, D., Stillwagon, M., Anderson, K.S. and Copeland, W.C. (2010) mip1 containing mutations associated with mitochondrial disease causes mutagenesis and depletion of mtDNA in Saccharomyces cerevisiae. *Hum Mol Genet*, 19, 2123-2133.

44. Baruffini, E. and Lodi, T. Construction and validation of a yeast model system for studying in vivo the susceptibility to nucleoside analogues of DNA polymerase gamma allelic variants. *Mitochondrion*, 10, 183-187.

45. Lee, Y.S., Lee, S., Demeler, B., Molineux, I.J., Johnson, K.A. and Yin, Y.W. (2010) Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity. *J Biol Chem*, 285, 1490-1499.

46. Lee, Y.S., Johnson, K.A., Molineux, I.J. and Yin, Y.W. (2010) A single mutation in human mitochondrial DNA polymerase Pol gammaA affects both polymerization and proofreading activities of only the holoenzyme. *J Biol Chem*, 285, 28105-28116.

47. Chan, S.S., Longley, M.J. and Copeland, W.C. (2006) Modulation of the W748S mutation in DNA polymerase gamma by the E1143G polymorphismin mitochondrial disorders. *Hum Mol Genet*, 15, 3473-3483.

48. Palin, E.J., Lesonen, A., Farr, C.L., Euro, L., Suomalainen, A. and Kaguni, L.S. Functional analysis of H. sapiens DNA polymerase gamma spacer mutation W748S with and without common variant E1143G. *Biochim Biophys Acta*, 1802, 545-551.

49. Naviaux, R.K. and Nguyen, K.V. (2005) POLG mutations associated with Alpers syndrome and mitochondrial DNA depletion. *Ann Neurol*, 58, 491.

CHAPTER 3: Mapping 136 Pathogenic Mutations into Functional Modules in Human DNA Polymerase γ Establishes Predictive Genotype-phenotype Correlations for the Complete Spectrum of POLG Syndromes.

Gregory A. Farnum¹, Anssi Nurminen², and Laurie S. Kaguni^{1,2}

¹Department of Biochemistry and Molecular Biology and Center for Mitochondrial Science and Medicine, Michigan State University, East Lansing, MI 48824-1319, USA; ²Institute of Biomedical Technology, University of Tampere, 33014 Tampere, Finland;

ABSTRACT

We establish for the first time genotype-phenotype correlations for the complete spectrum of POLG syndromes, by refining our previously described protocol for mapping pathogenic mutations in the human *POLG1* gene to functional clusters in the catalytic core of the mitochondrial replicase, Pol γ (1). We assigned 136 mutations to five clusters and identify segments of primary sequence that can be used to delimit the boundaries of each cluster. We report that compound heterozygotes with two mutations from different clusters manifested more severe, earlier onset POLG syndromes, whereas two mutations from the same cluster are less common and generally are associated with less severe, later onset POLG syndromes. We also show that specific cluster combinations are more severe than others, and have a higher likelihood to manifest at an earlier age. Our clustering method provides a powerful tool to predict the pathogenic potential and predicted disease phenotype of novel variants and mutations in *POLG1*, the most common nuclear gene underlying mitochondrial disorders. We propose that such a prediction tool would be useful for routine diagnostics for mitochondrial disorders.

INTRODUCTION

Mitochondrial dysfunction due to impaired energy production via oxidative phosphorylation (OXPHOS) causes a variety of diseases, known collectively as mitochondrial disorders (2). Base substitutions, deletions or depletion of mitochondrial DNA (mtDNA) resulting in production of dysfunctional and/ or depletion of OXPHOS proteins is an important cause of mitochondrial dysfunction (2). In animal mitochondria, Pol γ is the only known DNA polymerase, and therefore it is responsible for maintenance of mtDNA integrity associated with mtDNA replication and repair (3). Human Pol γ is a heterotrimer consisting of a catalytic subunit, Pol γA , and a dimer of an accessory subunit, Pol γ B (4). Encoded by the *POLG1* gene, Pol γ A, known as the catalytic core, is a 140 kDa polypeptide that contains the 5'-3' DNA polymerase (pol), 3'-5' exonuclease (exo), and 5'-dRP lyase activities (3). Mutations in the POLG1 gene lead to accumulation of mtDNA deletions as well as mtDNA depletion, manifesting mitochondrial disorders termed POLG syndromes (5). The most severe form of POLG syndrome, known as Alpers syndrome, is caused by compound heterozygosity of two recessive POLG mutations, and leads to hepatocerebral mtDNA depletion syndrome during infancy and death at an early age (6). Late onset POLG syndromes, such as progressive external ophthalmoplegia (PEO) can be caused by dominant or recessive compound heterozygous mutations, and are associated with varying degrees of tissue specific mtDNA depletion and/ or single or multiple mtDNA deletions (7).

The crystal structure of Pol γ was determined in its apoenzyme form (PDB code 3IKM, (8)). Its catalytic core has three major domains: an N-terminal Exo domain that contains the exo active site; a C-terminal Pol domain that contains the pol active site; and a spacer domain that separates the Exo and Pol domains in primary sequence. Three subdomains are defined in the Pol domain: the palm subdomain represents the most highly conserved structural module between non-

homologous DNA polymerases, and contains the pol catalytic site coordinating two Mg²⁺ ions; the fingers subdomain is involved in binding the incoming dNTP substrate, and the thumb subdomain forms a major surface of the DNA binding channel. The spacer domain comprises two subdomains: the accessory interacting domain (AID), which forms the major hydrophobic contact with the proximal accessory subunit, and the intrinsic processivity (IP) domain, which forms a region of the upstream DNA binding channel and does not contact the accessory subunits. Both subdomains form distinct regions of the upstream DNA binding channel, but the flexible AID contributes to DNA binding affinity only when the accessory subunit stabilizes it in the correct position (9). The extreme N-terminal residues of the catalytic core do not show sequence conservation with other Exo domains in the family A polymerase group and are thus designated as a separate domain N-terminal domain (NTD), which presumably contains the mitochondrial leader sequence.

Pol γ is a member of the family A polymerase group and shares the highest sequence similarity with bacteriophage T7 DNA polymerase (T7 Pol) (8). We generated a structural model of the Pol γ ternary complex by superimposition of the apo-holoenzyme Pol γ structure (PDB code 3IKM) (1) with the structure of T7 DNA polymerase (PDB code 1T8E) with bound primer template DNA, incoming ddNTP and Mg²⁺ ions. We then mapped 58 Alpers mutations onto this structural model and reported that they cluster into distinct regions, which we termed clusters 1-5 (1). Because Pol γ is a family A polymerase, functional insight can be extrapolated from the extensive studies of conserved elements in bacterial DNA polymerase I (Pol I) and T7 Pol (10-15). This and evaluation of the unique features of Pol γ by biochemical analysis of yeast, fly, and human Pol γ variants (16-18), revealed that each cluster defines a unique functional region, which exhibits a distinct biochemical defect when affected by a pathogenic mutation (1). Cluster 1

mutations are predicted to cause a primary defect in pol activity, and affect residues involved directly in catalysis, or indirectly by affecting architectural residues that may disrupt the position of catalytic residues. Catalytic residues include those binding the two Mg^{2+} ions, those that make contact with the incoming dNTP, and those that make contact with the first, second and third nucleotide pairs in the nascent dsDNA, which are critical for correct positioning of the substrate in the pol active site. For example, the human pathogenic mutation R943H affects an amino acid residue that contacts the y phosphate of the incoming dNTP, and has been shown *in vitro* to have a 30-fold increase in K_m for dNTPs and a 5-fold decrease in k_{cat}, which together reduce pol activity 150-fold without affecting DNA binding affinity (19). Cluster 2 mutations are predicted to cause a primary defect in DNA binding affinity, and affect residues of the IP and AID that form the upstream DNA binding channel, which enhance DNA binding affinity via contacts with nucleotides upstream of the third base pair of primer template. Biochemical analysis of recombinant fly Pol y with a triple alanine substitution in amino acid residues K768/D769/F770 are located in the DNA binding channel wall of the IP domain, and cause a small decrease in binding affinity (17). Cluster 3 mutations are predicted to affect the pol: exo activity ratio and may have variable defects in DNA binding affinity. Many of the cluster 3 mutations map to a helix-coil-helix module (residues 295-312) located in the Exo domain that has been termed the "orienter" module (18). L304R of the "orienter" module has been studied biochemically in recombinant yeast Pol γ , and exhibits 3-fold increased exo activity, 20-fold decreased pol activity and 10-fold decreased DNA binding affinity (18). In addition, cluster 3 mutations map to the partitioning loop, which is a novel module conserved in Pol γ (residues 1050-1095) that is located between the fingers and palm subdomains and is not present in any other known polymerase (1). To date, no biochemical data is available for residues of the partitioning loop, although it has been shown that yeast strains homozygous for a mutation that is equivalent to G1051R in human Pol γ cause a 10-fold increase in point mutational frequency *in vivo* (20). Cluster 4 mutations map to the Exo domain along the distal accessory subunit interface, and are predicted to cause a biochemical defect similar to the R232G variant, which was shown *in vitro* to have reduced pol rate, increased exo activity and wild type (WT) DNA binding affinity (21). Cluster 5 mutations are located in the periphery of the IP subdomain and have not been shown to cause a biochemical defect. For example, human recombinant Pol γ harboring R627Q or R627W mutations exhibited no defects *in vitro* when analyzed for DNA binding affinity, pol activity and stimulation by the accessory subunit (22).

In applying our clustering model to the reported Alpers mutations, we found that Alpers syndrome was found only in compound heterozygous patients bearing two mutations from different clusters in Pol γ (1). Since publication, several new reports have been published that have added dozens of new mutations and mutation combinations to the POLG syndrome library (23-26). Here, we examine potential genotype-phenotype correlations by analyzing all reported POLG mutations available to date, not restricted to a phenotype or age-of-onset, utilizing our clustering model. In total, we assign 136 mutations to the five clusters and identify segments of primary sequence that can be used to delimit the boundaries of each cluster. We demonstrate the validity of our clustering model as the first to establish genotype-phenotype correlations for POLG syndromes, and show that it can be used to predict the pathogenic potential and biochemical defects of novel mutations, and to provide information about the likely severity of the POLG syndrome for compound heterozygotes with novel combinations of mutations.

METHODS

Computational analysis

We docked primer-template DNA (ptDNA) into the putative DNA-binding channel of the Pol γ apo-holoenzyme crystal structure (PDB code 3IKM, (8)) by superposition of the closed ternary complex of T7 Pol bound to ptDNA and dNTP (PDB code 1T8E, (27)) using PyMOL (http://www.pymol.org/). As we described previously, we found that the best overall alignment was obtained by aligning the palm subdomain of Pol γ A (residues 815-910, 1095-1239) to the palm subdomain of T7 Pol (residues 409-487, 611-704) (1). This model was used to map and evaluate all reported POLG disease mutations.

Statistical analysis

We used Pearson's chi-square analysis to calculate the significance of each age of onset (infantile, childhood, juvenile, adult), and the mutation combination affecting the same (=0) or different cluster (=1) of Pol γ . Logistic regression analysis was used to calculate odds ratios for having an infantile-onset disease. We conducted all statistical analyses with Stata (version 11.0). Two-sided p values were used with a significance level of 0.05.

RESULTS

Cluster assignment of all reported POLG disease mutations

The Alpers mutations represent only a subset of the total library of reported diseasecausing mutations in POLG. To investigate further the utility of the five functional clusters in Pol γ that we defined earlier to evaluate Alpers mutations (1), we compiled a list of all pathogenic mutations reported date by combining the POLG database point to entries (http://tools.niehs.nih.gov/polg/) with new data from recent reports (23-26, 28, 29). The list comprises a total of 136 pathogenic point mutations, including both dominant and recessive mutations. We included in our analysis POLG mutations that represented the only disease-causing defect identified in the patient, exclusive of mutations in other genes associated with mitochondrial dysfunction, and excluded consideration of null mutations. By excluding patients with mutations in multiple genes associated with mitochondrial dysfunction, a clear link can be established between functional defects in Pol γ and the severity of disease manifestations. Furthermore, we considered a pathogenic mutation to be dominant only if the family history demonstrated a dominant mode of inheritance, leaving out potential de novo mutations.

We mapped each of the pathogenic mutations onto our structural model and assigned them to a cluster by evaluating them individually. Although a cluster may contain residues that are distant in primary sequence, assignment was straightforward in most cases because each cluster occupies a distinct structural region in the Pol γ tertiary structure. Some mutations mapped to areas equidistant from two different clusters, and because we did not define explicitly the boundaries of the structural regions for each cluster in our previous analysis, we made use of additional criteria for cluster assignment. To do so, we analyzed each point mutation based on reported biochemical data to predict the functional defect it would cause. In this way, cluster assignment is based not only on structural location but also upon functional insight. In total, we assigned 136 pathogenic mutations to the five functional clusters we defined earlier (1) as shown in Figure 32. Due to the large library of reported mutations, we concluded that a sufficient number of mutations were assigned to each cluster to use their map positions as the sole means of defining the structural region and primary sequence occupied by each cluster. Mutations from each cluster are divided into several subclusters of 10-100 amino acid residues, and are distributed across the length of the *POLG1* gene (Figure 32, upper schematic). Despite the large separation of subclusters in primary sequence, the elements of a cluster fold into compact structural regions as illustrated in Figure 32 (lower panel).

68 mutations were assigned to cluster 1, defining seven subclusters located in the NTD, Exo and Pol domains. The primary biochemical defect caused by a cluster 1 mutation is predicted to be reduced pol activity. As expected, most mutations affect residues of the Pol domain, including the five conserved motifs within the Pol domain of family A polymerases that are essential for 5'-3' DNA polymerase activity (Figure 33, upper schematic, (12)). Subcluster 1e (residues 914-966) spans the O-helix and its environs (Figure 33, lower panel), the function of which is to bind the correct dNTP substrate by transitioning between an open and closed complex (30). The conserved amino acid residues on the O-helix are known as the Pol B motif (residues 943-958), and function to establish specific contacts with correctly base paired dNTP in the closed conformation (12). POLG mutations that disrupt the specific contacts with the incoming dNTP (H932Y, R943H, K947R, Y951N, and Y955C) will reduce fidelity and increase K_m (dNTP), without affecting DNA binding affinity (19, 31). Thus, mutations affecting this site are most likely dominant because they are capable of competing for dNTP binding with wild type (WT) Pol γ but are unable to polymerize nucleotides effectively, and are predicted to cause mtDNA damage that is associated with enzyme stalling (32).



Figure 32. Clustering of 136 pathogenic mutations within five functional modules in the catalytic subunit of human Pol γ.



Upper panel, schematic diagram of the human *POLG1* gene illustrating the clustering of 136 pathogenic mutations into discrete blocks of amino acid residues, which we term subclusters. Mutant alleles and subclusters are colored according to the cluster to which they belong: cluster 1, green; cluster 2, yellow; cluster 3, red; cluster 4, blue; cluster 5, cyan Figure 32 (Cont'd) (see the text for details). The palm (residues 815-910 and 1096-1239), fingers (residues 911-1048), and thumb (Th, residues 440-475 and 785-814) subdomains of the Pol domain are colored pink, and the partitioning loop (PL, residues 1049-1095) is red. The accessory (subunit) interacting domain (AID, residues 476-570) and the intrinsic processivity (IP, residues 571-784) subdomains of the spacer domain are colored in magenta.

Figure 32 (cont'd) The N-terminal domain (NTD, residues 1-170) and the Exo domain (residues 171-439) are colored in purple. *Lower panel*, structural model of the human Pol γ apoholoenzyme (PDB code 3IKM) with docked primer template DNA shown as orange ribbon (see *Computational methods* for details). The catalytic subunit of Pol γ is shown as a cartoon representation of the secondary structural elements (SSEs), with regions defined by clusters illustrated as space-filled modules, colored according to the schematic. The proximal and distal accessory subunits are shown as surface representations in light and dark gray, respectively.

Subcluster 1d (residues 848-895) comprises the RR loop (residues 845-863) (1) and the conserved Pol A motif (residues 887-896) of the palm subdomain (Figure 33) (3). The RR loop is equivalent to motif 2 in Pol I, and is critical for binding correctly base-paired ptDNA in the minor groove and the template DNA backbone (12). The Pol A motif in Pol I contacts the primer strand, binds Mg²⁺ via amino acid residue D705 (D890 in Pol γ) and discriminates against ribonucleotide incorporation via residue E710 (E895 in Pol γ) (33). Mutations in the Pol A motif at either D890 or E895 in Pol γ would likely manifest a dominant lethal phenotype, and the only reported patient to date with the E895G mutation died immediately after birth (34). Motif 2 mutations should always cause reduced pol activity but may also cause a DNA binding defect depending on the residue, such as the 5-fold defect reported for G848S and R852C (35). The DNA binding affinity of T851A, R853Q, Q879H, T885S may decrease slightly, but the reduction in pol activity is much greater (35). Overall, we feel it is unlikely for dominant mutations to reside in motif 2 because these mutant forms of Pol γ would not compete with WT Pol γ for DNA binding.



Figure 33. Architectural and functional subclusters of the Pol domain. *Upper panel*, schematic diagram of the *POLG1* gene as shown in Figure 32, with an additional upper section indicating the location of the three Exo motifs, labeled as I, II, and III, and the six Pol motifs, labeled as 1 (motif 1), 2 (motif 2), A (Pol A motif), B (Pol B motif), 6 (motif 6) and C (Pol C motif), which are conserved throughout family A polymerases (see text). Motifs and subclusters that are illustrated in the bottom panels are in shown bold. *Bottom panels*, SSEs and transparent surface representation of the palm and fingers subdomains of Pol γ (PDB code 3IKM) are shown (subdomains are colored in gray in the middle section of the schematic). *Bottom-left panel*, the five motifs of the Pol domain, which are

Figure 33 (cont'd) colored in pink in the upper section of the schematic, are shown as pink SSEs and are labeled accordingly. Docked primer template DNA is shown as orange ribbon, Mg²⁺ ions are shown as orange spheres, and the incoming dNTP is shown as orange sticks (see *Computational methods* for details). *Bottom-right panel*, subclusters 1D, 1E and 1F encompass one or more of the five conserved motifs of the Pol domain and are colored in light green. Subclusters 1A, 1B, 1C and 1G, colored in dark green, are located further structurally from the pol active site, and are considered to play an architectural role.

Subcluster 1f (residues 1104-1138) comprises motif 6 (residues 1097-1110) and the Pol C motif (residues 1134-1141) in the palm subdomain (Figure 33) (according to the nomenclature defined for Pol I (12)). Motif 6 is located on the Q-helix and binds correctly base paired template strand and the ptDNA minor groove in Pol I via the residues N845 and Q849 ((13), equivalent to N1098 and Q1102 in Pol γ , respectively). The Pol C motif in Pol γ binds Mg²⁺ via D1135, whereas H1134 and E1136 contact the primer strand. Subcluster 1g (residues 1157-1196) maps to a C-terminal region of the palm subdomain and forms an anti-parallel beta strand adjacent to the Pol A motif. Only K1191 is predicted to contact the primer terminus, whereas the rest of subcluster 1g serves an architectural role (1). Mutations affecting motif 6 have been reported to cause DNA binding defects along with pol defects in Pol I Klenow (13). However, an alanine substitution at the equivalent residue to Pol γ H1134 in Klenow (H881A) caused a decrease in k_{cat} from 6- to 66-fold (13) but retained DNA binding affinity, suggesting that a mutation in H1134 might be dominant.

We assigned mutations to subclusters 1d-g with high confidence because there is extensive biochemical data available for the Pol domain, in addition to the presence of the highly conserved amino acid sequence motifs. In contrast, a subcluster of seven mutations (R417T, C418R, G426S, L428P, M430L, G431V, S433C) mapped to the G-helix (Figure 33, lower right panel), a structural element that had not been studied previously. According to the current Exo domain assignment of residues 170-440 by Lee *et al.* (8), this subcluster would reside in the Exo domain, although

earlier T7 Pol (15) and Pol I structures (36) label the G-helix as part of the palm subdomain within the Pol domain. Regardless of the domain assignment, this subcluster is adjacent to motif 2 and motif 6, and we propose it serves an architectural role that contributes to pol activity indirectly, and assign it as subcluster 1c. Four mutations from the NTD were also observed to be closer structurally to functional Pol motifs as compared to Exo motifs; we thus assigned D136E and A143V as subcluster 1b (residues 136-143), and both L83P and F88L were assigned as subcluster 1a (residues 83-88).

25 mutations from cluster 2, which we predict to cause primary defects in DNA binding affinity, outline the putative DNA binding channel, and map to four subclusters within the thumb subdomain of the Pol domain (2a, residues 463-468), and the AID (2b, residues 497-517) and IP (2c, residues 561-617 and 2d, 752-767) subdomains of the spacer domain. Subcluster 2c contains motif 1, which in Pol I was shown to fold into a loop and binds DNA in the channel (12). Motif 1, together with subcluster 2d, form the major face of the DNA binding channel. Subcluster 2a maps to a region of the thumb subdomain at the accessory subunit interface where A467T, N468D, and L463F are positioned. Our group has characterized residue A467T biochemically, and it was shown to retain 70% of WT DNA binding affinity in a reconstituted holoenzyme form (22). Extensive biochemical analysis of the spacer domain in general has demonstrated that the hydrophobic core of the IP subdomain is critical for shaping the DNA binding channel wall (17, 22). Mutations that alter the IP subdomain, such as A467T, are predicted to perturb the channel to block DNA from entering, which is then observed as reduced DNA binding affinity.

8 mutations from cluster 3 map to the orienter module (18) of the Exo domain, and another 12 mutations from the same cluster map to the partitioning loop (1) of the Pol domain, which are defined by two subclusters (3b, residues 303-319 and 3d, 1047-1096, respectively). Additionally, a recent report from the Foury group (37) has shown that variants of yeast Pol γ bearing substitutions in residues of the thumb subdomain M602I (R802 in *Hs*) and the Exo II motif of the Exo domain A228V, R231K, R233W (S272, R275, H277 in *Hs*) also produce the altered pol:exo activity ratio that characterizes cluster 3 mutations. Therefore, we assign pathogenic residues R807P, R807C, R807H, and R804T as subcluster 3c (residues 804-807), and the Exo II motif mutations G268A, R275Q, and H277L as subcluster 3a (residues 268-277). In support of the proposed subcluster 3c, we note that the SYW triple alanine substitution in recombinant fly Pol γ , which was found to exhibit increased exo activity with both decreased pol activity and DNA binding affinity (17), maps adjacent to subcluster 3c in human Pol γ (residues S799/F800/W801). Similarly, the Exo II motif has been demonstrated to alter the pol: exo activity ratio in biochemical variants of Pol I by decreasing the affinity of the primer strand for the exo active site, consistent with the decrease in exo activity and WT pol activity observed for the R233W variant in yeast Pol γ (H277 in human Pol γ) (37).

Six mutations from cluster 4 C224Y, R227P, R227W, R232G, R232H and L244P map to a single region on the subunit interface of the Exo domain (subcluster 4a, residues 224-244). Biochemical studies of a human R232G variant in reconstituted holoenzyme form showed a decrease in pol rate and an increase in exo activity, with unchanged DNA binding affinity that derive from loss of a direct stimulation of pol activity by the distal accessory subunit (21). We propose that the other residues of cluster 4 have similar biochemical characteristics.

Ten mutations from cluster 5 are also located in the IP domain, but they map to the distal region far removed from the DNA binding channel, and define subcluster 5a (residues 623-648) and subcluster 5b (residues 737-749).

Cluster assignment can be used to predict the pathogenicity of a novel mutation.

Our clustering model provides an annotated guide to assign function to the entire mutational spectrum in the POLG1 gene, and we propose that it can be used to evaluate the potential pathogenicity of novel POLG mutations. In particular, when a novel mutation maps to a specific subcluster within the five functional clusters we have defined, we predict that it will likely be pathogenic and result in the biochemical defect associated with that subcluster. At the same time, we would argue that mutations that do not map to these specific amino acid sequence blocks are more likely to be neutral polymorphisms. In support of this model, we found that only 12 out of 87 reported single nucleotide polymorphisms (SNPs) from the human dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/) map within subclusters (Figure 34). Three of these are conservative substitutions of highly similar amino acids (F1092L, F1164L and F1164I), and likely have no functional consequences. These examples would suggest that although a mutation may affect a residue within a functional subcluster, conservative substitutions with amino acids with highly similar chemical properties may not be pathogenic, and this possibility should be taken into consideration when evaluating novel mutations. Additionally, missense mutations reported as SNPs in the SNP database are not necessarily non-pathogenic, as they could be uncommon recessive mutations that are masked when paired with a wild type allele.

Although the mutation library is large, the *POLG1* mutational map is likely not yet saturated, so routine sequencing of patients with POLG syndrome will continue to fill out the spectrum. We predict that most novel mutations will map within our current clusters because they are by definition "mutational hotspots" and importantly, we observe from their positions within the crystal structure that they encompass the known functional motifs required for the biochemical activities of Pol γ . For example, based upon the high sequence conservation with family A polymerases, we note that the essential elements of the pol active site such as the Pol A-C motifs,

motif 2, and motif 6 are each populated by human pathogenic mutations that map within our current cluster 1 (Figures 32 and 33).



Figure 34. SNPs reported in the *POLG1* **gene rarely map to clusters**. Of 87 reported SNPs in the dbSNP database, 75 (86%, black boxes) map outside the defined pathogenic clusters (shown as colored regions in the schematic). The remaining 12 SNPs (magenta boxes) that map within the clusters can be considered as non-deleterious changes within a high-risk region of the POLG tertiary structure. However, in our view, the possible pathogenicity of the SNPs reported in cluster 4 (dark blue) warrants careful experimental reevaluation.

Cluster combinations of compound heterozygotes correlate with age of onset of POLG syndrome.

Previously, we evaluated the combination of mutations in compound heterozygotes that cause Alpers syndrome using our original clustering model, and found that Alpers was triggered exclusively by the combination of two mutations from different clusters (1). Here, we evaluate all mutations causing POLG syndromes (caused by compound heterozygosity) using the model refined by the designation of subclusters within the original five functional clusters. Whereas clinical presentation and age of onset are extremely variable for POLG syndromes, in general the severity of symptoms decreases with increasing age of onset (38). We were able to catalogue a total of 341 compound heterozygous patients harboring two pathogenic mutations and we classified each patient by the age of onset reported. Each combination was assigned to one of four age groups: infantile, < 3 years of age; childhood, ages 3-12; juvenile, 13-20 years; and adult, > 20 years. The data are compiled and evaluated in Figure 35. Compound heterozygotes with two mutations from the same cluster were less common (113/341) and were generally associated with late onset POLG syndromes (Figure 35, upper panel). Compound heterozygotes with two mutations from different clusters were much more common (228/341), and comprised 109/119 of the infantile onset combinations. Clearly, the combination of two different clusters is not just a specific trigger for Alpers syndrome; rather, it establishes a universal genotype-phenotype correlation for all POLG syndromes.



Figure 35. Analysis of mutation combinations as cluster combinations reveals predictive genotype-phenotype correlations. The data used to compile the information in this figure was

Figure 35 (cont'd) derived from the literature listed under supplemental references in appendix C. The number of mutation combinations manifesting POLG syndrome at each of the four age groups is shown in each panel for specific cluster combinations. Age of onset trends can be used to predict the severity of POLG syndrome for an individual with compound heterozygous mutations in *POLG1. Upper panel*, POLG syndrome age of onset trends for mutation combinations of two mutations from different clusters versus the same cluster. *Middle panel*, patient data show trends in which severe cluster combinations have an earlier age of onset, whereas less severe or uncommon cluster combinations have a later age of onset. *Lower panel*, earlier age of onset trends for more severe cluster/ subcluster combinations.

Critical functions: clusters 1, 3, and 4

We predict that mutations affecting cluster 1 will reduce primarily pol activity, and these have been associated with increased replication fork stalling in vivo (32). The stalling phenotype is characterized by slow progression of or suspended replication forks, and causes an accumulation of replication intermediates that can lead to double-stranded DNA breaks, mtDNA deletions and base substitution mutations (39-41). It is thus unlikely that a compound heterozygote with cluster 1 mutations would be viable unless the defects are very mild. Indeed, only nine of these combinations have been reported (Figure 36), and involve mutations such as G923D and A957S that retain 25% WT pol (19), which is a relatively mild pol defect compared to other cluster 1 mutations, such as the 100-fold reduction of pol activity engendered by the Y955C mutation (32). Although the partitioning loop has not been studied biochemically, cluster 3 mutations may produce a biochemical phenotype similar to that of the L260R variant in yeast Pol γ (L304R in Hs Pol γ), which is characterized by an increase in exo activity, and decreases in pol activity and DNA binding affinity (18). Only four combinations have been reported for compound heterozygotes with two cluster 3 mutations (Figure 36), which may indicate that most of these combinations are lethal. Cluster 4 mutations are predicted to increase exo activity and decrease pol rate, while retaining WT DNA binding affinity (21). Interestingly, cluster 4 mutations have not been reported in combination with other cluster 4 mutations, or together with cluster 3 mutations. When Pol γ is stalled, the primer strand can translocate between the exo active site the pol active site, a mode

described as idling by Atanassova *et al.* (32). We speculate that cluster 3 and 4 mutants may manifest an idling phenotype because of their enhanced exo: pol activity ratios, which is likely to increase non-productive turnover by mutant Pol γ , via exonucleolytic hydrolysis of correctly incorporated nucleotides. In compound heterozygotes, two idling Pol γ s may be lethal, explaining the absence of cluster 3 + 4 combinations, as well as cluster 4 + 4 combinations. Taken together, the lack of complementation observed for clusters 1, 3, and 4 suggests that they serve essential functions in mtDNA metabolism.



Figure 36. Analysis of all compound heterozygous mutation combinations as cluster combinations. The data used to compile the information in this figure was derived from the literature listed under supplemental references in appendix C. The number of mutation combinations manifesting POLG syndrome at each of the four age groups is shown for all cluster combinations for which patient data have been reported; no patient data have been reported for the three cluster combinations 3+4, 4+4 and 4+5. Age of onset trends can be used to predict the severity of POLG syndrome for an individual with compound heterozygous mutations in *POLG1*. Patient data show trends in which severe cluster combinations have an earlier age of onset, whereas less severe or uncommon cluster combinations have a later age of onset.

Moderately severe dysfunction: clusters 2 and 5

Mutations affecting cluster 2 will most likely cause a primary defect in DNA binding affinity. Processivity, defined as the number of nucleotides incorporated in a single DNA binding event, varies directly with pol rate and DNA binding affinity and as a consequence, cluster 2 mutations reduce processivity. Individuals carrying homozygous A467T mutations present as juveniles with POLG syndrome (26, 42, 43). As we showed that Pol γ carrying the A467T substitution shows a DNA binding affinity 70% of that of WT (22), this suggests that moderate reductions in DNA binding and pol activity may be tolerated throughout childhood. It thus seems likely that most same cluster combinations of cluster 2 mutations will not be lethal, and instead will lead to development of later-onset POLG syndrome. Consistent with this hypothesis, 42 such combinations have been reported, and the majority show juvenile or adult onset (Figure 35, middle panel).

W748S, R627Q and R627W are cluster 5 mutations and have been studied *in vitro* in reconstituted human Pol γ holoenzyme, but all three revealed no biochemical phenotype (22, 44). These mutations map to the distal surface of the IP domain, far removed from the DNA binding channel. Despite the lack of a documented biochemical phenotype, cluster 5 mutations have been reported in 140 compound heterozygotes with POLG syndrome manifesting at all ages (Figures 35 and 36). Therefore, this region clearly carries a biologically-relevant function in mtDNA metabolism because cluster 5 mutations cause mtDNA instability (44, 45). As for cluster 2 mutations, cluster 5 mutations are found together in compound heterozygote individuals with juvenile or adult onset POLG syndrome. Compound heterozygotes with a cluster 2 mutation and a cluster 5 mutation have been reported in 40 patients, and a majority of the combinations resulted in juvenile or adult onset POLG syndrome (Figure 35, middle panel). Notably, this cluster combination is the only one that did not typically cause infantile or childhood onset POLG

syndromes. To this point, 39/39 patients that were homozygous for W748S only developed a mild adult-onset POLG syndrome (46-48). In sum, cluster 2 and cluster 5 mutations, in contrast to those in clusters 1, 3 and 4, are observed to complement each other, suggesting they serve less critical functions in mtDNA metabolism.

Severe combination: cluster 1 + 2

Compound heterozygotes with cluster 1 and cluster 2 mutations have been reported in 70 patients, and are the most abundant of all cluster combinations (Figure 35, middle and lower panels). The majority (44/70) of these combinations caused infantile onset POLG syndrome. In fact, the most severe combinations observed in this study are the subcluster combinations 1D + 2A and 1E + 2A. These combinations cause infantile onset POLG syndrome in 17/18 and 19/20 cases, respectively. Cluster 1 mutants have reduced pol activity but will compete for DNA binding and cause enzyme stalling when bound. Cluster 2 mutations show reduced DNA binding affinity and would be unable to compete for DNA binding with a cluster 1 mutant. We predict this would result in more stalled enzyme complexes than for bound cluster 2 complexes alone. Because cluster 2 mutants show only modest decreases in pol activity, mtDNA depletion will progress gradually until a threshold level is reached to give rise to symptoms of POLG syndrome.

The two most common cluster 2 mutations are A467T and P587L, and the latter is always reported *in cis* with T251I in patients. The occurrence of two mutations together suggests either a genetic founder effect (as in Hakonen *et al.* (46)), or synergic or compensatory effect of the variants. T251I maps to the Exo domain, but we consider it to be a SNP for the following reasons: T251 is not a conserved amino acid, is located on the protein surface, and is not in a reasonable proximity to any functional region. In contrast, P587L is conserved in mammals and fly, is positioned in the putative DNA binding channel, and two pathogenic mutations, G588D and

P589L, are located directly adjacent to P587 on the same hairpin. By considering T251I as a SNP and excluding it from the functional analysis, P587L can be compared directly to A467T for its relative pathogenicity. Although no biochemical data is available for P587L, we suggest it will reduce DNA binding affinity to a much lesser extent than A467T. Direct comparison of the two mutations is possible in two different cluster 1 mutations. One patient was reported as compound heterozygous for G848S/A467T with infantile onset (43), whereas another was reported as compound heterozygous for G848S/P587L with adult onset (26). Homozygosity for P587L+T251I results in a late-onset myopathy phenotype ((7) and Suomalainen, unpublished). The same pattern is observed with K1191N (7, 26).

Severe combination: cluster 1+5

Compound heterozygotes with cluster 1 and cluster 5 mutations were reported showing that 23/42 mutation combinations had infantile onset POLG syndrome (Figure 35, middle panel). Such a high occurrence of the most severe form of POLG syndrome, together with the relatively large number of combinations documented, indicates that the combined defects from cluster 1 and cluster 5 are highly prone to cause mtDNA depletion at a young age. In particular, the most common cluster 5 mutation, W748S, manifested infantile onset POLG syndrome in 28/31 patients when found in combination with cluster 1 mutations. Two particularly severe combinations are the subcluster combinations 1D + 5B and 1E + 5B; both result in infantile onset POLG syndrome in >80% of affected patients (Figure 35, lower panel).

Other combinations

Cluster 4 mutations were reported in a total of 17 combinations with cluster 1 and cluster 2 (Figure 35, lower panel). The incidence of infantile onset POLG syndrome was high (15/17),

and may indicate that biochemical defects of cluster 4 mutations are uniquely severe *in vivo*. Infantile onset was less common for compound heterozygotes with cluster 3 mutations in combination with cluster 1 (9/17), cluster 2 (10/21) and cluster 5 (3/10) (Figure 36). These results suggest that cluster 4 mutations are more pathogenic than cluster 3 mutations.

Expected incidence of specific symptoms caused by POLG syndrome can be predicted by the age of onset.

The clinical manifestations reported for each compound heterozygous patient were compiled in effort to study the incidence and timeline of symptoms caused by POLG syndrome. To analyze the clinical descriptions in terms of individual symptoms, it was necessary to group similar and synonymous symptoms into symptom groups (Table 6). Very common symptoms, such as ataxia, seizures, hypotonia, and migraines, were left as individual entries because their sample size was sufficiently large, and these symptoms have been established as hallmarks for specific POLG syndromes. For other symptom groups, it was necessary to combine different symptoms into a group of related symptoms based on the affected tissue type resulting in the observed symptom. For example, the CPEO group includes all synonyms for ophthalmoplegia as well as diplopia and ptosis, two symptoms that both indicate impaired function of the extraocular muscles. Neurological disorders are divided into two subcategories; the neuropathy group includes symptoms caused by damage to peripheral neurons, and the CNS group includes symptoms caused by damage to neurons of the central nervous system. The developmental delay group includes symptoms related to an encephalopathy that involve an altered mental state. The myopathy, hepatopathy, and GI groups involve symptoms that affect or are an indicator of dysfunction in muscle tissue, the liver, or the gastrointestinal tract, respectively. Figure 37 shows the incidence of each symptom group in patients as a function of age of onset. The more severe symptoms, such as hepatopathy, hypotonia, and developmental delay have a high incidence among patients with infantile onset POLG syndrome, whereas less severe symptoms such as CPEO manifest mainly in patients with adult onset POLG syndrome. These data demonstrate that the symptoms caused by POLG syndromes represent a continuum of manifestations that become less severe as the age of onset increases. These syndromes have been characterized in the past according to specific groups of co-occurring symptoms that present in patients (Figure 38). For example, Alpers syndrome is characterized by seizures, developmental delay, and liver failure. Figure 38 displays the timeline of POLG syndromes, and outlines the co-occurring symptoms that define each of the specific syndromes that have been used to characterize patients with pathogenic POLG mutations.



Figure 37. Age of onset correlation for reported symptoms. Symptoms associated with POLG syndromes manifest typically at different ages. The symptoms reported in the patients have been grouped together based on affected tissue types, and ordered according to their ages of onset. The more severe symptoms and forms of POLG syndromes are at the left end of the figure, while the less severe symptoms are at the right end. Detailed information of the symptom grouping is shown in Table 6.


Figure 38. Symptoms associated with different POLG syndromes. Depending on the specific mutations and cluster combinations of the *POLG1* gene mutations in individual patients, POLG syndromes can become symptomatic at different ages and manifest as pathogenic conditions in different tissue types. The figure shows a continuing spectrum of decreasing symptom severity from top to bottom, as well as a delayed age of onset from left to right. MCHS, Childhood myocerebrohepatopathy spectrum; MEMSA, Myoclonic epilepsy myopathy sensory ataxia; ANS, Autonomic Nervous System Dysfunction; arPEO/ adPEO, Autosomal Recessive/ Dominant Progressive External Ophthalmoplegia.

Symptom	Symptoms included	Patient
group		number
Hypotonia	Hypotonic	47
Alpers	Alpers syndrome, Assumed Alpers	51
Developmental delay	Altered mental status, Cognitive delay, Dementia, Developmental delay, Encephalopathy, Failure to thrive (FTT), Growth retardation, Lowered consciousness, Psychomotor delay, Psychomotor regression, Retardation	148
Hepatopathy	Cholestasis, Hepatocerebral, Hepatomegaly, Jaundice, Lactic acidaemia, Lactic acidosis, Liver cirrhosis, Liver dysfunction, Liver failure	98
GI	Cyclic vomiting, Delayed gastric emptying, Diarrhea, GI dysmotility, GI problems, GI reflux, Pancreatitis, Vomiting	37
Seizures	Abrupt onset of seizure, Epilepsia partialis, Epilepsy, Focal seizures, Hemiparesis, Intractable seizure, Myoclonic seizures, Status epilepticus	162
Migraines	Headache/ migraine	45
CNS	Dysarthria, Dysphagia, Dystonia, Encephalomyopathy, Metabolic stroke, Nystagmus, Paralysis, Parkinson's disease, Spasticity, Stroke, Tremor	114
Myopathy	Abnormal muscle histology, Abnormal muscle ultrastructure, Cox-deficient muscle, Cox-negative, Exercise intolerance, Mitochondrial myopathy, Muscle weakness, Ragged red fibers	86
Neuropathy	Axonal sensorimotor polyneuropathy, Delayed myelination, Demyelinating neuropathy, Optic atrophy, Peripheral neuropathy, Polyneuropathy, Sensomotor neuropathy	80
Ataxia	Cerebellar ataxia, Cerebellar atrophy, Movement disorder (ataxia), Sensory ataxia	30
СРЕО	Diplopia, External ophthalmoplegia, Ocular myopathy, Ophthalmoplegia, PEO, Ptosis	168
Other	Areflexia, Arrhythmia, Atrial hypertrophy, Bilateral deafness, Bilateral lesions of thalami, CPK abnormalities, Cardiopathy, Chronic bronchitis, Cortical blindness, Distal muscle wasting, Hearing loss, Hypoglycemia, Hypothyroidism, Ischaemic episodes, Ketosis, Leigh syndrome, Microcephaly, Occipital strokes, Ocular bulbar weakness, Proximal weakness, Renal tubulopathy, Respiratory deficiency, Septicemia, Stroke-like episodes	74

Table 6. Grouping of POLG syndrome symptoms

DISCUSSION

Due to the mutational diversity in the *POLG1* gene and the variable clinical features of POLG syndromes, establishment of genotype-phenotype correlations has remained elusive (49). Furthermore, a number of *de novo* variants emerge in routine diagnosis laboratories, and interpretation of their pathogenic role is challenging. We show here that detailed protein structural and functional studies, with the aid of a crystal structure, combined with knowledge from experimental and disease mutations can be combined to predict clinical consequences of an identified variant, and to distinguish likely pathogenic mutations from polymorphic variants. In this report, we demonstrate clear genotype-phenotype correlations for predicting the severity and hence age of onset of compound heterozygous POLG syndromes. We refine our clustering model of Pol γ (1), proposing five functional clusters, each of which comprise subclusters that define structural regions of the catalytic core of Pol γ . Within these clusters, we can predict the likely biochemical defects and intrinsic pathogenicity of novel point mutations. In applying this clustering protocol to nearly 140 combinations of POLG compound heterozygous mutations, we showed that each cluster combination can be used to give information of the potential severity of disease outcome. Strikingly, we were able to predict correctly the age of onset in 90% of infantileonset POLG syndromes via the presence of two mutations from different clusters in compound heterozygotes of POLG. Furthermore, we show that age of onset can be used to predict the symptoms that a patient will manifest, and provide a timeline of the disease progression that may be expected.

Individuals carrying mutations in *POLG1* can manifest a disease at any age, depending on the mutation combination as well as the severity of the Pol γ functional defects (5, 26, 50). Presumably, the most severe Pol γ defects would be embryonic lethal, and this is likely the reason for example, that nucleotide changes that yield amino acid substitutions in the Mg²⁺ binding residues D890 and D1135 have never been reported. We found a similar trend in our clustering analysis showing a lower occurrence of compound heterozygotes from the same cluster for clusters 1, 3 and 4. The low occurrence of same-cluster compound heterozygotes may indicate that the majority of potential combinations of this type are lethal, and the lack of complementation supports our assignment of a common biochemical defect. This does not, however, apply to IP-domain mutations, which as homozygous are a common cause of severe progressive disorder that has a later disease-onset in juvenile age or early adulthood. Therefore, combinations of mutations that cause the most severe form of the POLG syndrome allow embryonic development, but lead to early infantile death.

We believe that our clustering model of *POLG1* mutations can be used as a tool to aid in the diagnosis of patients with mitochondrial disorders. The *POLG1* gene shows considerable variation, and novel polymorphic changes are common. A valuable existing tool, the polymerase gamma database (http://tools.niehs.nih.gov/polg/), lists reported mutations, but their pathogenic role sometimes remains unclear, especially for rare changes. Our prediction tool adds considerable power via structural analysis for interpretation of the consequences of identified rare variants, and can provide information for novel changes. Furthermore, it predicts the combinatorial effect of compound heterozygous changes for a patient, and provides information of typical clinical manifestations and ages of onset. We note that these predictions may not always be accurate, both because of the variability in disease manifestation introduced by genetic background, and the possibility that a number of the patients reported with POLG syndrome may also have other inherited mutations in their mitochondrial or nuclear DNAs in addition to the reported mutations in the POLG gene; although we excluded known cases of digenic mitochondrial diseases in this study, the latter possibility remains problematic and would be addressed by whole exome sequencing. Despite these potential drawbacks, we believe that our clustering protocol provides additional information that will prove important not only for diagnosis and genetic counseling, but also for treatment: *e.g.*, the epilepsy drug valproate is toxic to the liver of patients with *POLG1* mutations (51), and as epilepsy is a common initial symptom in children with POLG diseases, many patients have died before receiving a liver transplant. The ability to predict pathogenic likelihood for new POLG variants will be important when considering valproate as a treatment in severe epilepsies. Finally, because the predictions are based on reported patient data, the everincreasing number of documented *POLG1* mutations will serve to enhance the diagnostic power of the clustering protocol.

APPENDIX

APPENDIX

Subcluster Definitions

Cluster 1

Location in structure: Pol active site and environs

Biochemical phenotype: Reduced pol activity caused by a defect in the pol rate

Description: Cluster 1 mutations are predicted to cause a primary defect in pol activity, and affect residues involved directly in catalysis, or indirectly by affecting architectural residues that may disrupt the position of catalytic residues. Catalytic residues include those binding the two Mg^{2+} ions, those that make contact with the incoming dNTP, and those that make contact with the first, second and third nucleotide pairs in the nascent dsDNA, which are critical for correct positioning of the substrate in the pol active site (Euro et al, 2011; Farnum et al, this report).

Subcluster 1A

Location in structure: Outer layer of the holoenzyme, contacts subcluster 1G and is adjacent to subcluster 1F

Biochemical phenotype: Mild defect in pol activity or diminished stability/ solubility of the polypeptide

Description: No biochemical data are available for this region. We predict an architectural role via stabilization of subcluster 1G.

Subcluster 1B

Location in structure: Outer layer of the holoenzyme, contacts subclusters 1C, 1G, and 1F **Biochemical phenotype:** Mild defect in pol activity or diminished stability/ solubility of the polypeptide

Description: No biochemical data are available for this region. We predict an architectural role via stabilization of subclusters 1C, 1G, and 1F.

Recessive/ dominant mutation predictions: Recessive

Subcluster 1C

Location in structure: Mostly solvent exposed, contacts subclusters 1B, 1D, 1F, and 1G; equivalent to the G-helix of other family A polymerases

Biochemical phenotype: Mild defect in pol activity or diminished stability/ solubility of the polypeptide

Description: No biochemical data are available for this region. This subcluster is adjacent to motif 2 and motif 6 in the pol domain, and we propose it serves an architectural role that contributes to pol activity indirectly, by stabilizing subclusters 1B, 1D, 1F, and 1G. Patient data shows these mutations to be recessive.

Recessive/ dominant mutation predictions: Recessive

Subcluster 1D

Location in structure: This subcluster forms a large portion of the pol active site and contains two highly conserved motifs that are found in all family A polymerases: the RR loop (motif 2) and motif A (Loh and Loeb, 2005).

Biochemical phenotype: Reduced pol rate, possibly reduced DNA binding affinity

Description: Motif 2 and the Pol A motif are both critical for catalysis but contribute to catalysis by different mechanisms. Therefore, a mutation in subcluster 1D should be evaluated by the motif to which it is closest in primary sequence.

Motif 2 (residues 845-863) In Pol γ , motif 2 in the pol domain has been termed the "RR loop" because of two sequential arginines (R852 and R853) located on the tip of the loop, whereas family A polymerases have only a single arginine (equivalent to R853) (Euro et al, 2011). Overall, motif 2 is critical for binding correctly base-paired primer-template DNA in the minor groove, and to stabilize the template DNA backbone in the active site (Loh and Loeb, 2005). These residues form specific contacts with DNA in the pol active site and are major determinants for the catalytic efficiency and fidelity of DNA synthesis by Pol γ . Motif 2 mutations should always cause reduced pol activity but may also cause a DNA binding defect depending on the residue, such as the 5-fold defects reported for G848S and R852C (Kasiviswanathan et al, 2009). The DNA binding affinity of T851A, R853Q, Q879H, T885S may be decreased slightly, but the reduction in pol activity is much greater (Kasiviswanathan et al, 2009). Overall, it is unlikely for motif 2 mutations to present as dominant mutations because these mutant forms of Pol γ would not compete effectively with wild type Pol γ for DNA binding.

Pol A motif (residues 887-896) The Pol A motif of family A DNA polymerases forms the site of catalysis where it contacts closely the primer strand and binds the Mg²⁺ ion required for the chemical steps of nucleotide polymerization (D890 in Pol γ). Additionally, a residue of the Pol A motif discriminates against ribonucleotide incorporation (E895 in Pol γ), and mutation causes severe pol defects in addition to increased ribonucleotide incorporation (Astatke et al, 1998). Mutations in the Pol A motif at either D890 or E895 in Pol γ would likely manifest a dominant

lethal phenotype, and the only patient carrying the E895G mutation to date died immediately after birth (Spinazzola et al, 2009). In contrast to these two critical residues, adjacent residues within the Pol A motif play lesser roles in pol activity and mutations in them are observed more frequently, resulting in reduced pol activity without a significant effect on DNA binding affinity, and thus they can produce a dominant phenotype. Overall, though dominant mutations are fairly rare for subcluster 1D, novel mutations affecting residues of the Pol A motif would present a moderate risk for dominant inheritance.

Recessive/ dominant mutation predictions: Moderate risk of dominant mutations within the Pol A motif (residues 887-896)

Subcluster 1E

Location in structure: This subcluster comprises most of the fingers subdomain of the pol domain, including the O-helix and the Pol B motif (Loh and Loeb, 2005).

Biochemical phenotype: Reduced pol rate, and altered dNTP binding affinity and fidelity

Pol B motif (residues 943-958) The conserved amino acid residues on the O-helix are known as the Pol B motif, and function to establish specific contacts with correctly base paired dNTP in the closed conformation. POLG mutations that disrupt the specific contacts with the incoming dNTP (H932Y, R943H, K947R, Y951N and Y955C) will reduce fidelity and increase K_m (dNTP), without affecting DNA binding affinity (Graziewicz et al, 2004; Batabyal et al, 2010). Mutations affecting this site will most likely be dominant because they are capable of competing for dNTP binding with wild type Pol γ , but are unable to polymerize nucleotides effectively, and are predicted to cause mtDNA damage that is associated with enzyme stalling (Atanassova et al, 2011). **Recessive/ dominant mutation predictions:** High risk of dominant mutations in the Pol B motif (residues 943-958); subcluster 1E residues outside this motif can be dominant, but the risk is much lower.

Subcluster 1F

Location in structure: This subcluster forms a segment of the pol active site and contains two highly conserved motifs that are found in all family A polymerases, the Pol C motif and motif 6 (Loh and Loeb, 2005).

Biochemical phenotype: Reduced pol rate, possibly reduced DNA binding affinity

Description: Motif 6 (residues 1097-1110) is located on the Q-helix and binds correctly base paired template strand and the minor groove of the primer-template DNA. The Pol C motif (residues 1134-1141) binds Mg^{2+} via residue D1135, whereas H1134 and E1136 contact the primer strand. Mutations in the Pol C motif will affect the rate of the chemical step of dNTP addition to the primer terminus.

Recessive/ dominant mutation predictions: Moderate risk of dominant mutations in the Pol C motif (residues 1134-1141)

Subcluster 1G

Location in structure: Subcluster 1G (residues 1157-1196) maps to a C-terminal region of the palm subdomain, and forms an anti-parallel beta strand adjacent to the Pol A and Pol C motifs (Loh and Loeb, 2005).

Biochemical phenotype: Mild defect in pol activity or diminished stability/ solubility of the polypeptide.

Description: This subcluster is adjacent to motifs A and C, and we propose it serves an architectural role that contributes to pol activity indirectly, by stabilizing subclusters 1D and 1F. Only K1191 is predicted to contact the primer terminus, whereas the remainder of the subcluster 1G residues serve an architectural role. Patient data shows these mutations to be recessive.

Recessive/ dominant mutation predictions: Recessive

Cluster 2

Location in structure: Upstream DNA binding channel of the spacer domain

Biochemical phenotype: Reduced DNA binding affinity

Affected step in polymerization pathway: The k_{on} and k_{off} rates will be altered by cluster 2 mutations, which will manifest as an increase in the K_d of primer-template DNA binding.

Description: Cluster 2 mutations are predicted to cause a primary defect in the DNA binding affinity of Pol γ to the primer-template DNA to form the binary complex. They affect residues of the intrinsic processivity subdomain (IP) and accessory interacting domain (AID) of the spacer domain that form the upstream DNA binding channel, and which enhance DNA binding affinity via contacts with nucleotides upstream of the third base pair of primer template DNA.

Recessive/ dominant mutation predictions: Recessive

Subcluster 2A

Location in structure: Subcluster 2A maps to a region of the thumb subdomain of the pol domain at the accessory subunit interface where A467T, N468D and L463F are positioned.Biochemical phenotype: Reduced DNA binding affinity

Description: Extensive biochemical analysis of the spacer domain in general has demonstrated that the hydrophobic core of the IP subdomain is critical for shaping the DNA binding channel wall, and mutations are predicted to perturb the channel to block DNA from entering, which results in reduced DNA binding affinity.

Recessive/ dominant mutation predictions: Recessive

Subcluster 2B

Location in structure: Subcluster 2B maps to the region of the AID that is predicted to contact the upstream DNA duplex.

Biochemical phenotype: Reduced DNA binding affinity

Description: The AID is stabilized in the presence of the accessory subunit and enhances DNA binding affinity through many positively charged residues. Subcluster 2B maps to the region of the AID containing the positively charged residues, and mutations may alter their interaction with bound DNA.

Recessive/ dominant mutation predictions: Recessive

Subcluster 2C

Location in structure: Subcluster 2C contains motif 1, which in Pol I was shown to fold into a loop that binds DNA in the channel (Loh and Loeb, 2005). Motif 1, together with subcluster 2D, form the major face of the putative DNA binding channel.

Biochemical phenotype: Reduced DNA binding affinity

Description: Extensive biochemical analysis of the spacer domain in general has demonstrated that the hydrophobic core of the IP subdomain is critical for shaping the DNA binding channel

wall, and mutations are predicted to perturb the channel to block DNA from entering, which is then observed as reduced DNA binding affinity.

Recessive/ dominant mutation predictions: Recessive

Subcluster 2D

Location in structure: Subcluster 2D, together with subcluster 2C, form the major face of the putative DNA binding channel.

Biochemical phenotype: Reduced DNA binding affinity

Description: Extensive biochemical analysis of the spacer domain in general has demonstrated that the hydrophobic core of the IP subdomain is critical for shaping the DNA binding channel wall, and mutations are predicted to perturb the channel to block DNA from entering, which is then observed as reduced DNA binding affinity.

Recessive/ dominant mutation predictions: Recessive

Cluster 3

Location in structure: Cluster 3 defines a group of structural modules that are located between the pol and exo active sites: a region near the exo active site (3A), the orienter module (3B), a segment of the thumb subdomain (3C), and the partitioning loop (3D).

Biochemical phenotype: Predicted to affect the pol: exo activity ratio and may have variable defects in DNA binding affinity

Description: Cluster 3 mutations are predicted to affect the pol: exo activity ratio; they have been reported to cause DNA binding defects in most biochemical variants studied.

Subcluster 3A

Location in structure: A region near the exo active site that comprises the Exo II motif **Biochemical phenotype:** Reduced exo activity, wild type pol activity

Description: The Exo II motif has been demonstrated to alter the pol: exo activity ratio in biochemical variants of Pol I by decreasing the affinity of the primer strand for the exo active site, consistent with the decrease in exo activity and wild type pol activity observed for the R233W variant in yeast Pol γ (H277 in human Pol γ) (Foury and Szczepanowska, 2011).

Recessive/ dominant mutation predictions: Recessive

Subcluster 3B

Location in structure: A helix-coil-helix module (residues 295-312) located in the Exo domain that has been termed the "orienter" module

Biochemical phenotype: Reduced DNA binding affinity, reduced pol activity, possible increase in exo activity

Description: All biochemical variants of the orienter module have shown reduced DNA-binding affinity and reduced pol activity and in addition, the L304R variant showed a significant increase in exo activity (Foury and Szczepanowska, 2010).

Recessive/ dominant mutation predictions: Recessive

Subcluster 3C

Location in structure: A segment of the thumb subdomain of the pol domain located on the inner face of the DNA binding channel between the pol and exo active sites

Biochemical phenotype: Increased exo activity, decreased pol activity, possible decrease in DNA binding affinity

Description: Mutations in subcluster 3C have been reported to cause DNA binding defects in most biochemical variants studied. Notably, an SYW triple alanine substitution in recombinant fly Pol γ , which was found to exhibit increased exo activity with both decreased pol activity and DNA binding affinity (Luo and Kaguni, 2005), maps adjacent to subcluster 3C in human Pol γ (residues ⁷⁹⁹SFW⁸⁰¹). Biochemical study of a variant of yeast Pol γ M602I (R802 in Hs Pol γ) showed wild type DNA binding affinity, decreased pol activity, and increased exo activity (Foury and Szczepanowska, 2011).

Recessive/ dominant mutation predictions: Recessive

Subcluster 3D

Location in structure: The partitioning loop, which is a novel structural module conserved in Pol γ (residues 1050-1095) that is located between the fingers and palm subdomains of the pol domain, and is not present in any other known DNA polymerase (Euro et al, 2011)

Biochemical phenotype: Predicted to alter the pol: exo ratio and possibly DNA binding affinity **Description:** To date, no biochemical data are available for residues of the partitioning loop, although it has been shown that yeast strains homozygous for a mutation equivalent to G1051R in human Pol γ cause a 10-fold increase in point mutational frequency *in vivo* (Baruffini et al, 2007).

Cluster 4

Location in structure: These mutations map to the exo domain along the distal accessory subunit interface.

Biochemical phenotype: Increase in exo activity, decrease in pol activity

Description: Biochemical studies of a human R232G variant in reconstituted holoenzyme form showed a decrease in pol rate and an increase in exo activity, with unchanged DNA binding affinity that derives from loss of a direct stimulation of pol activity by the distal accessory subunit (Lee et al, 2010). We propose that the other residues of cluster 4 will have similar biochemical characteristics.

Recessive/ dominant mutation predictions: Recessive

Cluster 5

Location in structure: Located in the periphery of the IP subdomain of the spacer domain, distant from the DNA binding channel

Biochemical phenotype: None observed

Description: Cluster 5 mutations are located in the periphery of the IP subdomain and have not been shown to cause a biochemical defect in Pol γ *per se*. We predict that these mutations may affect as yet undetermined protein-protein interactions *in vivo* (Euro et al, 2011).

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Euro, L., Farnum, G.A., Palin, E., Suomalainen, A. and Kaguni, L.S. (2011) Clustering of Alpers disease mutations and catalytic defects in biochemical variants reveal new features of molecular mechanism of the human mitochondrial replicase, Pol {gamma}. *Nucleic Acids Res.*

2. Ylikallio, E. and Suomalainen, A. (2011) Mechanisms of mitochondrial diseases. *Ann Med.*

3. Kaguni, L.S. (2004) DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem*, 73, 293-320.

4. Yakubovskaya, E., Chen, Z., Carrodeguas, J.A., Kisker, C. and Bogenhagen, D.F. (2006) Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *J Biol Chem*, 281, 374-82.

5. Saneto, R.P. and Naviaux, R.K. (2010) Polymerase gamma disease through the ages. *Dev Disabil Res Rev*, 16, 163-74.

6. Nguyen, K.V., Sharief, F.S., Chan, S.S., Copeland, W.C. and Naviaux, R.K. (2006) Molecular diagnosis of Alpers syndrome. *J Hepatol*, 45, 108-16.

7. Horvath, R., Hudson, G., Ferrari, G., Futterer, N., Ahola, S., Lamantea, E., Prokisch, H., Lochmuller, H., McFarland, R., Ramesh, V. *et al.* (2006) Phenotypic spectrum associated with mutations of the mitochondrial polymerase gamma gene. *Brain*, 129, 1674-84.

8. Lee, Y.S., Kennedy, W.D. and Yin, Y.W. (2009) Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell*, 139, 312-24.

9. Lee, Y.S., Lee, S., Demeler, B., Molineux, I.J., Johnson, K.A. and Yin, Y.W. (2009) Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity. *J Biol Chem*, 285, 1490-9.

10. Singh, K. and Modak, M.J. (2005) Contribution of polar residues of the J-helix in the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I (Klenow fragment): Q677 regulates the removal of terminal mismatch. *Biochemistry*, 44, 8101-10.

11. McCain, M.D., Meyer, A.S., Schultz, S.S., Glekas, A. and Spratt, T.E. (2005) Fidelity of mispair formation and mispair extension is dependent on the interaction between the minor groove of the primer terminus and Arg668 of DNA polymerase I of Escherichia coli. *Biochemistry*, 44, 5647-59.

12. Loh, E. and Loeb, L.A. (2005) Mutability of DNA polymerase I: implications for the creation of mutant DNA polymerases. *DNA Repair (Amst)*, 4, 1390-8.

13. Singh, K. and Modak, M.J. (2003) Presence of 18-A long hydrogen bond track in the active site of Escherichia coli DNA polymerase I (Klenow fragment). Its requirement in the stabilization of enzyme-template-primer complex. *J Biol Chem*, 278, 11289-302.

14. Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. and Loeb, L.A. (2001) Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J Mol Biol*, 308, 823-37.

15. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature*, 391, 251-8.

16. Johnson, A.A., Tsai, Y., Graves, S.W. and Johnson, K.A. (2000) Human mitochondrial DNA polymerase holoenzyme: reconstitution and characterization. *Biochemistry*, 39, 1702-8.

17. Luo, N. and Kaguni, L.S. (2005) Mutations in the spacer region of Drosophila mitochondrial DNA polymerase affect DNA binding, processivity, and the balance between Pol and Exo function. *J Biol Chem*, 280, 2491-7.

18. Szczepanowska, K. and Foury, F. (2010) A cluster of pathogenic mutations in the 3'-5' exonuclease domain of DNA polymerase gamma defines a novel module coupling DNA synthesis and degradation. *Hum Mol Genet*, 19, 3516-29.

19. Graziewicz, M.A., Longley, M.J., Bienstock, R.J., Zeviani, M. and Copeland, W.C. (2004) Structure-function defects of human mitochondrial DNA polymerase in autosomal dominant progressive external ophthalmoplegia. *Nat Struct Mol Biol*, 11, 770-6.

20. Baruffini, E., Ferrero, I. and Foury, F. (2007) Mitochondrial DNA defects in Saccharomyces cerevisiae caused by functional interactions between DNA polymerase gamma mutations associated with disease in human. *Biochim Biophys Acta*, 1772, 1225-35.

21. Lee, Y.S., Johnson, K.A., Molineux, I.J. and Yin, Y.W. (2010) A single mutation in human mitochondrial DNA polymerase Pol gammaA affects both polymerization and proofreading activities of only the holoenzyme. *J Biol Chem*, 285, 28105-16.

22. Luoma, P.T., Luo, N., Loscher, W.N., Farr, C.L., Horvath, R., Wanschitz, J., Kiechl, S., Kaguni, L.S. and Suomalainen, A. (2005) Functional defects due to spacer-region mutations of human mitochondrial DNA polymerase in a family with an ataxia-myopathy syndrome. *Hum Mol Genet*, 14, 1907-20.

23. Ferreira, M., Evangelista, T., Almeida, L.S., Martins, J., Macario, M.C., Martins, E., Moleirinho, A., Azevedo, L., Vilarinho, L. and Santorelli, F.M. (2011) Relative frequency of known causes of multiple mtDNA deletions: two novel POLG mutations. *Neuromuscul Disord*, 21, 483-8.

24. Sato, K., Yabe, I., Yaguchi, H., Nakano, F., Kunieda, Y., Saitoh, S. and Sasaki, H. (2011) Genetic analysis of two Japanese families with progressive external ophthalmoplegia and parkinsonism. *J Neurol*, 258, 1327-32.

25. Tang, S., Dimberg, E.L., Milone, M. and Wong, L.J. (2011) Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)-like phenotype: an expanded clinical spectrum of POLG1 mutations. *J Neurol*.

26. Tang, S., Wang, J., Lee, N.C., Milone, M., Halberg, M.C., Schmitt, E.S., Craigen, W.J., Zhang, W. and Wong, L.J. (2011) Mitochondrial DNA polymerase gamma mutations: an ever expanding molecular and clinical spectrum. *J Med Genet*, 48, 669-81.

27. Brieba, L.G., Eichman, B.F., Kokoska, R.J., Doublie, S., Kunkel, T.A. and Ellenberger, T. (2004) Structural basis for the dual coding potential of 8-oxoguanosine by a high-fidelity DNA polymerase. *EMBO J*, 23, 3452-61.

28. Baruffini, E., Horvath, R., Dallabona, C., Czermin, B., Lamantea, E., Bindoff, L., Invernizzi, F., Ferrero, I., Zeviani, M. and Lodi, T. (2010) Predicting the contribution of novel POLG mutations to human disease through analysis in yeast model. *Mitochondrion*, 11, 182-90.

29. Khan, A., Trevenen, C., Wei, X.C., Sarnat, H.B., Payne, E. and Kirton, A. (2011) Alpers Syndrome: The Natural History of a Case Highlighting Neuroimaging, Neuropathology, and Fat Metabolism. *J Child Neurol*.

30. Li, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of Thermus aquaticus DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J*, 17, 7514-25.

31. Batabyal, D., McKenzie, J.L. and Johnson, K.A. (2010) Role of histidine 932 of the human mitochondrial DNA polymerase in nucleotide discrimination and inherited disease. *J Biol Chem*, 285, 34191-201.

32. Atanassova, N., Fuste, J.M., Wanrooij, S., Macao, B., Goffart, S., Backstrom, S., Farge, G., Khvorostov, I., Larsson, N.G., Spelbrink, J.N. *et al.* (2011) Sequence-specific stalling of DNA polymerase gamma and the effects of mutations causing progressive ophthalmoplegia. *Hum Mol Genet*, 20, 1212-23.

33. Astatke, M., Ng, K., Grindley, N.D. and Joyce, C.M. (1998) A single side chain prevents Escherichia coli DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc Natl Acad Sci U S A*, 95, 3402-7.

34. Spinazzola, A., Invernizzi, F., Carrara, F., Lamantea, E., Donati, A., Dirocco, M., Giordano, I., Meznaric-Petrusa, M., Baruffini, E., Ferrero, I. *et al.* (2009) Clinical and molecular features of mitochondrial DNA depletion syndromes. *J Inherit Metab Dis*, 32, 143-58.

35. Kasiviswanathan, R., Longley, M.J., Chan, S.S. and Copeland, W.C. (2009) Disease mutations in the human mitochondrial DNA polymerase thumb subdomain impart severe defects in mitochondrial DNA replication. *J Biol Chem*, 284, 19501-10.

36. Li, Y., Kong, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Sci*, 7, 1116-23.

37. Foury, F. and Szczepanowska, K. (2011) Antimutator Alleles of Yeast DNA Polymerase Gamma Modulate the Balance between DNA Synthesis and Excision. *PLoS One*, 6, e27847.

38. Suomalainen, A. and Isohanni, P. (2010) Mitochondrial DNA depletion syndromes-many genes, common mechanisms. *Neuromuscul Disord*, 20, 429-37.

39. Srivastava, S. and Moraes, C.T. (2005) Double-strand breaks of mouse muscle mtDNA promote large deletions similar to multiple mtDNA deletions in humans. *Hum Mol Genet*, 14, 893-902.

40. Fukui, H. and Moraes, C.T. (2009) Mechanisms of formation and accumulation of mitochondrial DNA deletions in aging neurons. *Hum Mol Genet*, 18, 1028-36.

41. Wanrooij, S., Goffart, S., Pohjoismaki, J.L., Yasukawa, T. and Spelbrink, J.N. (2007) Expression of catalytic mutants of the mtDNA helicase Twinkle and polymerase POLG causes distinct replication stalling phenotypes. *Nucleic Acids Res*, 35, 3238-51.

42. Wong, L.J., Naviaux, R.K., Brunetti-Pierri, N., Zhang, Q., Schmitt, E.S., Truong, C., Milone, M., Cohen, B.H., Wical, B., Ganesh, J. *et al.* (2008) Molecular and clinical genetics of mitochondrial diseases due to POLG mutations. *Hum Mutat*, 29, E150-72.

43. Stewart, J.D., Tennant, S., Powell, H., Pyle, A., Blakely, E.L., He, L., Hudson, G., Roberts, M., du Plessis, D., Gow, D. *et al.* (2009) Novel POLG1 mutations associated with neuromuscular and liver phenotypes in adults and children. *J Med Genet*, 46, 209-14.

44. Palin, E.J., Lesonen, A., Farr, C.L., Euro, L., Suomalainen, A. and Kaguni, L.S. (2010) Functional analysis of H. sapiens DNA polymerase gamma spacer mutation W748S with and without common variant E1143G. *Biochim Biophys Acta*, 1802, 545-51.

45. Uusimaa, J., Hinttala, R., Rantala, H., Paivarinta, M., Herva, R., Roytta, M., Soini, H., Moilanen, J.S., Remes, A.M., Hassinen, I.E. *et al.* (2008) Homozygous W748S mutation in the POLG1 gene in patients with juvenile-onset Alpers syndrome and status epilepticus. *Epilepsia*, 49, 1038-45.

46. Hakonen, A.H., Heiskanen, S., Juvonen, V., Lappalainen, I., Luoma, P.T., Rantamaki, M., Goethem, G.V., Lofgren, A., Hackman, P., Paetau, A. *et al.* (2005) Mitochondrial DNA polymerase W748S mutation: a common cause of autosomal recessive ataxia with ancient European origin. *Am J Hum Genet*, 77, 430-41.

47. Winterthun, S., Ferrari, G., He, L., Taylor, R.W., Zeviani, M., Turnbull, D.M., Engelsen, B.A., Moen, G. and Bindoff, L.A. (2005) Autosomal recessive mitochondrial ataxic syndrome due to mitochondrial polymerase gamma mutations. *Neurology*, 64, 1204-8.

48. Van Goethem, G., Luoma, P., Rantamaki, M., Al Memar, A., Kaakkola, S., Hackman, P., Krahe, R., Lofgren, A., Martin, J.J., De Jonghe, P. *et al.* (2004) POLG mutations in neurodegenerative disorders with ataxia but no muscle involvement. *Neurology*, 63, 1251-7.

49. Milone, M., Benarroch, E.E. and Wong, L.J. (2011) POLG-related disorders: Defects of the nuclear and mitochondrial genome interaction. *Neurology*, 77, 1847-52.

50. Blok, M.J., van den Bosch, B.J., Jongen, E., Hendrickx, A., de Die-Smulders, C.E., Hoogendijk, J.E., Brusse, E., de Visser, M., Poll-The, B.T., Bierau, J. *et al.* (2009) The unfolding clinical spectrum of POLG mutations. *J Med Genet*, 46, 776-85.

51. Stewart, J.D., Horvath, R., Baruffini, E., Ferrero, I., Bulst, S., Watkins, P.B., Fontana, R.J., Day, C.P. and Chinnery, P.F. (2010) Polymerase gamma gene POLG determines the risk of sodium valproate-induced liver toxicity. *Hepatology*, 52, 1791-6.