

In the format provided by the authors and unedited.

# Mitochondrial complex I structure reveals ordered water molecules for catalysis and proton translocation

Daniel N. Grba  and Judy Hirst  

---

The Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK. ✉e-mail: [jh@mrc-mbu.cam.ac.uk](mailto:jh@mrc-mbu.cam.ac.uk)

**Supplementary Table 1. Build summary of subunits in our *Y. lipolytica* complex I model and the homologous subunits of *M. musculus* (PDB 6G2J)<sup>20</sup>**

Subunit	Human/ bovine nomenclature	Chain	Total residues (mature protein numbering)	Residues modelled [%] (range built)	Mammalian homology reference	Residues without side chains	Modelled with side chains [%]
NUAM	NDUFS1/ 75 kDa	G	694 (35–728)	100 (35–728)	<sup>11</sup>	0	100
NUBM	NDUFV1/ 51 kDa	F	470 (19–488)	98 (29–488)	<sup>11</sup>	0	100
NUCM	NDUFS2/ 49 kDa	D	444 (23–466)	98 (30–466)	<sup>11</sup>	0	100
NUGM	NDUFS3/ 30 kDa	C	263 <sup>a</sup> (31–281)	92 (33–274)	<sup>11</sup>	0	100
NUHM	NDUFV2/ 24 kDa	E	216 (28–243)	100 (28–243)	<sup>11</sup>	0	100
NUIM	NDUFS8/ TYKY	I	198 (32–229)	96 (39–229)	<sup>11</sup>	0	100
NUKM	NDUFS7/ PSST	B	183 (28–210)	97 (34–210)	<sup>11</sup>	0	100
NU1M	NU1M/ ND1	H	341 (1–341)	100 (1–341)	<sup>11</sup>	215–217	99
NU2M	NU2M/ ND2	N	469 (1–469)	100 (1–469)	<sup>11</sup>	0	100
NU3M	NU3M/ ND3	A	128 (1–128)	100 (1–128)	<sup>11</sup>	38, 42–46	95
NU4M	NU4M/ ND4	M	486 (1–486)	100 (1–486)	<sup>11</sup>	0	100
NU5M	NU5M/ ND5	L	655 (1–655)	100 (1–655)	<sup>11</sup>	609–612	99
NU6M	NU6M/ ND6	J	185 (1–185)	100 (1–185)	<sup>11</sup>	0	100
NULM	NULM/ ND4L	K	89 (1–89)	100 (1–89)	<sup>11</sup>	0	100
NUEM	NDUFA9/ 39 kDa	P	355 (21–375)	100 (21–375)	<sup>11</sup>	353–364	97
NUFM	NDUFA5/ B13	V	136 (9–144)	93 (19–144)	<sup>11</sup>	0	100
NUMM	NDUFS6/ 13 kDa	R	118 (19–136)	100 (19–136)	<sup>11</sup>	0	100
NUYM	NDUFS4/ 18 kDa/AQDQ	Q	137 (25–161)	91 (37–161)	<sup>11</sup>	0	100
NUZM	NDUFA7/ B14.5a	c, r ( <i>Mm</i> )	182 (1–182)	100 (1–182)	<sup>12</sup>	0	100
N7BM	NDUFA12/ B17.2	f, q ( <i>Mm</i> )	137 (2–138)	100 (2–138)	<sup>11</sup>	0	100
NB4M	NDUFA6/ B14	W	123 (2–124)	100 (2–124)	<sup>11</sup>	0	100
ACPM1	NDUFAB1/ SDAP	T	84 (26–109)	96 (29–109)	<sup>11</sup>	0	100
NI8M	NDUFA2/ B8	S	86 (2–87)	100 (2–87)	<sup>11</sup>	0	100
NUPM	NDUFA8/ PGIV	X	171 (2–172)	100 (2–172)	<sup>11</sup>	0	100
NUJM	NDUFA11/ B14.7	Y	185 (14–198)	97 (20–198)	<sup>11</sup>	0	100
NB6M	NDUFA13/ B16.6	Z	122 (2–123)	100 (2–123)	<sup>11</sup>	0	100
NIPM	NDUFS5/ 15 kDa/PFFD	e	88 (2–89)	77 (2–69)	<sup>11</sup>	0	100
NUXM	Absent from <i>Mm</i>	O	168 (2–169)	100 (2–169)	<sup>11</sup>	0	100

NI9M	NDUFA3/ B9	b	78 (1–78)	100 (1–78)	<sup>11</sup>	0	100
NIMM	NDUFA1/ MWFE	a	86 (2–87)	100 (2–87)	<sup>11</sup>	0	100
NEBM	NDUFC2/ B14.5b	d	73 (2–74)	92 (2–68)	Structural homologue <sup>c</sup> <sub>35</sub>	0	0
ST1	Absent from <i>Mm</i>	q (PDB 6YJ5)	314 (2–315)	97 (2–307)	<sup>11</sup>	30–36, 45, 54–67, 139, 144–149, 169–192, 197, 200– 225, 234– 236–244, 246–254, 292–307	63
NESM	NDUFB11/ ESSS	g	204 (46–249)	97 (46–243)	<sup>11</sup>	75–85	94
NIAM	NDUFB8/ ASHI	l	125 (25–149)	100 (25–149)	<sup>11</sup>	0	100
NUNM	NDUFB5/ SGDH	h	119 (32–150)	95 (38–150)	Structural homologue <sup>c</sup> <sub>35</sub>	0	100
NB2M	NDUFB3/ B12	k	59 (2–60)	80 (2–48)	<sup>11</sup>	0	100
NB5M	NDUFB4/ B15	m	92 (2–93)	100 (2–93)	<sup>11</sup>	0	100
NB8M	NDUFB7/ B18	o	98 (2–98)	85 (2–84)	<sup>11</sup>	0	100
ACPM2	NDUFAB1/ SDAP	U	88 (45–132)	100 (45–132)	<sup>11</sup>	0	100
NIDM	NDUFB10/ PDSW	p	91 (2–92)	100 (2–92)	<sup>11</sup>	0	100
NI2M	NDUFB9/ B22	n	108 (2–109)	100 (2–109)	<sup>11</sup>	0	100
NIGM	NDUFB2/ AGGG	j	91 <sup>b</sup> (31–83) 67 <sup>c</sup> (1–67)	58 (31–83) 79 (7–59)	Structural homologue <sup>d</sup> <sub>19</sub>	0	100
NUUM	NDUFB6/ B17	i	89 (2–90)	97 (2–87)	Same position in structure <sup>e</sup>	0	100
	NDUFV3/10 kDa	s ( <i>Mm</i> )			Absent from <i>Yl</i>		
	NDUFA10/42 kDa	O ( <i>Mm</i> )			Absent from <i>Yl</i>		
	NDUFC1/ KFYI	c ( <i>Mm</i> )			Absent from <i>Yl</i>		
	NDUFB1/ MNLL	f ( <i>Mm</i> )			Absent from <i>Yl</i>		

<sup>a</sup>Includes the hexa-Ala-hexa-His tag; <sup>b</sup>Includes the MTS predicted by MitoFates<sup>64</sup>; <sup>c</sup>The mature protein predicted by Parey and coworkers,<sup>19</sup> as used in the final model; <sup>d</sup>Little sequence homology but excellent secondary structure overlap; <sup>e</sup>Little sequence homology and some secondary structure overlap (situated at equivalent site in model). MTS-cleavage sites were assumed from previously published information<sup>62,63</sup>.

**Supplementary Table 2. Reference table for the numbering of the water molecules used in this manuscript, relative to the system used in the model (PDB 6YJ4).**

Manuscript water number	Numbering in 6YJ4	Nearest Subunit	Subunit Chain
1	612	NDUFS2	D
2	614	NDUFS2	D
3	605	NDUFS2	D
4	602	NDUFS2	D
5	407	NDUFS7	B
6	413	NDUFS7	B
7	401	NDUFS7	B
8	705	ND1	H
9	408	NDUFS7	B
10	706	ND1	H
11	704	ND1	H
12	405	NDUFS7	B
13	710	ND1	H
14	202	ND6	J
15	602	ND3	A
16	103	ND4L	K
17	2514	ND2	N
18	102	ND4L	K
19	2511	ND2	N
20	2513	ND2	N
21	2501	ND2	N
22	2515	ND2	N
23	2507	ND2	N
24	2535	ND2	N
25	2526	ND2	N
26	2536	ND2	N
27	2523	ND2	N
28	2527	ND2	N
29	2537	ND2	N
30	2512	ND2	N
31	608	ND4	M
32	612	ND4	M
33	603	ND4	M
34	610	ND4	M
35	604	ND4	M
36	606	ND4	M
37	602	ND4	M
38	1125	ND5	L
39	1106	ND5	L
40	1116	ND5	L
41	1102	ND5	L
42	1123	ND5	L
43	1115	ND5	L

**Supplementary Table 3. Comparison of the  $\pi$  bulges detected in structures of complex I from different species**

	$\pi$ bulge	<i>M. musculus</i> (active)	<i>M. musculus</i> (deactive)	<i>H. sapiens</i>	<i>S. scrofa</i>	<i>Y. lipolytica</i>	<i>E. coli</i>	<i>T. thermophilus</i>	<i>T. elongatus</i>
		6G2J <sup>20</sup>	6G72 <sup>20</sup>	5XTC <sup>44</sup>	5GPN <sup>45</sup>	6YJ4	3RKO <sup>28</sup>	4HEA <sup>26</sup>	6NBQ <sup>46</sup>
ND1	$\pi$ 1	289–295	288–293 288–292	287–293 286–293	288–293	311–315	n/a		344–350
	$\pi$ 2					204–208	n/a		
	$\pi$ 3						n/a	189–193	
	$\pi$ 4						n/a	211–215	
	$\pi$ 5						n/a	276–280	
ND2	$\pi$ 1			129–133			242–249	211–215	
	$\pi$ 2	298–302							
	$\pi$ 3		325–335						
	$\pi$ 4								261–267
	$\pi$ 5								286–292
	$\pi$ 6								430–434
ND4	$\pi$ 1			231–235		246–250	259–263		250–254
	$\pi$ 2						271–275 272–277	241–245	264–270
	$\pi$ 3	132–136	132–136						
ND5	$\pi$ 1	245–251 247–251	245–251 246–251	244–251 <sup>b</sup>	244–251	248–252 250–254	254–258 252–257	239–245	264–274 269–274
	$\pi$ 2	560–565 560–565	546–560 <sup>1</sup> 560–564	560–564		561–566 561–566	547–553 546–550		614–618
	$\pi$ 3	546–552 553–557	546–560 <sup>a</sup> 550–557	546–557 548–554		576–582 574–583			
	$\pi$ 4 <sup>c</sup>	427–431	427–431 426–430	428–432		434–438			
	$\pi$ 5	362–366							
	$\pi$ 6	484–488							
	$\pi$ 7								491–495
	$\pi$ 8			389–393					
ND6	$\pi$ 1		56–63 59–63	57–63	56–61	60–68 59–67	58–63	63–70 55–63	61–68 63–70
	$\pi$ 2					154–158 153–157			

Black: results from ASSP program<sup>74</sup>; red: results from DSSP<sup>75</sup>. Shaded rows contain  $\pi$  bulges that were detected in more than one organism. <sup>a</sup>Region containing multiple  $\pi$  bulges. <sup>b</sup> $\pi$  bulge not detected in the published model, but revealed following local remodelling using Coot. <sup>c</sup>On the periphery of the subunit so unlikely to be functionally relevant.

**Supplementary Table 4. The structural elements representing decreasing activation levels in the active mouse, ubiquinone-bound *Y. lipolytica*, DDM-bound *Y. lipolytica*, and deactive mouse complex I models**

	<b>Active mouse PDB 6G2J<sup>20</sup></b>	<b>Ubiquinone-bound <i>Y. lipolytica</i> PDB 6RFR<sup>19</sup></b>	<b>DDM-bound <i>Y. lipolytica</i> PDB 6YJ4</b>	<b>Deactive mouse PDB 6G72<sup>20</sup></b>
	<b>Active</b>	<b>Active/deactive hybrid</b>	<b>Mild deactive</b>	<b>Deactive</b>
<b>ND1-TMH4</b>	bent	straight	straight	straight
<b>ND1-TMH5–6 loop</b>	ordered	ordered but shifted compared to 6G2J	shifted and partially disordered (residues 215–217 no side chain features)	disordered (residues 204–215 absent from the model)
<b>ND1 cavity</b>	relatively small and closed to matrix	expanded due to TMH4 movement but not matrix exposed	expanded due to TMH4 movement and matrix exposed	expanded due to TMH4 movement and matrix exposed
<b>ND3-TMH2–3 loop</b>	ordered	ordered	shifted and partially disordered (residues 38, 42–46 no side chain features)	disordered (residues 28–48 absent from the model)
<b>ND3-Cys40(<i>Mm</i>39)-binding pocket</b>	present	destabilised due to straightening of ND1-TMH4	absent	absent
<b>ND6-TMH3 <math>\pi</math> bulge</b>	absent	present	present	present
<b>NDUFS2-<math>\beta</math>1–<math>\beta</math>2 loop</b>	ordered	ordered	flexible (resolved but to a lower resolution than its surroundings)	disordered (residues 52–60 absent from the model)