A model for Batten disease protein CLN3: Functional implications from homology and mutations

Robert W. Janes[1,a,*], Patricia B. Munroe[b], Hannah M. Mitchison[b], R. Mark Gardiner[b], Sara E. Mole[b], B.A. Wallace[a]

[a]*Department of Crystallography, Birkbeck College, University of London, London WC1E 7HX, UK*
[b]*Department of Paediatrics, University College London Medical School, The Rayne Institute, London WC1E 6JJ, UK*

Received 10 September 1996; revised version received 21 October 1996

Abstract In an attempt to understand the molecular nature of Batten disease, we have examined the amino acid sequence of the affected *CLN3* gene product (The International Batten Disease Consortium (1995) Cell 82, 949–957) and the site-specific mutations which give rise to the biological defect. Homology searches and molecular modeling have led to the development of a model for the folding and disposition of the protein, possibly within a mitochondrial membrane. High homology with a yeast protein of unknown function suggests a strong evolutionary conservation of function. We speculate that a possible role for the protein may be in chaperoning the folding/unfolding or assembly/disassembly of other proteins, specifically subunit c of the mitochondrial ATP synthase complex.

Key words: Molecular modeling; Chaperone; Genetic disease; Membrane protein; Neurodegenerative disease

1. Introduction

Batten disease (juvenile onset ceroid lipofuscinosis [JNCL]) is one type of a group of autosomal recessive inherited neurodegenerative disorders characterized by the accumulation of autofluorescent lipopigment (ceroid and lipofuscin) in many tissues [2]. The clinical features include visual failure, seizures and psychomotor deterioration leading to a vegetative state and premature death in the third decade. The main component of the storage material deposited is subunit c of the mitochondrial ATP synthase complex [3,4]. The gene defective in Batten disease, *CLN3*, has recently been identified [1] and encodes a predicted protein of 438 amino acids. To rationalize the possible structural basis of the protein defects underlying Batten disease, we have constructed a model of the CLN3 protein (Fig. 1).

2. Materials and methods

The sequence homology studies were performed using the BLAST [5] algorithm. Sequence alignment with the yeast protein, SWISS-PROT [6] entry YJF9_YEAST, was performed using Multalin [7] with a gap weight of 8 and gap length weight of 0. PSORT [8] was used to identify mitochondrial recognition and cleavage site data. The FOAMV86 [9] package, using the GES [9], Kyte and Doolittle [10], and Wolfenden [11] scales, each with a window size of 20, was used to produce hydrophobicity plots for visual analysis. The scanPROSITE [12] tool was used to find motifs consistent with *N*-glycosylation sites, and potential phosphorylation sites by cAMP- and cGMP-dependent protein kinase, protein kinase c, and casein kinase II. NetOglyc [13] was used to find motifs consistent with *O*-glycosylation sites. All other searches were performed manually.

3. Results and discussion

The first aim was to establish whether the CLN3 protein was a membrane or soluble protein. Calculation of the overall average hydrophobicity of the sequence using the GES [9] and Eisenberg [14] scales, gave values of 0.256 and 0.253, respectively. Using the same scales, myoglobin, a typical soluble protein, produced values of −0.094 and 0.008, and bacteriorhodopsin, an extremely hydrophobic membrane protein, gave 0.315 and 0.363. Thus, the CLN3 protein is significantly hydrophobic in character, consistent with it being an integral membrane protein. Hydrophobicity plots using GES, Kyte and Doolittle [10], and Wolfenden [11] scales suggest the existence of six transmembrane segments (TMS), as shown in Fig. 1 and Table 1. The sequence appears to have a mitochondrial targeting site at residue 11, with a cleavage site at residue 19 for mitochondrial proteins, which suggests that CLN3 is associated with a mitochondrial membrane. As expected for a realistic model, the predicted *N*-glycosylation sites at residues 49, 71, 85, and 310, were found in our model structure to lie only on one surface (Fig. 1), a surface which could be considered topologically equivalent to the exoplasmic surface. In this paper we will refer to this as the 'exo' surface. Also consistent with a reasonable model, none of the predicted phosphorylation sites were found to be in a TMS (Fig. 1). Two putative *O*-glycosylation sites exist at residues 80 and 256, and our model placed one of these on the N-terminal exo surface and the other on the opposite face, in a 'cyto' loop. However, as *O*-glycosylation sites are poorly defined and not necessarily utilized [15], this is not considered a significant problem. There is only a single cysteine on the cyto surface; all other extramembranous cysteine residues are on the exo loops. This means that no disulphide bonds could be formed on the cyto surface of the protein although some could form on the exo surface, again consistent with the expectations for a correct model. There are seven positively charged residues in the cyto loops, but only six in the exo loops, which is consistent with the general observation of a more positively charged cyto surface. Furthermore, for mitochondrial proteins such a charge asymmetry is not expected to be very pronounced, as is the case here. Tryptophan residues within the TMS of our model are at, or very near to, the ends of the transmembrane domains, which is consistent with known X-ray structures of

*Corresponding author. Fax: (44) (171) 631-6803.
E-mail: ubcg06w@ccs.bbk.ac.uk

[1]*Present address:* School of Biological Sciences, Queen Mary and Westfield College, University of London, London E1 4NS, UK.

PII S0014-5793(96)01290-2

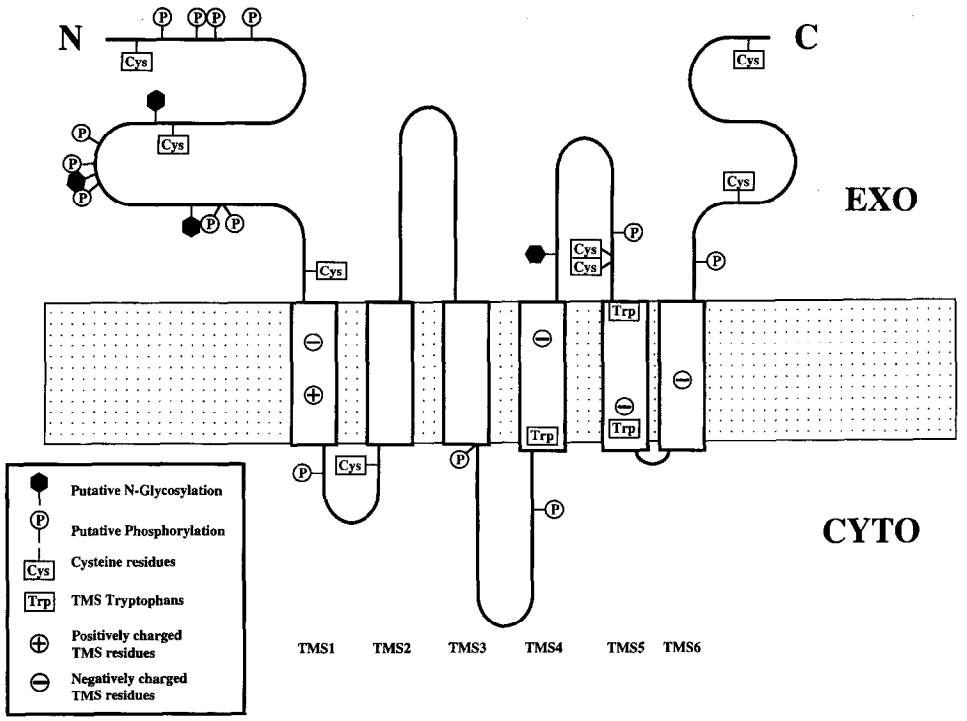

Fig. 1. Schematic model for the putative structural arrangement of CLN3 Batten disease protein showing the six transmembrane segments (TMS) and highlighting locations of potential extramembranous *N*-glycosylation and phosphorylation sites and Cys residues, and the positions of Trp and charged residues within the TMS.

proteins containing TMS. Five potentially charged residues fall within the TMS of our model, which at first would appear to be energetically unfavourable, but their spatial arrangements are such that ion-pairs could be formed between Lys^{112} (TMS1), and either Asp^{362} (TMS5), or Glu^{381} (TMS6), thus leaving only three unpaired charges. It is notable that if either of the proposed salt bridges is formed, this would impose constraints on the 3-dimensional folding of the polypeptide.

A homology search has identified a yeast protein in *S. cerevisiae*, (SWISS-PROT [6] entry YJF9_YEAST) which has 36% identity and 56% similarity with the CLN3 protein (Fig. 2). This is the only protein of known sequence in the public databases to date with which CLN3 has significant homology, and is a further example of the trend [16] of correspondence between yeast and human disease genes. A subsequent examination of the alignments of the individual N-termini, surface loops, TMS and C-termini domains, of the CLN3 and yeast proteins showed that the entire CLN3 and yeast proteins possess a high level of homology, with slightly higher homology in the exo loops and in three of the six TMS. One noteworthy feature in this case is that all of the known single point mutation sites which have been found in the CLN3 protein of Batten disease patients thus far (Munroe, P.B. et al., in preparation), at residues 101, 170, 295, 330, and 334, are found at sites which have identical residues in the yeast protein, three of which are in TMS, and two in exo loops. Of the five TMS charges found in the CLN3 protein, four of them are found in the yeast protein. That the equivalent residue to Asp^{362} of the CLN3 protein is not a charged residue in the yeast protein, suggests that the charged pairing in the CLN3 protein involves TMS1 and TMS6, and not TMS5.

Finally, it is of note that no significant homology is found with any prokaryotic protein known to date, suggesting a function that is conserved and specific to eukaryotes. The slow rate of evolutionary divergence between the CLN3 and yeast protein suggests it is involved in interactions with other proteins. The percentage identity is comparable to that found for other pairs of human/yeast proteins which are involved in assembly processes. Therefore, we postulate that amongst other possibilities, the function of the CLN3 protein could be as a chaperone involved in the folding or unfolding pathways of subunit c of the ATP synthase complex. This is consistent with recent biochemical evidence suggesting a chaperoning role for the Batten disease gene product [17]. The residues in CLN3 which show mutations in Batten disease patients, and are identical to residues in the yeast protein, clearly play an important role in the function and/or structural integrity of the CLN3 protein, and would indicate the

Table 1
Homology analysis of the individual domains of CLN3 compared with the yeast protein YJF9_YEAST

Domain	Residues	Percentage identity
N-terminal	1–97	22 (21/97)
TMS1	98–119	55 (12/22)
Loop 1	120–138	32 (6/19)
TMS2	139–160	45 (10/22)
Loop 2	161–212	52 (27/52)
TMS3	213–234	18 (4/22)
Loop 3	235–280	20 (9/46)
TMS4	281–303	57 (13/23)
Loop 4	304–345	40 (17/42)
TMS5	346–368	30 (7/23)
Loop 5	369–370	50 (1/2)
TMS6	371–393	61 (14/23)
C-terminal	394–438	40 (18/45)

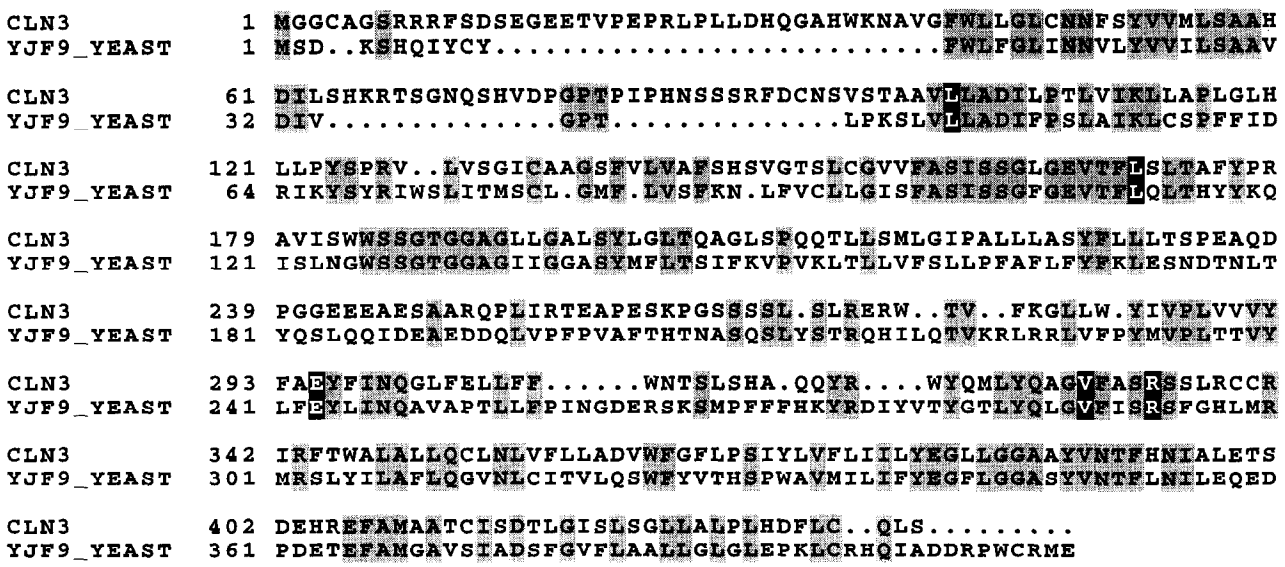

Fig. 2. Sequence alignment between the CLN3 Batten disease protein and the yeast protein, code YJF9_YEAST in SWISS-PROT. The identical residues between the proteins are highlighted by a grey background while the CLN3 protein point mutation sites found in Batten disease patients are shown with a black background.

CLN3 protein interacts with other molecules via its exo surface.

Acknowledgements: The work at Birkbeck College was supported in part by a Biomolecular Sciences Centre grant from the BBSRC. The work at University College London Medical School was supported by grants from the U.S. Public Health Service National Institutes of Health (Grant No. NS28722), the MRC (UK), and the Research Trust for Metabolic Diseases in Childhood (UK).

References

[1] The International Batten Disease Consortium (1995) Cell 82, 949–957.
[2] Santavuori, P. (1988) Brain Dev. 10, 80–83.
[3] Fearnley, I.M., Walker, J.E., Martinus, R.D., Jolly, R.D., Kirkland, K.B., Shaw, G.J. and Palmer, D.N. (1990) Biochem. J. 268, 751–758.
[4] Palmer, D.N., Fearnley, I.M., Walker, J.E., Hall, N.A., Lake, B.D., Wolfe, L.S., Haltia, M., Martinus, R.D. and Jolly, R.D. (1992) Am. J. Med. Genet. 42, 5641–567.
[5] Altschul, S.F, Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[6] Bairoch, A. and Boeckmann, B. (1994) Nucl Acids Res. 22, 3578–3580.
[7] Corpet, F. (1988) Nucl. Acids Res. 16, 10881–10890.
[8] Nakai, K. and Kanehisa, M. (1992) Genomics 14, 897–911.
[9] Engelman, D.M., Steitz, T.A. and Goldman A. (1986) Annu. Rev. Biophys. Chem. 15, 321–353.
[10] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[11] Wolfenden, R. (1983) Science 222, 1087–1093.
[12] scanPROSITE, a world wide web tool utility, current address:http://expasy.hcuge.ch/sprot/scnpsite.html
[13] Hansen, J.E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J.O. Hansen, J.E.S. and Brunak, S. (1995) Biochem. J. 308, 801–813.
[14] Eisenberg, D. (1982) Faraday Symp. Chem. Soc. 17, 109–120.
[15] Creighton, T.E. (1993) in: Proteins: Structures and Molecular Properties, Freeman, New York.
[16] Bassett, D.E., Jr, Boguski, M.S. and Hieter, P. (1996) Nature 379, 589–590.
[17] Tanner, A.J. and Dice, J.F. (1996) Biochem. Mol. Med. 57, 1–9.