The Role of Protein Modelling in Predicting the Disease Severity of Cystinuria

Kathie A Wong[a,1,*], Mark Wass[b,1], Kay Thomas[a]

A Urology Centre, Guy's and St. Thomas' NHS Foundation Trust, London, SE1 9RT, UK
B School of Biosciences, University of Kent, Canterbury Kent, UK
*Corresponding author. Urology Centre, Guy's and St. Thomas' NHS Foundation Trust, London SE1 9RT, UK. Tel. +44 79 12225938; Fax: +44 20 71997188.
E-mail address: kathie.wong@gmail.com (K.A. Wong).

1 These authors are joint first authors.

Mutations in one of two genes SLC3A1 and SLC7A9 are responsible for the majority of cystinuria phenotypes. A defective renal tubular protein transporter causes high urinary levels of *cystine* and the dibasic amino acids*, lysine, arginine and ornithine.* Cystine is relatively insoluble in urine and forms stones. We have found 57 different mutations in our UK cohort[1]. Most are missense mutations and it is unclear what effect they have on protein function and how this translates to phenotype.

The aim of this study was to use protein modelling to investigate how missense mutations may affect protein function. Limited experimental data is available as experimental techniques are time-consuming and results would lag significantly behind the rate at which new mutations are being reported. Understanding how individual mutations can cause protein dysfunction could allow us to predict a patient's disease severity and tailor individual management more effectively. We modelled the b(0+)AT protein encoded by SLC7A9 using the Phyre2 web server, and other validated software [2-5].

The missense mutations were scored by an investigator blind to the clinical data into those predicted to cause a low/low-medium effect=1 or high/high-medium effect=2. This was based on several factors including the proximity of the mutation to the predicted functional sites and size of conformational change. Large genomic rearrangements were assumed to cause significant protein dysfunction therefore all other (non-missense) mutations were

assigned a score of 2. An overall severity score was calculated for 26 patients based on the sum of the score of each individual mutation. For example, a patient with a predicted low-effect missense mutation and a frameshift mutation would score 3 (1+2).

When comparing patients with a score of 4 versus a score of 3, there was no difference between the levels of cystine(201 µmol/mmolCr IQR 161 to 231 vs 154 µmol/mmolCr IQR 133.7 to 191.3, p=0.2545) or lysine (629.5 µmol/mmolCr IQR 593.2 to 814.6 vs 569.5 µmol/mmolCr IQR 403.6 to 807.5, p=0.2887). Patients who scored 4 had higher levels of *arginine*(383 µmol/mmolCr IQR 283 to 392.7 vs 70 µmol/mmolCr IQR 36.38 to 237.3, p=0.0151) and *ornithine*(120.2 µmol/mmolCr IQR 97.67 to 153.1 vs 94 µmol/mmolCr IQR 68.38 to 111, p=0.0482) than patients who scored 3. They also experienced a higher number of stone episodes(0.5/yr IQR 0.0 to 1.0 vs 0.0/yr IQR 0 to 0.3, p=0.0451). Only three patients scored less than 3, precluding statistical analysis.

Our results suggest our model may help determine a patient's phenotype. The lack of statistical difference for cystine and lysine may be explained by limitations in cystine measurements and other transporter mechanisms involved in lysine transport.

Clearly, a patient's genotype can only determine part of a patient's disease profile, which may also be influenced by modifier genes and complex genetic and environmental interactions. More collaborative work is needed to explore

our hypothesis. This approach has given us insight into how the different missense mutations may cause the range of phenotypes seen in Cystinuria and is a step closer to a personalised approach to the management of these patients.

References

[1] Wong KA, Mein R, Wass M, et al. The Genetic Diversity of Cystinuria in a UK Population of Patients. BJU Int. 2014 Aug 11:

[2] Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010 Jul: **38**:W529-33

[3] Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res. 2010 Jul: **38**:W469-73

[4] Lopez G, Maietta P, Rodriguez JM, Valencia A, Tress ML. firestar--advances in the prediction of functionally important residues. Nucleic Acids Res. 2011 Jul: **39**:W235-41

[5] Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun: **10**:845-58

Figures

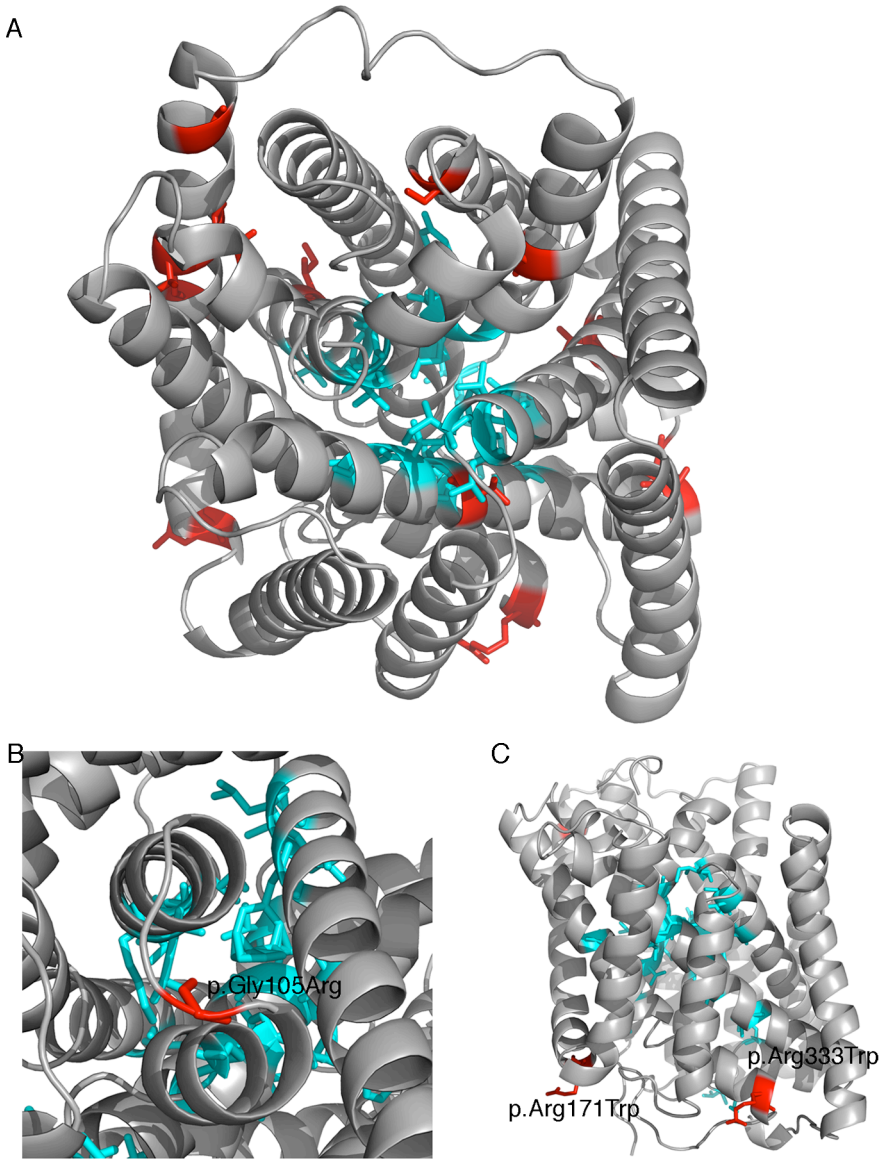

Figure 1 Structural analysis of mutations present in b(0+)AT. A) A structural model of b(0+)AT was generated and residues involved in amino acid transport identified (cyan) and mutations mapped on (red) to analyse their potential structural and functional effect. B) p.Gly105Arg is located at the end of the channel where amino acids are transported. C) Some mutations (p.Arg171Trp, p.Arg333Trp) are located close to the end of the membrane and could affect stability in the membrane.