Supplementary Materials

Large-scale In Vitro Functional Testing and Novel Variant Scoring via Protein Modeling Provide Insights into Alkaline Phosphatase Activity in Hypophosphatasia

Guillermo del Angel; John Reynders; Christopher Negron; Thomas Steinbrecher; Etienne Mornet

CONTENTS

METHODS

RESULTS

SUPPLEMENTARY TABLES (Excel spreadsheet)

METHODS

ALPL variant selection

Variants to be tested were generated using *ALPL* canonical RefSeq cDNA transcript NM_000478.6 (equivalent to Ensembl transcript ENST00000374840.7). Missense, frameshift, in-frame indels, and nonsense *ALPL* variants were collected as follows:

- 65 variants present in both the Genome Aggregation Database (gnomAD) (Karczewski et al., 2019) and on either the HPP patient cohort or on the HPP Gene Mutations Database.
- 36 variants present only in gnomAD but absent in HPP patients.
- 51 variants present in the HPP patient cohort only but absent in gnomAD
- 3 variants from other sources reported to the Unité de Génétique Constitutionnelle Prénatale et Postnatale (Centre Hospitalier de Versailles, Le Chesnay, France) as potentially pathogenic.

Plasmid preparation, transfection, and TNSALP activity measurement

Cells to be transfected were seeded the day before transfection at a cell density corresponding to 60,000 cells/cm^2 in 24 well plates (wells of 2 cm^2 with 500 µL of Dulbecco's Modified Eagle's Medium + 10% fetal calf serum) and incubated at 37°C with 5% of CO_2. The next day, the medium was replaced with 500 µL of fresh medium, and the cells were transiently transfected with 100% of mutant plasmid or a mix corresponding to mutant plasmid: wild-type (WT) plasmid (ratio 50:50). A plasmid expressing β-galactosidase was co-transfected in all conditions to allow normalization of the *ALPL* signal. All transfections were performed in MDCK II cells. The DNA:Lipofectamine 3000 ratio was 1:3 (1 µg DNA plasmids/3 µL Lipofectamine 3000), and the transfection protocol was followed according to manufacturer specifications (ThermoFisher Scientific, Regensburg, Germany).

Forty-eight hours after transfections, cells were lysed and centrifuged, and the lysate was transferred into 96-well, round-bottom dilution plates. β-galactosidase or TNSALP substrates were added to the plates, sealed, and incubated at 37°C. p-Nitrophenyl

phosphate (PNPP) was used as the TNSALP substrate. Absorbance at 405 nm was read for both *ALPL* and β-galactosidase.

Predicting in vitro activity of patient genotypes

For compound heterozygous patients, the rationale for the estimate of TNSALP residual activity for patients with a dominant negative effect is that about 25% of formed homodimers are expected to be composed of 2 copies of allele 1, and about 50% of formed homodimers would be composed of a copy of allele 1 and a copy of allele 2. In either case, the dominant negative effect of allele 1 would render resulting dimers incorporating allele 1 inoperative, thus leaving only the 25% of alleles composed of two copies of allele 2 operational.

TNSALP homology model and variant scoring based on stability/affinity changes

A protein homology model for human TNSALP was built with the Prime software package version 2016-2 (Schrödinger, Inc., New York, NY, USA) based on the methods described by Jacobson et al (Jacobson, Friesner, Xiang, & Honig, 2002; Jacobson et al., 2004). The *PDB* homolog structure *1EW2* was used as a template, which corresponds to the human Placental ALP (*PLAP*), encoded by amino acid sequence UniProt P05187-1 (Le Du, Stigbrand, Taussig, Menez, & Stura, 2001). The *ALPL* amino acid sequence NP_000469.3 (equivalent to UniProt P05186-1) was used as query sequence.

The template chosen from a Basic Local Alignment Search Tool (BLAST) had a sufficiently high sequence identity to ALP (57%) and is an x-ray crystal structure of high resolution (1.8Å). The first 17 amino acids on the sequence, corresponding to the signal peptide and not forming part of the final TNSALP dimeric structure, were removed from the model. Similarly, the last 22 amino acids from the sequence, corresponding to the C-terminus in structure *1EW2*, were also omitted, as they were unstructured in the original template (i.e., electron density was absent in the crystal structure). The template structure was obtained as a dimer using the *PDB* biologic assembly download option.

The structure was prepared for residue scanning using Schrödinger's Protein Preparation Wizard (Sastry, Adzhigirey, Day, Annabhimoju, & Sherman, 2013). N-acetyl-glucosamine molecules were removed from the template structure, but all ions were left in place. Phosphate ions were modelled as protonated with hydrogen positions manually adjusted to allow for metal ion coordination. Homology models were built using the Structure Prediction Wizard in Prime, using default settings throughout in the Homo-Multimer mode to build both chain A and B simultaneously. ClustalW was used for sequence alignment, and all phosphate and metal ion cofactors were left in place. Prime loop refinement was conducted on modeled loops.

Each possible *ALPL* missense variant was generated, and its effect on protein stability and dimer binding affinity were predicted using MM-GBSA, a physics-based approach to measuring changes in protein stability (Beard, Cholleti, Pearlman, Sherman, & Loving, 2013), as implemented in the BioLuminate software package (Schrödinger, Inc., New York, NY, USA). MM-GBSA predicts the relative change in folding free energy of an amino acid variant using an implicit solvent model integrated with a molecular mechanics force-field energy function. The method details and its application and assessment of a nonhuman system with a large amount of validation data is described in Negron et al (Negron, Pearlman, & Del Angel, 2019).

$\Delta_s(a_i)$, $\Delta_a(a_i)$, and $\Delta_{max}(a_i)$ were computed for all possible TNSALP missense variants using the residue scanning functionality in the BioLuminate software. Because of limitations of the residue scanning application, variants breaking covalent bonds in the structural model, such as Cys-Cys disulfide bridges or coordinate bonds with metal cofactors like Mg or Zn ions, were assigned values of $\Delta_s(a_i) = \Delta_a(a_i) = \infty$. This value was chosen given the existing evidence that variants at these positions severely disrupt TNSALP functionality (Satou et al., 2012). Variants on the C- or N-terminus of the amino acid chain that were not part of the homology model were assigned a value of $\Delta_s(a_i) = \Delta_a(a_i) = 0$.

RESULTS

Protein structural model and missense variant scoring

There were 485 amino acid positions amenable to the residue scanning process, corresponding to the 524 positions on the canonical TNSALP transcript, minus the 17 positions corresponding to the signal peptide at the N-terminus and the 22 positions at the C-terminus missing from the model.

Sixteen positions where either side chains interacted with metal cofactors or disulfide bridges were present were also excluded from the residue scanning process (see Methods): disulfide bridges were present on positions $Cys^{139} - Cys^{201}$ and $Cys^{489} - Cys^{497}$ (based on UniProt amino acid sequence numbering). One Mg^{2+} ion had coordinate covalent bonds with the side chains of residues Glu^{235}, Glu^{291}, and Asp^{289}, as well as with the backbone of Phe^{290}. Another Mg^{2+} ion had coordinate bonds with the side chains of residues Asp^{60}, Thr^{173}, and Glu^{332}. One Zn^{2+} ion had coordinate bonds with the side chains of residues Asp^{337}, His^{341}, and His^{454}. Another Zn^{2+} ion had coordinate bonds with the side chains of residues Asp^{60}, Ser^{110}, Asp^{378} and His^{389}, as well as with the PO_4 ion.

REFERENCES

Beard, H., Cholleti, A., Pearlman, D., Sherman, W., & Loving, K. A. (2013). Applying physics-based scoring to calculate free energies of binding for single amino acid mutations in protein-protein complexes. *PloS One,* **8**(12), e82849. doi:10.1371/journal.pone.0082849

Jacobson, M. P., Friesner, R. A., Xiang, Z., & Honig, B. (2002). On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology,* **320**(3), 597-608.

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A hierarchical approach to all-atom protein loop prediction. *Proteins,* **55**(2), 351-367. doi:10.1002/prot.10613

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., … MacArthur, D. G. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv*, https://doi.org/10.1101/531210

Le Du, M. H., Stigbrand, T., Taussig, M. J., Menez, A., & Stura, E. A. (2001). Crystal structure of alkaline phosphatase from human placenta at 1.8 A resolution. Implication for a substrate specificity. *Journal of Biological Chemistry,* **276**(12), 9158-9165. doi:10.1074/jbc.M009250200

Negron, C., Pearlman, D. A., & Del Angel, G. (2019). Predicting mutations deleterious to function in beta-lactamase TEM1 using MM-GBSA. *PloS One,* **14**(3), e0214015. doi:10.1371/journal.pone.0214015

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design,* **27**(3), 221-234. doi:10.1007/s10822-013-9644-8

Satou, Y., Al-Shawafi, H. A., Sultana, S., Makita, S., Sohda, M., & Oda, K. (2012). Disulfide bonds are critical for tissue-nonspecific alkaline phosphatase function revealed by analysis of mutant proteins bearing a C(201)-Y or C(489)-S substitution associated with severe hypophosphatasia. *Biochimica et Biophysica Acta,* **1822**(4), 581-588. doi:10.1016/j.bbadis.2012.01.007