HHS Public Access

Author manuscript

J Inherit Metab Dis. Author manuscript; available in PMC 2024 February 16.

Published in final edited form as:

J Inherit Metab Dis. 2023 November ; 46(6): 1170–1185. doi:10.1002/jimd.12667.

Beyond genetics: Deciphering the impact of missense variants in CAD deficiency

Francisco del Caño-Ochoa[1], **Bobby G. Ng**[2], **Antonio Rubio-del-Campo**[1], **Sonal Mahajan**[2], **Matthew P. Wilson**[3], **Marçal Vilar**[4], **Daisy Rymen**[5], **Paula Sánchez-Pintos**[6,7], **Janna Kenny**[8], **Myriam Ley Martos**[9], **Teresa Campos**[10], **Saskia B. Wortmann**[11,12,*], **Hudson H. Freeze**[2,*], **Santiago Ramón-Maiques**[1,13,*]

[1]Structure of Macromolecular Targets Unit. Instituto de Biomedicina de Valencia (IBV), CSIC. Valencia, Spain

[2]Human Genetics Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA

[3]Laboratory for Molecular Diagnosis, Center for Human Genetics, KU Leuven, 3000 Leuven, Belgium

[4]Molecular Basis of Neurodegeneration Unit. Instituto de Biomedicina de Valencia (IBV), CSIC. Valencia, Spain

[5]Department of Pediatrics - Center for Metabolic Diseases, University Hospitals of Leuven, Belgium

[6]Unidad de Diagnóstico y Tratamiento de Enfermedades Metabólicas Congénitas. C.S.U.R. de Enfermedades Metabólicas. MetabERN. Hospital Clínico Universitario de Santiago de Compostela, La Coruña, Spain

Send correspondence to: Santiago Ramón-Maiques, Insituto de Biomedicina de Valencia (IBV), CSIC, Jaime Roig, 11. 46010-Valencia, Spain, sramon@ibv.csic.es; Hudson H. Freeze, Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA, hudson@SBPDiscovery.org; Saskia B. Wortmann, University Children's Hospital, Paracelsus Medical University (PMU), Strubergasse 21, 5020 Salzburg, Austria, s.wortmann@salk.at.

*Joint Senior Authors

Contributions of individual authors

- H.H. Freeze, S.B. Wortmann, B. Ng, P. Sánchez-Pintos, D. Rymen, M. Wilson, J. Kenny, M. Ley Martos and T. Campos provided genetic data and clinical analysis.
- F. del Caño-Ochoa and S. Mahajan performed site-directed mutagenesis and functional validation assays.
- M. Wilson performed and analyzed splicing experiments.
- F. del Caño-Ochoa, and A. Rubio-del Campo produced and characterized the DHO mutants, and F. del Caño-Ochoa and S. Ramón-Maiques determined the crystal structures.
- M. Vilar and S. Ramón-Maiques performed molecular dynamics simulations.
- S. Ramón-Maiques, S. Wortmann, and H.H Freeze wrote the initial manuscript, and all authors read it and contributed to its final writing.

Competing interest statement

Francisco del Caño-Ochoa, Bobby G. Ng, Antonio Rubio-del-Campo, Sonal Mahajan, Matthew P. Wilson, Marçal Vilar, Daisy Rymen, Paula Sánchez-Pintos, Janna Kenny, Myriam Ley Martos, Teresa Campos, Saskia B. Wortmann, Hudson H. Freeze and Santiago Ramón-Maiques declare that they have no conflict of interest related to this work. Saskia B. Wortmann receives funding from Paracelsus Medical University (PMU-FFF A-20/01/040-WOS) and ERAPERMED2019-310 – Personalized Mitochondrial Medicine (PerMiM): Optimizing diagnostics and treatment for patients with mitochondrial diseases (Austrian Science Funds (4704-B)). S.B. Wortmann declares that over the last 3 years in the area of inherited metabolic diseases she has received accommodation support from Nutricia Metabolics.

Ethics approval

Evaluation of the splice variant was performed in the context of validation of the genetic investigation for which written informed consent was obtained by the parents/legal guardian of the subject.

[7]Instituto de Investigación Sanitaria Santiago de Compostela (IDIS), La Coruña, Spain

[8]Children's Health Ireland at Crumlin, Ireland

[9]Pediatric Neurology Unit. Hospital Universitario Puerta del Mar, Cádiz, Spain

[10]Reference Center of Inherited Metabolic Diseases of Hospital de São João, Porto, Portugal

[11]University Children's Hospital, Paracelsus Medical University (PMU), Salzburg, Austria

[12]Amalia Children's Hospital, Radboudumc, Nijmegen, The Netherlands

[13]Group 739, Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER)–Instituto de Salud Carlos III, Valencia, Spain

Abstract

CAD is a large, 2,225 amino acid multi-enzymatic protein required for *de novo* pyrimidine biosynthesis. Pathological CAD variants cause a developmental and epileptic encephalopathy which is highly responsive to uridine supplements. CAD deficiency is difficult to diagnose because symptoms are nonspecific, there is no biomarker, and the protein has over 1,000 known variants. To improve diagnosis, we assessed the pathogenicity of 20 unreported missense CAD variants using a growth complementation assay that identified 11 pathogenic variants in 7 affected individuals; they would benefit from uridine treatment. We also tested 9 variants previously reported as pathogenic and confirmed the damaging effect of 7. However, we reclassified two variants as likely benign based on our assay, which is consistent with their long-term follow-up with uridine. We found that several computational methods are unreliable predictors of pathogenic CAD variants, so we extended the functional assay results by studying the impact of pathogenic variants at the protein level. We focused on CAD's dihydroorotase (DHO) domain because it accumulates the largest density of damaging missense changes. The atomic-resolution structures of 8 DHO pathogenic variants, combined with functional and molecular dynamics analyses, provided a comprehensive structural and functional understanding of the activity, stability, and oligomerization of CAD's DHO domain. Combining our functional and protein structural analysis can help refine clinical diagnostic workflow for CAD variants in the genomics era.

Keywords

Developmental and epileptic encephalopathy; epilepsy; uridine; treatment; functional validation assay; pyrimidine metabolism; variant of uncertain significance; inborn metabolic disease; dihydroorotase; protein structure-function; X-ray crystallography; molecular dynamics

INTRODUCTION

Genome-wide sequencing is indispensable for diagnosing inborn metabolic disorders and identifies millions of genetic variants.[1] However, the sequencing capacity is far greater than our ability to interpret the significance of genetic variation, and the classification of variants is a major task and a moving field.[2] Computational methods are commonly used to quantitate the potentially damaging effect of a variant on the gene product, but they are less reliable in pathogenicity evaluation. Studies have shown that up to one-third of benign

variations could be predicted as disease-causing and that even state-of-art *in silico* prediction tools, which may disagree with each other, should not be used as standalone evidence to infer variant pathogenicity.[3,4] Alternatively, functional validation assays offer a powerful and reliable way to assess the pathogenicity of genetic variants,[5] especially in poorly characterized proteins that present significant challenges for computational predictors. Regrettably, functional testing is not feasible for the majority of cases, not only due to the lack of functional assay itself but also because of the lack of specialized laboratories to perform them in a diagnostic context and often in the complete absence of funding from health care systems for these "post-exome" confirmatory diagnostics. Therefore, acquiring as much knowledge as possible about a protein's mechanisms and the effects of known deleterious mutations can help assess whether a variant of uncertain clinical significance (VUS) is connected to a health condition. In this context, the analysis of protein structures is an excellent tool for understanding disease-associated mutations, but its potential to predict deleterious effects needs to be fully exploited.

CAD deficiency (developmental and epileptic encephalopathy-50, OMIM #616457) is a recently discovered autosomal recessive neurometabolic disorder characterized by early-onset refractory epilepsy, developmental delay, severe developmental regression, and normocytic anemia.[6-8] The disease is caused by biallelic variants in *CAD*, encoding a 2,225 amino acid protein responsible for initiating the *de novo* synthesis of uridine 5-monophosphate (UMP), the precursor of all pyrimidine nucleotides for DNA and RNA synthesis, glycosylation, and other important cellular processes.[9] While the course of the disease left untreated is often fatal, patients respond well to the oral supplement of a uridine source (uridine, uridine monophosphate, uridine triacetate), which fuels pyrimidine nucleotide requirements through CAD-independent salvage pathways. In some treated patients, the improvement is striking, with a cessation of seizures and significant improvement in the clinical phenotype, and even regaining of lost skills.[6,8,10] However, the diagnosis is difficult since there are neither distinctive biomarkers nor a distinctive clinical phenotype, as the terms neurodevelopmental disorder and epilepsy are non-specific and commonly seen in a broad range of other diseases. Furthermore, genetic testing is challenged by ~1,000 *CAD* missense variants reported in the population.[11] Predicting whether a missense variant is pathogenic or benign is hampered by the complexity of the protein and how little we know about its functioning. The protein CAD results from the fusion of four enzymatic domains, glutamine amidotransferase (GLN), carbamoyl phosphate synthetase (CPS-2), aspartate transcarbamylase (ATC), and dihydroorotase (DHO), each catalyzing one of the initial sequential reactions for *de novo* UMP synthesis.[9] This already large multi-enzymatic protein further oligomerizes into hexameric particles,[12,13] subjected to a complex regulation through allosteric effectors and phosphorylation cascades.[14-19] So far, our understanding of the architecture of this mega-enzyme is limited to the crystal structures of the DHO and ATC domains,[20-23] and a paralogous human CPS (CPS-1),[24] and thus, our ability to predict the pathogenicity of the variants is greatly diminished.

To help assess the damaging potential of missense *CAD* variants, we developed a functional assay using CAD-knockout (KO) cells unable to grow in media without uridine.[25] The cells are transiently transfected with a cDNA encoding a green fluorescent protein (GFP)-tagged CAD carrying the individual-specific variant. Protein expression is confirmed

by fluorescence microscopy, and cell proliferation is monitored for one week in uridine-deprived conditions. CAD-KO cells expressing a pathogenic variant do not proliferate without uridine, whereas normal growth indicates that the variant is benign and compatible with CAD activity. Using this assay, we previously evaluated 34 missense variants found in 25 individuals suspected of CAD deficiency, out of which only 11 were proven to carry biallelic pathogenic variants and, therefore, suffer disease-causing CAD deficiency and could benefit from the uridine treatment.[26] The study also revealed the low reliability of sometimes conflicting computational approaches to predict the pathogenicity of CAD variants compared to the functional assay. Therefore, the CAD-KO rescue assay has remained the most reliable and feasible approach to discern pathogenic versus benign missense variants, and new suspects were brought to our attention. for evaluation.

In this study, we report the evaluation of 20 new missense variants found in 15 individuals with clinical phenotypes compatible with CAD deficiency. We also analyzed 9 variants reported as disease-causing in other studies to confirm their detrimental effect. To understand the molecular bases of disease-casuality by the identified pathogenic variants, we investigated the impact of those changes mapping n the DHO domain of CAD, which exhibits the greatest number of deleterious variants relative to its size. Our studies revealed the importance of the affected residues on the DHO active site, in stabilizing the protein, or as interactors with other domains in the unresolved architecture of CAD.

MATERIALS AND METHODS

Site-directed mutagenesis

Variants were introduced in plasmid pcDNA3.1-GFPhuCAD, encoding an N-terminal histidine-tagged enhanced green fluorescent protein (GFP) followed in-frame by human CAD,[25] using site-directed mutagenesis and specific primers (Supplementary Table S3) as reported,[25] or it was outsourced to GenScript (https://www.genscript.com).

Growth complementation assay

Growth complementation assays were performed as published.[25] In brief, U2OS cells knockout for *CAD* (*CAD*-KO) were grown in DMEM (Lonza), 10% fetal bovine serum (FBS; Sigma) dialyzed to remove uridine,[25,27] 2 mM L-glutamine (Lonza), and 50 $U \cdot ml^{-1}$ penicillin and 50 $\mu g \cdot ml^{-1}$ streptomycin (Invitrogen), at 5% CO_2 and 37 °C. Media was supplemented with 30 μM uridine (Sigma) to allow the growth of CAD-deficient cells. Cells were transfected with WT or mutated pcDNA3.1-GFPhuCAD using FuGene6 transfection reagent (Promega). One day before transfection, 1.5–2 x 10^5 cells in a final volume of 500 μl of medium were transferred to 24-well plates to reach approximately 50-80% confluence. For transfection, 2 μg of DNA in 50 μl of DMEM and 50 μl FuGene6 at $1 mg \cdot ml^{-1}$ in DMEM were incubated separately for 5 min at room temperature and then mixed and incubated at room temperature for an additional 10 min. The 100 μl mix was added to the wells drop by drop, followed by a 16 h incubation at 37°C and 5% CO_2. One day after transfection, 1 x 10^5 cells were seeded by duplicate in 24-well plates using media without uridine. Every 24 h, cells from one well were trypsinized and counted using Neubauer chamber or an automated cell counter (Luna II Automated cell counter, Logos Biosystems).

The efficiency of transfection (>95%) and the expression of the GFP-CAD protein variant were verified two days after transfection by fluorescence microscopy.

Protein production

The DHO coding region (aa 1456–1846) in the mutated *CAD* constructs was PCR amplified using specific primers (Supplementary Table S3) and inserted into pOPIN-M (Oxford Protein Production Facility) by IVA-cloning.[28] WT and mutated DHO proteins were expressed in HEK293 cells and purified as reported.[20,21,29]

Enzymatic assays

Enzymatic activity was assayed spectrophotometrically following the production of DHO by absorbance at 230 nm.[21] Reactions were carried out in a final volume of 100 μl containing 50 mM sodium phosphate pH 5.7, 150 mM NaCl, 20 μM $ZnSO_4$, 0.1 mg·ml^{-1} of bovine serum albumin, and 5 mM CA-asp. Protein concentrations were 0.25 μM for the WT and were increased to 2.5-3 μM for some mutants. The reaction was triggered by adding CA-asp to a mix with the enzyme pre-incubated for 5 min at the reaction temperature. Data analysis was performed with GraphPad Prism.

Differential scanning fluorimetry (DSF)

Protein stability was measured in triplicates by differential scanning fluorometry[30] using a 7500 Real-Time PCR system (Applied Biosystems) in a 96-well reaction plate with 20 μl sample, containing 5 μM protein, 5x SYPRO Orange (Invitrogen) and GF buffer with or without 2 mM 5-fluoroorotate (FOA). Fluorescence changes were monitored every 1°C in a temperature ramp from 20 to 95°C using the extrinsic fluorescence of SYPRO Orange ($\lambda_{excitation}$ = 465 nm, $\lambda_{emission}$ = 580 nm). Curves were normalized, and the melting temperature (T_m) was determined as the midpoint of the unfolding transition. Data were analyzed with GraphPad.

Splicing analysis

EBV-immortalized lymphoblastoid cells (LCLs) were cultured at 37°C under 5% CO_2 in DMEM-F12 with HEPES (Gibco, 11039-021) supplemented with 10% FBS (Clone III, HyClones). Puromycin treatment was performed by incubating cells for 2 h at 37°C with 100 μg·ml^{-1} puromycin before RNA extraction. Total RNA was isolated using an RNeasy Mini kit (Qiagen) alongside DNase treatment (Roche Diagnostics). A First-Strand cDNA synthesis kit (GE Healthcare) was used for reverse transcription from 2 μg RNA. PCR amplification and RT-qPCR cDNA were performed from the resulting cDNA with the oligonucleotides specified in Supplementary Table S4. RT-qPCR was performed using the Lightcycler 480 SYBR Green I Master Kit.

Structure determination

Crystals of the mutated DHO alone or in the presence of 4 mM dihydroorotate, carbamoyl aspartate or FOA were obtained as reported.[20,21,29] Optimal crystallization conditions consisted of 2–3 mg ml^{-1} protein and 2.0–3 M sodium formate, and 0.1 M HEPES pH 6.5-7.5 as the mother liquor. Crystals were cryoprotected with 20% glycerol and

flash-frozen. X-ray diffraction datasets were collected at XALOC (ALBA, Barcelona) and ID23-2 (ESRF, Grenoble). Data processing and scaling were performed automatically with autoPROC.[31,32] Crystallographic phases were obtained by molecular replacement with PHASER[33] and the WT DHO structure (PDB 4C6C and 4C6J) as search model. Protein models were traced with COOT[34] and refined with PHENIX[35] or Refmac5.[36,37]

Molecular dynamics

MD simulations were performed for the DHO dimer using GROMACS 2022.6 software package[38] and the CHARMM36m force field.[39]. Initial models with point mutations and files for GROMACS were prepared with CHARMM-GUI (http://www.charmm-gui.org).[40] The carboxylated lysine and the Zn cations were not included due to parametrization problems. The protein was solvated with water, and counterions (150 mM KCl) were added by replacing a corresponding number of water molecules to achieve a neutral condition. Van der Waals interactions were smoothly switched off at 10–12 Å by a force-switching function, and long-range electrostatic interactions were treated with the particle mesh Ewald method.[41] Simulations were performed at 310.15 K using a Nose-Hoover thermostat with τT = 1.0 ps. Pressure was maintained constant at 1 bar with a Parrinello-Rahman algorithm with a semi-isotropic coupling constant τP = 5.0 ps and compressibility = 4.5 Å~ 10^{-5} bar^{-1}. The LINCS method was used to constrain bond lengths. A time step of 2 fs was used for numerical integration. Coordinates were saved every 5 ps for analysis, and the simulation time was at least 400 ns. Analysis was performed using the GROMACS suite programs. The trajectories were visualized with PyMol (Schrödinger; https://pymol.org/2/).

RESULTS

Functional validation of *CAD* variants

Using a growth complementation assay,[25] we tested the functionality of 20 unreported *CAD* missense variants found in 15 individuals with a clinical phenotype compatible with CAD deficiency (Table 1). All patients referred for functional testing of CAD variants presented one or more of the features suggestive of CAD deficiency (seizures, epilepsy, developmental delay, intellectual disability, abnormalities on blood smear and/or anemia), and all variants were considered as VUS or likely pathogenic based on next-generation sequencing analysis. Upon transfection of CAD-KO cells, variants p.G58S, p.R238C, p.R1475Q, p.K1482M, p.S1538L, p.W1581R, p.R1617Q, p.H1687R, p.R1722W, and p.R1785H did not rescue the growth in media without uridine and were considered pathogenic as they impair CAD activity and *de novo* pyrimidine synthesis (Figure 1A). As previously seen, inactivating mutants in the GLN domain showed a partial rescue, indicating that free ammonia can, to some extent, feed the CPS-2 reaction.[25] Similar slow growth was observed for CAD variants p.K1482M and p.W1581R in the DHO domain, suggesting that these mutations were damaging, although to a lesser extent than other changes in the same domain. In turn, variants p.L56V, p.Q195K, p.R496W, p.I713T, p.A779S, p.V1783I, p.R2122H, p.P2137L, and p.A2201T rescued the growth phenotype and were regarded as benign (Figure 1A and Table 1). Interestingly, variant c.2030A>G introduced the change Q677R but also modified the last base of exon 13 and was predicted to affect splicing (Table 1). Since variant p.Q677R rescued the growth phenotype (Figure 1A), the alternative splicing was further

explored and confirmed (Supplementary Figure S1). Thus, the variant was considered pathogenic, although the change Q677R did not harm CAD activity. Overall, 11 of the 20 tested variants were diagnosed as pathogenic, and 7 out of 15 patients were confirmed as CAD-deficient subjects for carrying biallelic damaging variants (Table 1).

We also used the functional assay to validate 9 missense CAD variants reported as pathogenic by previous studies: p.M33R, p.Y60H, p.V762I, p.S948N, p.Y967C, p.V1259M, p.R1789Q, p.R1810Q, and p.E1954K.[6,8,42,43] We confirmed the damaging effect of all variants, except p.V762I and p.Y967C, that rescued the growth phenotype and were therefore reconsidered as likely benign (Figure 1A).

The results of the functional validation of the 63 CAD variants reported to date are summarized in Figure 1B and Supplementary Table S1. The comparison of these functional validation results with the predictions from state-of-the-art computational methods is striking (Figure 2 and Supplementary Figure S1). The scores calculated with CADD[44] for pathogenic and likely benign variants practically overlap, whereas values obtained from VARITY[45] for pathogenic variants are widely spread. Thus, unlike the functional assay, computational indicators cannot be used confidently to assess the disease-causing potential of CAD variants. Yet, these computational methods are expected to improve their predictions if properly trained with data resulting from a better understanding of the protein structure and function. Interestingly, 12 of the 34 (35%) pathogenic missense variants in CAD map within the DHO domain (Figure 1B), which represents only 1/6 of the entire protein and accumulates the largest number of deleterious changes. This, and the fact that we previously determined the crystal structure of this domain,[20,29,46] led us to produce the pathogenic DHO variants and study their functionality, molecular properties, and structure to start uncovering the specific impacts of pathogenic variants on CAD.

Production and characterization of DHO pathogenic variants

To start understanding the effect of the pathogenic variants, we produced and characterized the isolated DHO domain of human CAD bearing 8 damaging point mutations and compared them with the wild-type (WT) protein. Four of the 12 reported DHO pathogenic variants –p.S1538L, p.W1581R, p.H1687R, and p.R1785H– were recently identified, and although their characterization is underway, they could not be included in this study. The mutated DHO domains were produced in human cells using the same protocol as for the WT, and the yield and purity were similar (Figure 3A), indicating that the mutations did not severely affect the domain's folding and solubility. Size-exclusion chromatography also showed that the WT and mutated proteins eluted as single peaks at the expected positions for a homodimer (Figure 3B), and only R1810Q showed peak tailing, possibly suggesting partial dimer dissociation. These results indicated that, with the possible exception of R1810Q, the mutations did not impair DHO oligomerization.

Next, we compared the enzymatic activity of the WT and DHO mutants. The enzyme catalyzes a reversible pH-dependent reaction.[20,21,47,48] We measured the activity at an acidic pH (5.7) to favor the reaction in the forward direction (dihydroorotate synthesis). Although all the tested variants cause CAD deficiency (Figure 1), only R1475Q and K1556T were inactive, while K1482M exhibited ~5-fold decreased specific activity relative to the WT (v =

4.22 ± 0.2 U·mg^{-1}) (Figure 3C). All other tested mutants exhibited specific activities similar to WT.

We compared the stability of the different mutants relative to the WT by differential scanning fluorimetry (DSF) in the absence and presence of the product mimic 5-fluoroorotate (FOA) (Figure 3D).[20,21] The denaturing curves of WT without FOA were approximately sigmoidal with a midpoint temperature (Tm) of ~65 °C. The curve became biphasic in the presence of FOA, indicating a grossly ~10 °C increase in T_m. In contrast, the denaturing curves of the inactive mutants R1475Q and K1556T did not change with FOA, suggesting a lack of binding, whereas the partially active K1482M showed a small Tm increase, suggesting decreased substrate affinity (Figure 3D). The mutants with normal activities exhibited Tm changes with FOA addition (Figure 3D), suggesting unaffected substrate affinities and consistent with specific activities similar to WT (Figure 3C). However, R1617Q, R1789Q and R1810Q showed decreases of 3, 5 and 15°C in respective Tm values without FOA, pointing to decreased stability as a potential cause of their pathogenic nature. This was particularly marked for R1810Q, which showed ~4-fold lower enzymatic activity than the WT at 42°C and was nearly inactive at 50 °C (Figure 3E). This result confirmed that R1810Q decreases thermal stability of the enzyme.

Overall, these experiments indicated that mutations R1475Q, K1482M, and K1556T inactivate the DHO domain by impeding the binding of the substrate and that mutations R1789Q and R1810Q cause a marked decrease in stability. The effect of variants R1617Q, R1722W, and R1785C remained unclear as the protein stability or activity changes were subtle.

Structures of the inactivating variants K1482M, R1475Q and K1556T

We further studied the impact of the variants by solving the crystal structure of the DHO domain bearing the point mutations. Previously, we reported the structure of the isolated human DHO domain, showing that it is a homodimer in solution, with each subunit folded in a $(\alpha/\beta)_8$-barrel, with a Zn-containing active site enclosed by a flexible loop, an adjacent domain of unknown function, and a C-terminal extension placing the N- and C-termini on opposite sides of the protein (Figure 4A).[20,21] Using similar conditions as for the WT, we crystallized and determined the structures of the DHO mutants R1475Q, K1482M, K1556T, R1617Q, R1722W, R1785C, R1789Q, and R1810Q, free or in complex with FOA or with the substrate (carbamoyl aspartate or dihydroorotate). The statistics for the diffraction and refinement data of the 14 new structures are summarized in Supplementary Table S2a,b.

The inactivating mutations K1482M, R1475Q, and K1556T affect elements at or near the active site. In the WT, the side chain of K1556 undergoes a posttranslational carboxylation required to coordinate two catalytic Zn^{2+} cations and form the active site (Figure 4B). Thus, any other residue at this position would hamper the correct binding of the metals and the formation of the active site. However, to our surprise, the structures of K1556T showed an intact active site, with the substrate or FOA bound, the Zn^{2+} cations in the correct position, the Thr replacing K1556 well-defined in the electron density map, and a molecule of formate from the crystallization condition mimicking the posttranslational modification of the missing lysine (Figure 4C). Thus, the presence of 2–3 M sodium formate, 1–2 mM

$ZnSO_4$, and 2 mM FOA/substrate in the crystallization solution favored the assembly of the active site in an inactive mutant that failed to bind the substrate in more physiological conditions (Figure 3C,D).

On the other hand, the WT structure showed that residue R1475, N1505, and H1690 are responsible for recognizing the α-COOH group of the substrate (Figure 4B). Thus, mutation R1475Q was expected to hamper the binding and correct positioning of the substrate, as confirmed by the stability assay (Figure 3D). Indeed, despite increasing the concentration of FOA or dihydroorotate to 10 mM in the crystallization condition, the structure showed a fuzzy electron density at the active site, suggesting weak or no binding of the ligands, the side chain of the mutated Gln pointing outside the active site, the loop containing the mutated residue (loop 1) appeared flexibly disordered, and an extra Zn^{2+} cation (from the crystallization condition) occupied the position of K1482 in the interaction with H1690 and E1694 in loop 8 (Figure 4D). This zigzagging interaction between loop 1 (residues H1481 and K1482), loop 8 (E1694 and H1690), and the substrate α-COOH group (Figure 4B), was also disrupted by mutation K1482M. The structure of this mutant showed the mutated Met facing outside the active site, the side chain of H1481 occupying the position of the missing Lys, and an extra Zn^{2+} cation coordinated by H1481, E1694, H1690 and a formate molecule (Figure 4E). These results strongly suggested that the activity loss of R1475Q and K1482M is caused by the increased flexibility of loops 1 and 8 and the incorrect positioning of H1690 and R1475Q to interact with the substrate. However, this effect is possibly minimized in the crystal structures because the high concentration of formate, Zn^{2+}, and dihydroorotate/FOA stabilize the active site elements.

Structures of variants R1617Q, R1722W and R1785C

Next, we studied the impact of variants R1617Q, R1722W, and R1785C on the protein structure. Residue R1617 locates three residues downwards the Zn-coordinating H1614, and interacts with E1592, which also binds the Zn-coordinating H1590 (Figure 4B). Thus, mutation R1617Q was predicted to affect the coordination of Zn and decrease the catalytic efficiency. However, the activity of R1617Q was similar to WT (Figure 3C), and the structure, obtained in complex with the substrate or FOA, showed no alterations in the active site (Figure 4F). The side chain of the replacing Gln is well-defined in the electron density map, and two water molecules occupy the positions of the missing Arg and interact with E1592. To better understand the effect of R1617Q, we performed molecular dynamics (MD) simulations. The 400 ns trajectories calculated for the WT showed that the side chain of residue R1617 does not occupy the fixed position observed in the crystal structure but rather shifts its interactions between E1592, and the nearby negatively charged residues, E1619, E1620, E1663, and D1669 (Supplementary Figure S2A). The formation of any of these strong salt bridges between different structural elements in the (α/β)-barrel would be impeded by mutation R1617Q, which may explain the decreased stability of the mutant (Figure 3D).

Residue R1722, on the other hand, is in an invariant loop (^{1721}GR^{1722}) exposed to the solvent between the last helix of the $(\alpha/\beta)_8$-barrel and the helix connecting with the adjacent domain (Figure 4G, left panel). The crystal structure of R1722W showed that replacing

the positively charged Arg with a large and hydrophobic residue did not alter the local arrangement of the loop (Figure 4G, right panel). Rather than a stacking interaction between the guanidine group of R1722 and the side chain of W1674, the mutant showed a π-cation interaction between the two Trp side chains, and an additional density adjacent to the mutated Trp was interpreted as a glycerol molecule from the cryogenic solution. However, analysis of MD trajectories showed that crystal packing favors a conformation that masked the effect of the mutation. Indeed, early in the simulated trajectory, the side chain of residue E1671 moves from the position favoring crystal lattice contacts to interact with R1722 forming a stable salt bridge (Supplementary Figure S2B-D). Mutation R1722W would hamper this strong linkage between adjacent structural elements, explaining the decrease, although small, in protein stability (Figure 3D).

Residue R1785 locates in the middle β-strand of a five-stranded β-sheet formed by the adjacent domain and the beginning of the C-terminal extension (Figure 5A). The side chain of R1785 makes electrostatic interactions with adjacent structural elements, and thus mutation R1785C was expected to disrupt these interactions and destabilize the protein. Instead, R1785C showed a small increase in thermal stability (Figure 3D), and the structure of the mutant showed that despite the lack of interactions, the introduced Cys did not alter the adjacent domain (Figure 5B). Since the mutation does not seemingly damage the stability nor the activity of the protein, one possibility is that it affects the interaction between the DHO and the CPS-2 domain of CAD. A short 3 aa loop bridges the CPS-2 with the adjacent domain, and thus R1785 and other conserved elements at this interface could be involved in interdomain contacts and assembly of the full-length protein (Supplementary Figure S3).

Structure of the destabilizing variants R1789Q and R1810Q

The C-terminal extension is an important structural element, absent in the DHOs of prokaryotes and plants, which places the N- and C-termini at opposite ends of the domain (Figure 4A). In the WT structure, the side chain of R1789 projects from the adjacent domain and makes H-bonds with the main and side chains of Y1805 in the C-terminal extension and a salt bridge with D1530 in the central β-barrel (Figure 5A). The structure of R1789Q showed that the replacing Gln forms an H-bond with D1530 rather than a strong salt bridge and lacks interactions with the C-terminal extension (Figure 5C).

On the other hand, mutation R1810Q affects a pivotal residue anchoring the C-terminal extension to the bottom of the $(\alpha/\beta)_8$-barrel (Figure 5D, left panel). The side chain of R1810 forms H-bonds with carbonyl oxygens of residues A1552, A1553, F1740, and H1741 in the $(\alpha/\beta)_8$-barrel, and with G1815 downstream the C-terminal extension, which follows a wavy trace. The crystal structure of mutant R1810Q showed that the replacing Gln only interacts with A1552 and with a water molecule that mimics the interactions of the missing Arg with A1553 and F1740 (Figure 5D, right panel), and there was no electron density beyond residue 1811, indicating that the C-terminal extension was flexibly disordered.

Overall, these results indicate that mutations R1789Q and R1810Q damage the anchoring of the C-terminal extension to the DHO domain, explaining the decreased stability of these mutants (Figure 3D).

DISCUSSION

We identified 13 previously unreported damaging changes in *CAD*, including 10 missense, 2 splicing, and 1 loss of function variants. This confirmed the diagnosis of CAD deficiency in 7 patients with suggestive clinical features and biallelic variants in CAD (Table 1). Less than 50 CAD-deficient subjects have been reported worldwide, and as new cases are brought regularly to our attention, we are confident that it is still grossly underdiagnosed. Being a severe and often fatal neurometabolic disease with a safe treatment option available, we aim to develop rapid and reliable tools for its diagnosis. Additionally, biomarkers and the use of fluxomic assays to diagnose the disease are currently being investigated but rely on patient-derived material (e.g., fibroblasts, urine). Hence, the proliferation assay remains the most informative approach, and we used it to complete the functional validation of the 63 missense clinical *CAD* variants reported thus far (Figure 1B and Supplementary Table S1).

As new variants continue appearing, we are concerned about the limited availability of this functional assay, which is currently only done in two research groups (S. Ramón-Maiques and H. H. Freeze) but needs to be implemented and available in diagnostic laboratories. The assay has no complications other than handling cell cultures, the CAD-KO cells and GFP-CAD cDNA used are available upon request, and troublesome mutagenesis can be outsourced (see Methods). But, in any case, the purpose of the functional validation is to complete the previous and necessary genetic analysis since a direct assessment at the protein level could be misleading. This is exemplified here by the variant c.2030A>G, which introduces a benign change at the protein level, p.Q677R, but modifies the last position of the exon, favoring aberrant splicing (Table 1 and Supplementary Figure S1). Once again, we emphasize that erroneously variant interpretation in cases of CAD deficiency would mean that an available and effective treatment is not given to severely diseased patients. Conversely, patients with CAD variants lacking functional impact receive unnecessary daily uridine supplements, which may lead to false hopes of improvement.

Another aspect that should be considered is the possibility that exogenously overexpressed hypomorphic variants could rescue cell growth and go undetected. Until we manage to tune the transient expression to the endogenous protein levels, pathogenic variants can be convincingly annotated, whereas those rescuing growth should be labeled as "likely benign" and evaluated alongside clinical data, including the possible response to uridine treatment.[49] In this line, two variants reported as pathogenic, V762I and V967C, rescued cell proliferation and were thus reclassified as likely benign (Figure 1A,B). Homozygous V967C variant occurred in two siblings with mild developmental issues, hematological issues, and no epilepsy.[8] Follow-up on these cases (now aged 7 and 4.5 years) revealed that both attended regular school/kindergarten and did not develop epilepsy. Despite uridine treatment, their mild anemia and anisopoikilocytosis, a frequent trait in CAD-deficient patients, persisted. Thus, there are arguments that the patients might not be CAD-deficient and that the variant could be benign. The other dubious variant, V762I, was found along with the confirmed pathogenic variant S948N in a patient with a story of refractive seizures who showed a better development with uridine treatment.[8] Follow-up on this patient revealed that he is on the same anti-epileptic drug regime and no longer has absence seizures under uridine supplementation. Also, mild anemia normalized. After an initial developmental

spurt, this went back to the same pace as before the uridine supplementation. Hence there are arguments for and against the pathogenicity of the variant leading to CAD deficiency.

There is also the possibility that the combination of two "benign" variants could be damaging and cause CAD deficiency. This could be well the case for changes at the protein interfaces that are insufficient to have an effect by themselves but that, in combination, could alter the assembly of CAD into an active oligomeric particle.

Treatment with a uridine source (uridine, uridine monophosphate or uridine triacetate) could be recommended when in doubt since no secondary effects have been reported. However, other inputs to the functional assay would be desirable to help discern the possible mild yet damaging effect of hypomorphic variants. There are high expectations on the rapidly evolving computational algorithms to aid in evaluating VUS, or even to reliably predict their pathogenicity without establishing a battery of disease-specific functional assays. However, despite advances in the field, computational approaches could be limited by our knowledge of the protein mechanisms and the "molecular sociology" that dictates interactions with other cellular components.[50] The potential to enhance the accuracy of computational predictions in CAD is evident (Figure 2). We lack the expertise to delve into the reasons behind inaccurate predictions, but at the moment, we strongly advise against using these computational methods to predict variant pathogenicity in CAD. Computational methods could benefit from the comprehensive structural and functional characterization of the four CAD domains, their assembly into a multienzyme particle, the detailed comprehension of the catalytic and regulatory properties, and deciphering the molecular mechanisms of already known pathogenic variants. To this aim, in this study, we started investigating the impact of pathogenic variants at the protein level.

Among the 997 missense variants found in *CAD*,[11] 220 are located in the DHO domain, the second most variable domain after CPS (390 variants), but before GLN (206 variants), ATC (136 variants) and the DHO-ATC linker (45 variants). The DHO domain exhibits the highest number of changes relative to its size, with 47 % of its residues having at least one missense variant. This variability might suggest that DHO is less important than other domains, with more missense variants being fixed in the population. However, the current study shows a greater accumulation of pathogenic variants within the DHO domain (Figure 1B), suggesting a prominent role in CAD's functioning mechanisms. In addition to its enzymatic function, DHO plays a crucial role in CAD's oligomerization, occupying a core position in the hexameric particles.[22] This scaffolding function is particularly evident in fungi. These organisms have a CAD-like protein with an inactive DHO-like domain, and an additional gene encodes for an independent, active DHO enzyme.[9] The structure of the isolated fungal DHO-like domain revealed strong similarity to the human CAD's DHO domain, except for the absence of the active site.[22] These results suggest that apart from its enzymatic activity, DHO serves a conserved structural function from fungi to humans. Much remains to be understood about CAD's DHO domain and its participation in assembling the mega-enzymatic particles. The abundance of pathogenic variants in this domain, including the 8 pathogenic changes characterized in this study, offers invaluable insights into identifying crucial elements of the protein's structure and function. This wealth of information should also facilitate the interpretation of new clinical variants. For instance, mutation K1556T

confirmed the central role of this residue with a post-translational modification in organizing the Zn-containing active site, whereas R1475Q and K1482M highlighted the importance of preserving the interactions between loops 1 and 8 to bind the substrate. On the other hand, R1617Q and R1722Q helped us identify electrostatic interactions between secondary structural elements that passed unadvertised in the initial structural characterization, but that proved important for stability. Similarly, R1789Q and R1810Q revealed the importance of correctly anchoring the C-terminal extension to preserve the integrity of the domain. Additionally, R1785C drew attention to the adjacent domain that had previously been assigned no function other than covering one side of the $(\alpha/\beta)_8$-barrel, and which is now proposed to provide the interactions with the CPS-2 domain.

Integrating structural knowledge with clinical and functional genomics can accelerate the diagnosis of new CAD-deficient patients and enable their timely treatment. However, uncovering the structure, mechanisms, and abnormalities of this mega-enzyme within the context of human disease poses a formidable challenge. This is an example of the ongoing challenges on reaching meaningful diagnoses in the genomics era and the unmet need of integrating functional validation assays or other “post-exome” investigations into the clinical diagnostic workflow that needs attention and financial coverage by the healthcare systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the clinicians, especially Koen Devriendt, Kristin Lindstrom, Matthew Troester, Kimihiko Oishi, Kaan Boztug, Marianne Rohrbach, Birgit Kaufmann, and Gregory Constain, who provided patient information, Jordi Pijuan for help with genetic analysis, and Manuel Soriano for computer support. This work was supported by grant RTI2018-098084-B-I00 financed by MCIN/AEI/10.13039/501100011033/ and "FEDER Una manera de hacer Europa", by grant PID2021-128468NB-I00 financed by MCIN/AEI /10.13039/501100011033 and by a grant from Fundación Ramón Areces Ciencias de la Vida (XX National Call) to SR-M. FdC-O is a postdoctoral fellow of the Generalitat Valenciana (APOSTD 2021). AR-d-C is supported by salary from the European Commission–Next Generation EU through the CSIC Global Health Platform (PTI+ Salud Global) established by EU Council Regulation 2020/2094. HHF, BN, SMP were supported by The Rocket Fund, R01DK099551, and U54 NS115198. SMP is also supported by a training component of U54 NS115198. MPW is supported by an MSCA individual fellowship (894669) and a FWO Senior Postdoctoral Fellowship (1289023N). X-ray diffraction experiments at synchrotrons were done through the participation of SR-M in the BAG proposals 2017082302, 2018082950, 2019093709, 2020074406, 2021075216, and 2022075911 at ALBA, and MX-2076, MX-2351 and MX-2452 at ESRF. We thank the ALBA synchrotron staff and Max H. Nanao at beamtime ID23-2 at ESRF synchrotron for assistance.

Data sharing statement

CAD-knockout cells and cDNA encoding human GFP-CAD for functional assays and molecular dynamics trajectories are available from S. Ramón-Maiques upon request. Crystal structures and structural factors have been deposited in the Protein Data Bank with the accession numbers indicated in Supplementary Table S2a,b.

REFERENCES

1. Biesecker LG, Green RC. Diagnostic Clinical Genome and Exome Sequencing. N Engl J Med. 2014;370:2418–2425. [PubMed: 24941179]

2. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–424. [PubMed: 25741868]
3. Ernst C, Hahnen E, Engel C, et al. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. BMC Med Genomics. 2018;11:35. [PubMed: 29580235]
4. Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? PLoS Comput Biol. 2019;15:e1006481. [PubMed: 30742610]
5. Rodenburg RJ. The functional genomics laboratory: functional validation of genetic variants. J Inherit Metab Dis. 2018;41:297–307. [PubMed: 29445992]
6. Koch J, Mayr JA, Alhaddad B, et al. CAD mutations and uridine-responsive epileptic encephalopathy. Brain. 2017;140:279–286. [PubMed: 28007989]
7. Ng BG, Wolfe LA, Ichikawa M, et al. Biallelic mutations in CAD, impair de novo pyrimidine biosynthesis and decrease glycosylation precursors. Human molecular genetics. 2015;24:3050–3057. [PubMed: 25678555]
8. Rymen D, Lindhout M, Spanou M, et al. Expanding the clinical and genetic spectrum of CAD deficiency: an epileptic encephalopathy treatable with uridine supplementation. Genetics in Medicine. 2020;22:1589–1597. [PubMed: 32820246]
9. Del Caño-Ochoa F, Ramón-Maiques S. Deciphering CAD: Structure and function of a mega-enzymatic pyrimidine factory in health and disease. Protein Sci. 2021;30:1995–2008. [PubMed: 34288185]
10. Al-Otaibi A, AlAyed A, Al Madhi A, et al. Uridine monophosphate (UMP)-responsive developmental and epileptic encephalopathy: A case report of two siblings and a review of literature. Mol Genet Metab Rep. 2022;30:100835. [PubMed: 35242569]
11. ENST00000264705 | gnomAD v2.1.1 | gnomAD. Available at: https://gnomad.broadinstitute.org/transcript/ENST00000264705?dataset=gnomad_r2_1. Accessed June 2, 2023.
12. Coleman PF, Suttle DP, Stark GR. Purification from hamster cells of the multifunctional protein that initiates de novo synthesis of pyrimidine nucleotides. The Journal of biological chemistry. 1977;252:6379–6385. [PubMed: 19472]
13. Lee L, Kelly RE, Pastra-Landis SC, Evans DR. Oligomeric structure of the multifunctional protein CAD that initiates pyrimidine biosynthesis in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America. 1985;82:6802–6806. [PubMed: 2995985]
14. Ben-Sahra I, Howell JJ, Asara JM, Manning BD. Stimulation of de novo pyrimidine synthesis by growth signaling through mTOR and S6K1. Science. 2013;339:1323–1328. [PubMed: 23429703]
15. Carrey EA, Campbell DG, Hardie DG. Phosphorylation and activation of hamster carbamyl phosphate synthetase II by cAMP-dependent protein kinase. A novel mechanism for regulation of pyrimidine nucleotide biosynthesis. The EMBO journal. 1985;4:3735. [PubMed: 4092695]
16. Graves LM, Guy HI, Kozlowski P, et al. Regulation of carbamoyl phosphate synthetase by MAP kinase. Nature. 2000;403:328–332. [PubMed: 10659854]
17. Jones ME. Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. Annual review of biochemistry. 1980;49:253–279.
18. Robitaille AM, Christen S, Shimobayashi M, et al. Quantitative phosphoproteomics reveal mTORC1 activates de novo pyrimidine synthesis. Science. 2013;339:1320–1323. [PubMed: 23429704]
19. Shin J, Mir H, Khurram MA, et al. Allosteric regulation of CAD modulates de novo pyrimidine synthesis during the cell cycle. Nat Metab. 2023;5:277–293. [PubMed: 36747088]
20. Del Cano-Ochoa F, Grande-Garcia A, Reverte-Lopez M, D'Abramo M, Ramon-Maiques S. Characterization of the catalytic flexible loop in the dihydroorotase domain of the human multi-enzymatic protein CAD. J Biol Chem. 2018;293:18903–18913. [PubMed: 30315107]
21. Grande-Garcia A, Lallous N, Diaz-Tejada C, Ramon-Maiques S. Structure, functional characterization, and evolution of the dihydroorotase domain of human CAD. Structure. 2014;22:185–198. [PubMed: 24332717]

22. Moreno-Morcillo M, Grande-Garcia A, Ruiz-Ramos A, Del Cano-Ochoa F, Boskovic J, Ramon-Maiques S. Structural Insight into the Core of CAD, the Multifunctional Protein Leading De Novo Pyrimidine Biosynthesis. Structure. 2017;25:912–923 e5. [PubMed: 28552578]
23. Ruiz-Ramos A, Velazquez-Campoy A, Grande-Garcia A, Moreno-Morcillo M, Ramon-Maiques S. Structure and Functional Characterization of Human Aspartate Transcarbamoylase, the Target of the Anti-tumoral Drug PALA. Structure. 2016;24:1081–1094. [PubMed: 27265852]
24. de Cima S, Polo LM, Diez-Fernandez C, et al. Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. Sci Rep. 2015;5:16950. [PubMed: 26592762]
25. del Caño-Ochoa F, Ng BG, Abedalthagafi M, et al. Cell-based analysis of CAD variants identifies individuals likely to benefit from uridine therapy. Genetics in Medicine. May 2020:1–8.
26. Del Caño-Ochoa F, Ng BG, Abedalthagafi M, et al. Cell-based analysis of CAD variants identifies individuals likely to benefit from uridine therapy. Genet Med. 2020;22:1598–1605. [PubMed: 32461667]
27. Patterson D, Carnright DV. Biochemical genetic analysis of pyrimidine biosynthesis in mammalian cells: I. Isolation of a mutant defective in the early steps of de novo pyrimidine synthesis. Somatic Cell Genet. 1977;3:483–495. [PubMed: 21463]
28. García-Nafría J, Watson JF, Greger IH. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. Scientific Reports. 2016;6:27459. [PubMed: 27264908]
29. Lallous N, Grande-Garcia A, Molina R, Ramon-Maiques S. Expression, purification, crystallization and preliminary X-ray diffraction analysis of the dihydroorotase domain of human CAD. Acta crystallographica Section F, Structural biology and crystallization communications. 2012;68:1341–1345. [PubMed: 23143245]
30. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2:2212–2221. [PubMed: 17853878]
31. Vonrhein C, Flensburg C, Keller P, et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol Crystallogr. 2011;67:293–302. [PubMed: 21460447]
32. Monaco S, Gordon E, Bowler MW, et al. Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. J Appl Crystallogr. 2013;46:804–810. [PubMed: 23682196]
33. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of applied crystallography. 2007;40:658–674. [PubMed: 19461840]
34. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography. 2010;66:486–501. [PubMed: 20383002]
35. Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography. 2010;66:213–221. [PubMed: 20124702]
36. Murshudov GN, Skubak P, Lebedev AA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67:355–367. [PubMed: 21460454]
37. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011;67:235–242. [PubMed: 21460441]
38. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19–25.
39. Huang J, Rauscher S, Nawrocki G, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14:71–73. [PubMed: 27819658]
40. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. Journal of Computational Chemistry. 2008;29:1859–1865. [PubMed: 18351591]
41. Darden T, Tork D, Pedersen L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. J Comput Chem. 1997;18:1463–1472.
42. Russo R, Marra R, Andolfo I, et al. Uridine treatment normalizes the congenital dyserythropoietic anemia type II-like hematological phenotype in a patient with homozygous mutation in the CAD gene. American Journal of Hematology. 2020;95:1423–1426. [PubMed: 32720728]
43. Zhou L, Xu H, Wang T, Wu Y. A patient with CAD deficiency responsive to uridine and lietrature review. Front Neurol. 2020;11:5. [PubMed: 32116996]

44. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47:D886–D894. [PubMed: 30371827]
45. Wu Y, Liu H, Li R, Sun S, Weile J, Roth FP. Improved pathogenicity prediction for rare human missense variants. The American Journal of Human Genetics. 2021;108:1891–1906. [PubMed: 34551312]
46. Ruiz-Ramos A, Grande-García A, Ramón-Maiques S. Dihydroorotase domain of human CAD. In: John Wiley & Sons L, ed. Encyclopedia of Inorganic and Bioinorganic Chemistry, Online. John Wiley & Sons, Ltd.; 2015.
47. Christopherson RI, Jones ME. Interconversion of carbamayl-L-aspartate and L-dihydroorotate by dihydroorotase from mouse Ehrlich ascites carcinoma. The Journal of biological chemistry. 1979;254:12506–12512. [PubMed: 40997]
48. Porter TN, Li Y, Raushel FM. Mechanism of the dihydroorotase reaction. Biochemistry. 2004;43:16285–16292. [PubMed: 15610022]
49. Shen JJ, Wortmann SB, de Boer L, et al. The role of clinical response to treatment in determining pathogenicity of genomic variants. Genetics in Medicine. 2021;23:581–585. [PubMed: 33087887]
50. Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature. 2007;450:973–982. [PubMed: 18075576]
51. Yarahmadi SG, Morovvati S. CAD gene and early infantile epileptic encephalopathy-50; three Iranian deceased patients and a novel mutation: case report. BMC Pediatrics. 2022;22:125. [PubMed: 35277149]

Synopsis

Integrating functional genomics with structural and functional protein knowledge, we identified pathogenic variants in the multi-enzymatic protein CAD and explained the underlying disease-causing mechanisms.

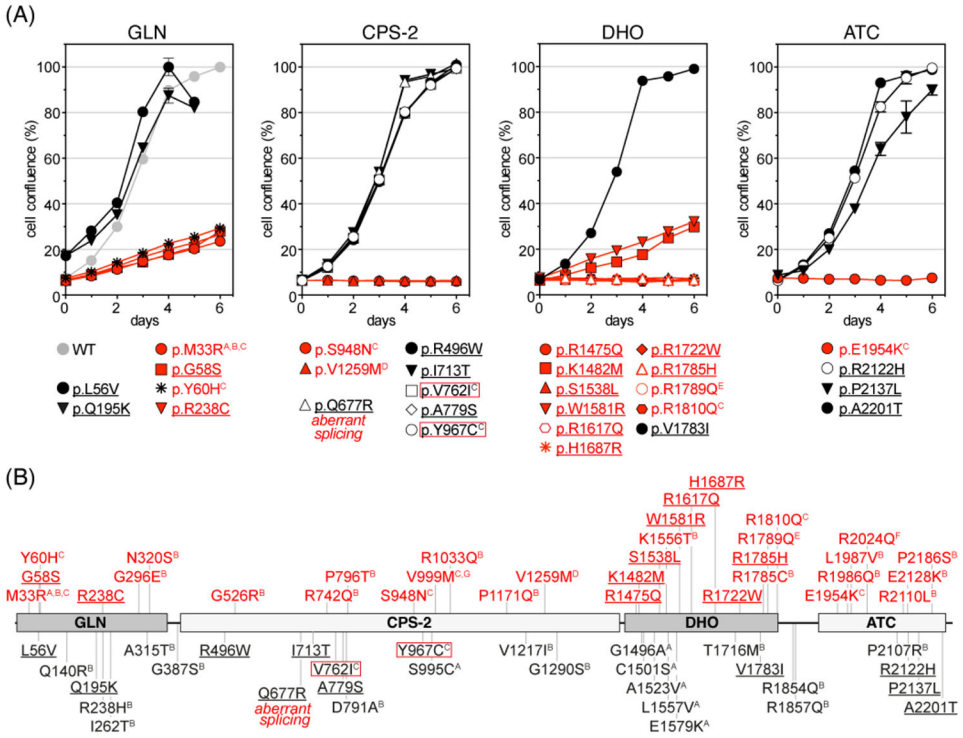

Figure 1. Functional validation of CAD variants.

A) Growth complementation assay of *CAD*-knockout (KO) cells grown without uridine and transfected with GFP-CAD WT (grey) or bearing point variants in the GLN, CPS-2, DHO, or ATC domains. Cell proliferation is shown as % confluence to KO-cells transfected with GFP-CAD WT. Each point represents the mean and standard deviation of two independent experiments, counted in duplicates, except for L56V, Q195K, and P2137L, which were assayed once and counted in triplicates. **B)** Scheme of CAD mapping all missense changes functionally validated. Pathogenic and likely benign changes are shown in red and black, respectively. Those reported for the first time in this study are underscored, whereas others reported in different studies are indicated with superscripts: A,[6] B,[25] C,[8] D,[43] E,[42] F,[7], and G.[51] Variants p.V762I and p.Y967C, shown with a red frame, were reported as pathogenic[8] but reconsidered as likely benign based on the cell assay. Variant p.Q677R rescues cell growth and is shown in black, although the nucleotide change is pathogenic for inducing aberrant splicing.

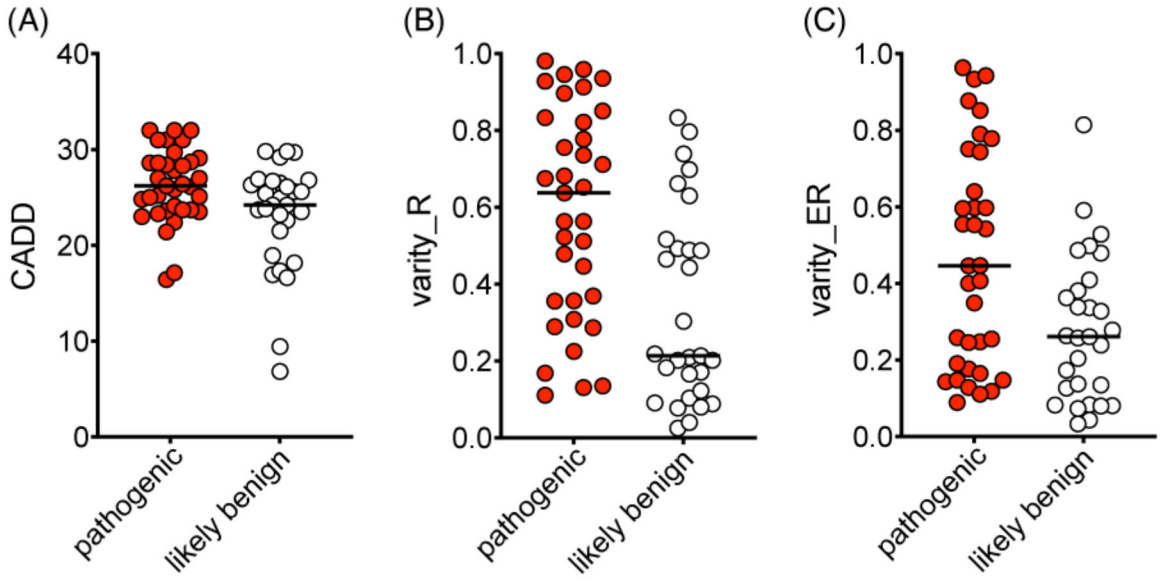

Figure 2. Performance of computational predictors contrasted with results of the KO-rescue assay.

Scatter plots of values calculated with computational predictors CADD[44] **(A)** and VARITY[45] **(B,C)** for variants validated as pathogenic (red dots) or likely benign (white dots) using the KO-rescue assay. The horizontal bar represents the median of the values.

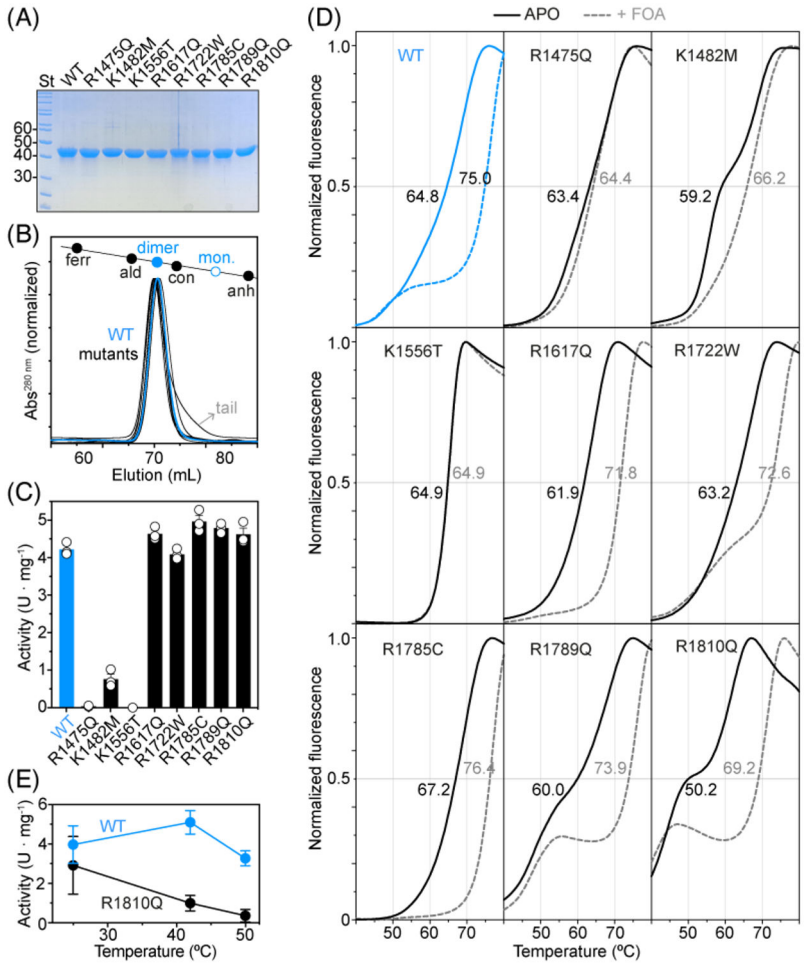

Figure 3. Characterization of pathogenic variants in CAD's DHO domain.

A) SDS-PAGE of purified recombinant DHO domains bearing point mutations. St, molecular weight standards (kDa). **B)** Size exclusion chromatography of purified DHO proteins eluting as a single peak at the position expected for a homodimer. The tailed peak of R1810Q is indicated. On top, column calibration with proteins of known molecular weight (ferritin, 440 kDa; aldolase, 158 kDa; conalbumin, 75 kDa; carbonic anhydrase 29 kDa), indicating in blue the positions for the DHO dimer (closed circle) and the expected position for the monomer (open circle). **C)** Enzymatic activity of WT and mutants measured at 25 °C. The scattered plot shows three independent measurements. Activity units (U) are nmol of dihydroorotate per min. **D)** Denaturing curves in the absence (APO) and presence of fluoroorotate (FOA) measured by differential scanning fluorimetry. The midpoint temperatures (T_m) for the curves are indicated. **E)** Activity assay at different temperatures for the WT and R1810Q.

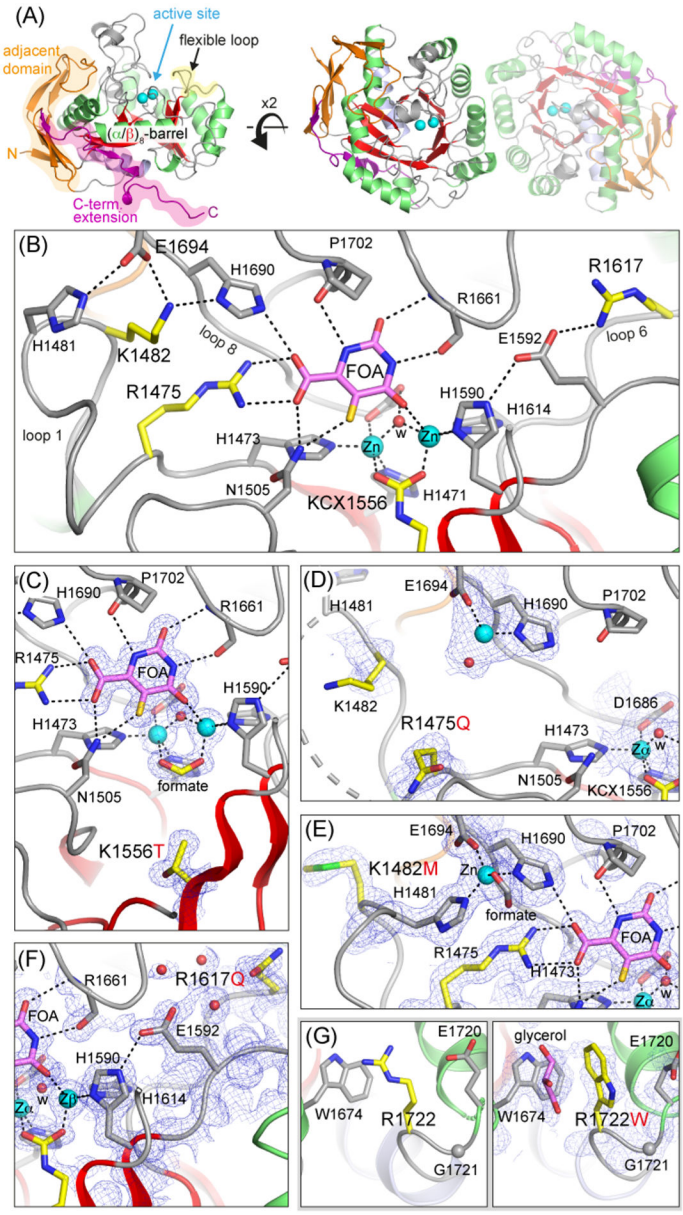

Figure 4. Characterization of DHO variants by X-ray crystallography.

A) Cartoon representation of the crystal structure of CAD's DHO subunit and dimer (PDB ID 4C6C). **B)** Detail of the DHO's active site with Zn^{2+} shown as cyan spheres and a molecule of FOA depicted with carbon atoms in pink. The carboxylated lysine (KCX) and the other residues bearing pathogenic substitutions are depicted with carbon atoms in yellow. Dashed lines represent relevant electrostatic interactions. **C–G)** Details of the crystal structures for mutants K1556T (C), R1475Q (D), K1482M (E), R1617Q (F), and R1722W (G). The $2F_{obs}$-F_{calc} electron density map for regions of interest is shown as a blue mesh contoured at 1.0 σ. Flexible disordered loops are represented as thick dashed lines. Panel G) shows the structure of the WT and R1722W side by side for comparison.

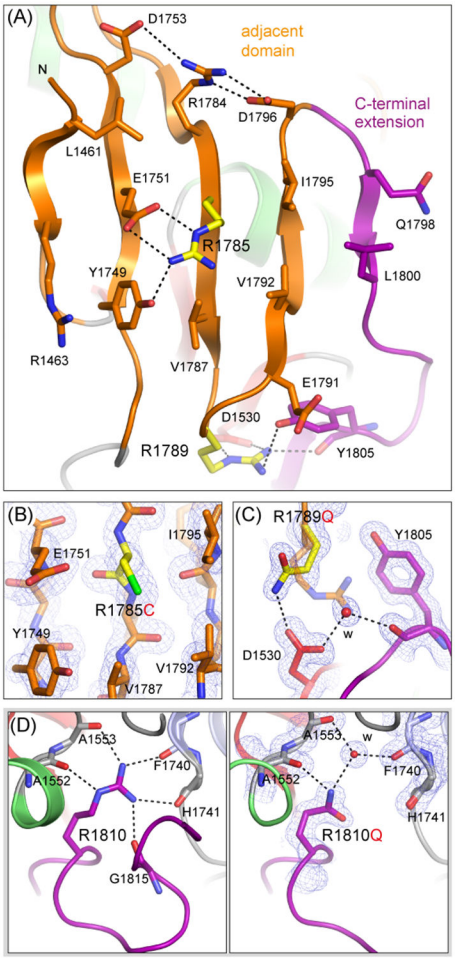

Figure 5. Structure of DHO variants in the adjacent domain and C-terminal extension.

A) Cartoon representation of the WT DHO adjacent domain and beginning of the C-terminal extension. **B–D)** Detailed views of the crystal structures of mutants R1785C (B), R1789Q (C), and R1810Q (D). $2F_{obs}$-F_{calc} electron density maps are shown as a blue mesh contoured at 1.0 σ. Panel D) shows the structure of the WT and R1810Q side by side for comparison.

J Inherit Metab Dis. Author manuscript; available in PMC 2024 February 16.

Table 1.

Summary of *CAD* variants reported in this study.

Patient[a]	gDNA[b]	cDNA[c]	Protein	CADD[d]	Varity_R[e]	Varity_ER[e]	gnomAD[f]	KO rescue
1	g.27440834	c.172G>A	p.G58S	31.0	0.946	0.934	NA	pathogenic
	g.27464966	c.6071G>A	p.R2024Q[#]	32.0	0.936	0.744	4/280,398	pathogenic
2	g.27445808	c.712C>T	p.R238C	16.5	0.287	0.177	5/25,149	pathogenic
	g.27461045	c.4850G>A	p.R1617Q	23.3	0.225	0.09	26/282,200	pathogenic
3	g.27449499	c.2138T>C	p.I713T	26.8	0.214	0.24	NA	benign
	g.27449499	c.2138T>C	p.I713T	26.8	0.214	0.24	NA	benign
4	g.27462292	c.5347G>A	p.V1783I	9.5	0.04	0.034	15/282,344	benign
	g.27462292	c.5347G>A	p.V1783I	9.5	0.04	0.034	15/282,344	benign
5	g.27460274	c.4424G>A	p.R1475Q	28.6	0.564	0.255	14/251,428	pathogenic
	g.27465630	c.6365G>A	p.R2122H	23.2	0.091	0.074	4/282,570	benign
6	g.27460295	c.4445A>T	p.K1482M	28.3	0.479	0.259	NA	pathogenic
	g.27460295	c.4445A>T	p.K1482M	28.3	0.479	0.259	NA	pathogenic
7	g.27440828	c.166C>G	p.L56V	24.9	0.493	0.529	1/250,026	benign
	g.27445474	c.583C>A	p.Q195K	17.0	0.104	0.083	NA	benign
8	g.27465771	c.6410C>T	p.P2137L	29.8	0.488	0.328	NA	benign
	g.27465771	c.6410C>T	p.P2137L	29.8	0.488	0.328	NA	benign
9	g.27447977	c.1486C>T	p.R496W	26.3	0.219	0.258	8/282,712	benign
	g.27462299	c.5354G>A	p.R1785H	28.6	0.712	0.447	1/251,026	pathogenic
10	g.27454383	c.2335G>T	p.A779S	24.2	0.489	0.337	5/282,840	benign
	g.27462011	c.5164C>T	pR1722W	32.0	0.834	0.598	NA	pathogenic
11	g.27449186	c.2030A>G[i]	p.Q677R	29.7	0.171	0.205	NA	benign[*]
	g.27449186	c.2030A>G[i]	p.Q677R	29.7	0.171	0.205	NA	benign[*]
12	g.27460936	c.4741T>C	p.W1581R	26.2	0.929	0.779	2/251,048	pathogenic
	g.27461498	c.5060A>G	p.H1687R	25.0	0.822	0.543	8/272,670	pathogenic
13	g.27460635	c.4613C>T	p.S1538L	25.1	0.447	0.165	2/282,726	pathogenic
	g.27440885	c.222+1G>A	splicing	35.0	NA	NA	NA	ND
14	g.27445181	c.472del	p.L158Wfs*52	NA	NA	NA	NA	ND

Patient[a]	gDNA[b]	cDNA[c]	Protein	CADD[d]	Varity_R[e]	Varity_ER[e]	gnomAD[f]	KO rescue
	g.27462599	c.5429G>A	p.R1810Q[#]	23.7	0.37	0.129	94/267,140	pathogenic
15	g.27466309	c.6601G>A	p.A2201T	29.8	0.74	0.591	14/242,494	benign
	g.27444233	c.352+18G>A	unknown	2.04	NA	NA	6/233,752	ND

[a] Information on the subject's variants was referred by Hudson Freeze (patients 1–8), Birgit Kaufmann/Saskia Wortmann (9 and 10), Marianne Rohrbach (11), Paula Sánchez-Pintos (12), Janna Kenny (13), Myriam Ley Martos (14), and Teresa Campos (15).

[b] Genome build GRCh37 / hg19

[c] Human CAD cDNA (NM_004341.4), UniProt P27708, NP_004332.2

[d] CADD (https://cadd.gs.washington.edu). A value of 20 puts variant in top 1% of deleterious variants, 30 in top 0.1%.

[e] VARITY_R (http://varity.varianteffect.org). Higher values indicate increased pathogenicity prediction. VARITY_R and VARITY_ER include only rare (R) or extremely rare (ER) ClinVar variants for calculations.

[f] Allele count per total number of called high-quality genotypes.[11]

[#] Variants already tested: R2024,[7,25] and R1810Q.[8]

[*] This variant changes the last base of the exon and affects splicing