Molecular Genetics and Metabolism Reports 37 (2023) 101016

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism Reports

journal homepage: www.elsevier.com/locate/ymgmr

Predicting the pathogenicity of missense variants based on protein instability to support diagnosis of patients with novel variants of ARSL

Eriko Aoki [a,1], Noriyoshi Manabe [b,1], Shiho Ohno [b,1], Taiga Aoki [c], Jun-Ichi Furukawa [d], Akira Togayachi [a], Kiyoko Aoki-Kinoshita [a], Jin-Ichi Inokuchi [b], Kenji Kurosawa [e], Tadashi Kaname [c,*], Yoshiki Yamaguchi [b,**], Shoko Nishihara [a,***]

[a] Glycan & Life Systems Integration Center (GaLSIC), Soka University, Hachioji 192-8577, Japan
[b] Division of Structural Glycobiology, Institute of Molecular Biomembrane and Glycobiology, Tohoku Medical and Pharmaceutical University, Sendai 981-8558, Japan
[c] Department of Genome Medicine, National Center for Child Health and Development, Tokyo 157-8535, Japan
[d] Institute for Glyco-Core Research (iGCORE), Nagoya University, Nagoya 466-8550, Japan
[e] Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama 232-8555, Japan

ARTICLE INFO

Keywords:
Arylsulfatase L
Rare disease
Missense variant
Stabilization energy
Sulfatase activity
Protein instability

ABSTRACT

Rare diseases are estimated to affect 3.5%–5.9% of the population worldwide and are difficult to diagnose. Genome analysis is useful for diagnosis. However, since some variants, especially missense variants, are also difficult to interpret, tools to accurately predict the effect of missense variants are very important and needed. Here we developed a method, "VarMeter", to predict whether a missense variant is damaging based on Gibbs free energy and solvent-accessible surface area calculated from the AlphaFold 3D protein model. We applied this method to the whole-exome sequencing data of 900 individuals with rare or undiagnosed disease in our in-house database, and identified four who were hemizygous for missense variants of arylsulfatase L (ARSL; known as the genetic cause of chondrodysplasia punctata 1, CPDX1). Two individuals had a novel Ser89 to Asn (Ser89Asn) or Arg469 to Trp (Arg469Trp) substitution, respectively predicted as "damaging" or "benign"; the other two had an Arg111 to His (Arg111His) or Gly117 to Arg (Gly117Arg) substitution, respectively predicted as "damaging" or "possibly damaging" and previously reported in patients showing clinical manifestations of CDPX1. Expression and analysis of the missense variant proteins showed that the predicted pathogenic variants (Ser89Asn, Arg111His, and Gly117Arg) had complete loss of sulfatase activity and reduced protease resistance due to destabilization of protein structure, while the predicted benign variant (Arg469Trp) had activity and protease resistance comparable to those of wild-type ARSL. The individual with the novel pathogenic Ser89Asn variant exhibited characteristics of CDPX1, while the individual with the benign Arg469Trp variant exhibited no such characteristics. These findings demonstrate that VarMeter may be used to predict the deleteriousness of variants found in genome sequencing data and thereby support disease diagnosis.

1. Introduction

Rare diseases are epidemiologically estimated to affect 3.5%–5.9% of the population, equating to about 350 million people worldwide. They are usually severe, chronic and disabling, with onset in childhood or young adulthood, and approximately 80% have a genetic origin [1]. Genetic information on >9000 gene-related human diseases has been registered in the Online Mendelian Inheritance in Man [2,3] and it is estimated that this number will reach tens of thousands in the future [4]. Currently, comprehensive genome analysis for patients with these

[*] Correspondence to: T Kaname, Department of Genome Medicine, National Center for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan.
[**] Correspondence to: Y Yamaguchi, Division of Structural Glycobiology, Institute of Molecular Biomembrane and Glycobiology, Tohoku Medical and Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai 981-8558, Miyagi, Japan.
[***] Correspondence to: S Nishihara, Glycan & Life Systems Integration Center (GaLSIC), Soka University, 1-236 Tangi-machi, Hachioji, Tokyo 192-8577, Japan.
E-mail addresses: kaname-t@ncchd.go.jp (T. Kaname), yyoshiki@tohoku-mpu.ac.jp (Y. Yamaguchi), shoko@soka.ac.jp (S. Nishihara).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ymgmr.2023.101016
Received 25 July 2023; Received in revised form 20 October 2023; Accepted 21 October 2023
Available online 29 October 2023

diseases is gradually clarifying their causes. For example, among the undiagnosed cases for which whole-exome sequencing analysis (WES analysis) has been performed at one department in the National Center for Child Health and Development, Japan, the cause has been identified in about 40%, while 60% of cases remain unexplained [in preparation].

The most common cause of genetic disorders are single nucleotide variants (SNVs), including missense, stop-gain, stop-loss, indel, splicing and other variants, of which the most frequent are missense variants leading to amino acid substitutions within a protein [5]. However, although there are several in silico prediction tools available, missense variants are difficult to estimate their effects as compared with other SNVs such as indel or splicing variants. Therefore, accurate prediction tools for effects of missense variants on proteins from genomic data are very important in their interpretation, which results can provide us useful information for the diagnosis of rare diseases.

Arylsulfatase L (ARSL), formerly known as arylsulfatase E, is a member of the sulfatase family encoded on the X-chromosome [6] and is well-established genetic cause of X-linked recessive brachytelephalangic chondrodysplasia punctata (CDPX1) (OMIM no. 302950). This pan-ethnic congenital rare disorder affects males and it is characterized by chondrodysplasia punctata or stippled epiphyses, brachytelephalangy or shortening of the distal phalanges, nasomaxillary hypoplasia and mixed conductive and sensorineural hearing loss. ARSL, comprising 589 amino acids, is expressed in multiple tissues [6] and localizes in the Golgi apparatus [7]. The 60-kDa precursor is converted into the mature 68-kDa form by *N*-glycosylation at four potential sites (Asn58, Asn125, Asn258, and Asn344). ARSL has a conserved catalytic domain and two predicted transmembrane helices, through which it is likely to be anchored to the Golgi cisternae. Although the physiological substrate of ARSL is unknown, its sulfatase activity is generally measured by using an artificial substrate, fluorogenic 4-methylumbelliferyl sulfate (4-MU sulfate) [8].

In this study, we developed a method for predicting deleterious missense variants based on Gibbs free energy and solvent-accessible surface area using the 3D structure of a protein. The prediction method was established by using information on previously reported missense ARSL variants [7–10], chitobiosyldiphosphodolichol β-mannosyltransferase (ALG1) variants [11], and mannose-6-phosphate isomerase (MPI) variants [12–19]. We applied our newly developed method, named VarMeter (VARiant impact-predicting MEthod combining muTation Energy and solvent-accessible surface aRea), to in-house genomic data associated with rare or undiagnosed disease held in the Department of Genome Medicine, National Center for Child Health and Development, leading to the identification of a novel pathogenic variant of ARSL possessing the amino acid substitution Ser89 to Asn (Ser89Asn). This variant showed no sulfatase activity and loss of resistance to trypsin digestion, verifying the destabilization of protein structure. Since the individual with the Ser89Asn variant showed the clinical features CDPX1, suggesting that the variant was pathogenic and causative. Thus, our prediction method using Gibbs free energy and solvent-accessible surface area may be useful for predicting the deleteriousness of variants in genomic data from individuals with undiagnosed diseases.

2. Materials & methods

2.1. Patients and genomic data analysis

The present study used data from patients with rare or undiagnosed disease who had given written informed consent to participate in research studies involving comprehensive genome analysis [20]. Genomic DNA extracted from the patients' blood cells was used for whole-exome sequencing, as described previously [20]. The presence of the *ARSL* gene variants was investigated in the in-house whole-exome sequencing data of 900 patients with rare or undiagnosed disease, and hemizygous missense variants were identified in male patients.

2.2. Selection of pathogenic and benign variants of ARSL and clinical criteria in the literature

Based on CDPX1 publications reported to date and data included in the Human Gene Mutation Database (HGMD), pathogenic variants of the *ARSL* gene were selected as variants of patients with clinically diagnosed chondrodysplasia punctata that were experimentally confirmed to have reduced activity of the product of ARSL enzyme. Benign variants were also selected as variants that had a high allele frequency in the population (> 0.05) and were found in healthy males, based on the gnomAD database (v2.1.1) (https://gnomad.broadinstitute.org/). The clinical criteria of chondrodysplasia punctata (CDP) were as follows: clinical manifestations characteristic of CDP, such as punctate calcifications in radiographs, characteristic facial features (e.g. nasal hypoplasia, depressed nasal bridge), and diagnosis by a clinical geneticist. These clinical criteria were also applied to our patients.

2.3. Selection of known pathogenic and benign variants of ALG1 and MPI

To determine more appropriate thresholds of pathogenicity prediction, we used ALG and MPI variant data selected from HGMD or the gnomAD database, in addition to ARSL variant data. For ALG1 and MPI, pathogenic variants experimentally confirmed to have less activity [11–19] were selected from those included in HGMD. Variants with a clinical significance of "benign" and a homozygote count of at least 1 were selected as benign from the gnomAD database.

2.4. Preparation of 3D models of variant proteins

The amino acid sequence of human ARSL was obtained from the Universal Protein Resource database (UniProt ID: P51690). The 3D structural models of ARSL, ALG1 [11] and MPI [12–19] were obtained from the AlphaFold2 Protein Structure Database [21], and a 3D model of each variant was prepared using the Calculate Mutation Energy/Stability module in Discovery Studio 2021 (BIOVIA, Dassault Systèmes, San Diego, CA, USA).

2.5. Calculation of mutation energy

Mutation energy (ΔΔ*G*mut), indicating the effect of variant on protein stability, was calculated with the Calculate Mutation Energy/Stability module in Discovery Studio 2021. ΔΔ*G*mut corresponds to the difference in Gibbs free energy of folding (ΔΔ*G*folding) between wild-type (WT) and variant proteins, while ΔΔ*G*folding is the energy difference between the folded and unfolded states of the protein (Δ*G*folded and Δ*G*unfolded):

$$\Delta\Delta G\text{mut} = \Delta\Delta G\text{folding (mutant)}–\Delta\Delta G\text{folding (WT)}$$

$$\Delta\Delta G\text{folding} = \Delta G\text{folded}–\Delta G\text{unfolded}$$

The total Gibbs free energy of folded or unfolded state is calculated as the weighted sum of energy terms:

$$\Delta G\text{total} = 0.5 \times \Delta G\text{vdW} + 0.5 \times \Delta G\text{elec}(T) + 0.8 \times \Delta G\text{entr}$$

where Δ*G*total is total Gibbs free energy, Δ*G*vdW is van der Waals interaction, Δ*G*elec is electrostatic interaction, and Δ*G*entr is an entropy contribution related to side-chain mobility. These energy terms were calculated with the CHARMM (version 44.2) force field [22,23], and electrostatics energy was calculated using a Generalized Born implicit solvent model [24,25]. The mutation energy of the *ARSL* variant with a substitution in the putative transmembrane region (Gly245Arg) was calculated after adding an implicit membrane to the corresponding region in Discovery Studio 2021.

2.6. Calculation of normalized solvent-accessible surface area

Solvent-accessible surface area (SASA) was defined as the contact area with the water molecule, which was assumed to be a rigid sphere with a radius of 1.4 Å. SASA of each residue was calculated using Discovery Studio 2021 based on the AlphaFold2 models of WT ARSL, ALG1 and MPI. Substitution at the putative transmembrane region of ARSL (Gly245Arg) was treated as SASA = 0 Å^2. The normalized SASA (nSASA) for each residue was calculated as follows:

$$\text{nSASA} = (\text{SASA obtained from AlphaFold2 model}) / (\text{SASA under the denatured state})$$

where SASA under the denatured state was based on previously reported values [26].

2.7. Statistical analysis

Each group of continuous variables (mutation energy [pathogenic, benign], nSASA [pathogenic, benign]) was tested for normal distribution by the Shapiro-Wilk test [27] and Anderson-Darling test [28]. Between-group (damaging, benign) differences for continuous variables were tested by Mann-Whitney *U* test.

2.8. Receiver operating characteristic (ROC) analysis

ROC curve analysis was performed to determine an appropriate threshold of nSASA and mutation energy in discriminating pathogenic and benign variants. Variant information (missense variant and its effect benign or pathogenic) were extracted from the literatures for three proteins: ARSL (pathogenic; $n = 25$, benign; $n = 7$), ALG1 (pathogenic; $n = 30$, benign; $n = 6$), and MPI (pathogenic; $n = 15$, benign; $n = 3$). ROC curves were plotted and analyzed using GraphPad Prism 7.05 (GraphPad Software). The thresholds for nSASA and mutation energy were determined to provide the shortest distance of the curve from the top-left corner [29].

2.9. Plasmid construction

All expression vectors were generated using the Gateway cloning system (Thermo Fisher Scientific, Waltham, MA, USA). In brief, the gene encoding arylsulfatase L (*ARSL*) was amplified from plasmid pFN21AE6750 (Kazusa DNA Research Institute, Chiba, Japan) using attB adaptor primers, and recombined into pDONR201 (Thermo Fisher Scientific, Waltham, MA, USA) to generate an entry clone. The *ARSL* gene encoded in this vector contains two nucleotide variants, including one missense variant relative to NCBI reference sequence NM_000047. The variant leading to amino acid substitution was reverted so that the translated sequence was the same as that from NCBI reference sequence NM_000047; the resulting protein is referred to as WT in this study. Nucleotide substitutions (Supplementary Table 1) were introduced by site-directed mutagenesis using the QuikChange II Site-directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). The entry clones encoding WT or variant *ARSL* genes were recombined via LR reaction into plasmid pCAGI-puro (a kind gift of Professor Kumiko Ui-Tei), which has a FLAG tag at the C-terminus via the linker (Asp-Pro-Ala-Phe-Leu-Tyr-Lys-Val-Val-Asp).

2.10. HeLa cell transfection

HeLa cells were maintained in DMEM (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% FBS (BioWest, Nuaillé, France), 100 units/mL of penicillin and 100 μg/mL of streptomycin (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C and 5% CO_2. Cells were seeded in a 10-cm dish at 1.8×10^6 cells and incubated for 24 h. Cultured cells were transfected with 12 μg of expression vector using 24 μL of Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Cells were grown for an additional 24 h, and the medium was replaced with fresh medium containing 2 μg/mL of puromycin (*Merk, Darmstadt, Germany*) to select transfected cells. After 24 h of puromycin selection, cells were washed with phosphate-buffered saline (PBS) and detached by treatment with 0.05% Trypsin-EDTA (Thermo Fisher Scientific, Waltham, MA, USA). The harvested cells were washed three times with PBS and stored at −20 °C.

2.11. Protein extraction and purification

The stored cell pellet was thawed on ice, resuspended in 50 mM Tris-HCl (pH 7.5) and 150 mM NaCl, and lysed by an additional freeze–thaw treatment. The suspension was centrifuged at 20,000 ×*g* for 10 min to remove the soluble fraction. The precipitate was resuspended in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl and 1% Triton X-100, and incubated on ice with occasional mixing for 30 min. The supernatant containing solubilized protein by Triton X-100 was obtained by centrifugation at 20,000 ×*g* for 10 min and incubated with Anti-FLAG M2 Magnetic Beads (*Merk, Darmstadt, Germany*) for 1 h on ice. After removing the supernatant, the beads were washed with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl and 1% Triton X-100. The protein was eluted with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% Triton X-100 and 100 μg/mL of DYKDDDDK Peptide (FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan). The eluted fractions were analyzed by SDS-PAGE, and the purified WT and variant ARSL proteins were separated via SDS-PAGE, stained with Lumitein (Biotium, Fremont, CA, USA), and detected using Sayaca Imager (DRC CO., LTD., Tokyo, Japan). The purified protein was quantified using ImageJ [30], and the concentration was estimated by comparing band intensity with a BSA or BamA fragment (residues 421–810) of known concentration.

2.12. ARSL activity assay

Enzyme activity of ARSL was measured as described previously with a slight modification whereby we used purified ARSL protein whose concentration was estimated using BSA as a standard [8,9,31]. In brief, 0.015 pmol/μL of purified ARSL in 50 mM Tris-HCl (pH 7.5), 30 mM NaCl, 20 μg/mL of DYKDDDDK Peptide and 1% Triton X-100 was added to a double volume of 0.4 mM 4-MU sulfate (*Merk, Darmstadt, Germany*). During incubation at 37 °C, an aliquot of reaction solution was withdrawn at various time points, the reaction was stopped by adding a 9-fold volume of 200 mM Glycine-NaOH (pH 10.7), and 100 μL was transferred to a 384-well plate. Fluorescence measurements were performed on a Varioskan LUX instrument (Thermo Fisher Scientific, Waltham, MA, USA) at 25 °C. The excitation wavelength was 360 nm, and the emission wavelength was 449 nm. The bandwidth was 12 nm for excitation, and the measurement time was 100 ms. Standard curves were prepared by measuring the fluorescence of different concentrations of 4-MU and 4-MU sulfate. The fluorescence observed in the enzyme reaction mixture included fluorescence from 4-MU produced by enzymatic reaction and 4-MU sulfate remaining after the reaction. The concentration of 4-MU produced by enzymatic reaction was estimated by using the standard curves of 4-MU and 4-MU sulfate.

2.13. Susceptibility to trypsin digestion assay

Wild-type and variant ARSL proteins, whose concentrations were estimated using a BamA fragment (residues 421–810) as a standard, were diluted to 0.8 ng/μL in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1% Triton X-100 and 100 μg/mL DYKDDDDK peptide, and mixed with 20 ng/μL of sequencing-grade modified trypsin (Promega, Madison, WI, USA) in an 8:1 ratio. Samples were also prepared using the same buffer without trypsin as a control. The reaction was incubated at 37 °C for 2 h, and stopped by adding PMSF to a final concentration of 1 mM. The samples were subjected to SDS–PAGE followed by Lumitein staining. In the absence of trypsin, the band corresponding to the molecular weight of full-length ARSL protein was quantified using ImageJ, while the trypsin-resistant fragment of ARSL, whose molecular weight is slightly lower than that of full-length ARSL protein, was quantified in the presence of trypsin. The amount of ARSL protein remaining after trypsin digestion was estimated as a proportion of the band intensity of trypsin-resistant fragment to that of full-length ARSL protein quantified from trypsin-untreated samples. For the variant ARSL proteins, this value was normalized to that of ARSL WT protein. The data were analyzed using Kaleida Graph (Synergy Software, PA, *USA*), and differences were assessed by one-way ANOVA followed by Tukey HSD test ($\alpha = 0.05$).

3. Results

3.1. Estimation of the instability of missense ARSL variants using mutation energy and nSASA

We investigated the effects of missense variants from the viewpoint of protein 3D structure, with a particular focus on two parameters: “mutation energy”, calculated as the difference in Gibbs free energy between WT and variant proteins; and “nSASA” at the variant site. Our rationale for using these parameters comes from the fact that protein function is highly correlated with energetic stability (Gibbs free energy) [32]. In addition, variant of the inner hydrophobic core will greatly affect protein stability [33]. To establish a prediction method based on these parameters, we first chose a model protein, arylsulfatase L (ARSL), whose missense pathogenic variants have been well characterized experimentally and discussed in relation to CDPX1 [7,9,10,31]. The variants were categorized into two groups based on sulfatase activity or population survey: a pathogenic group (25 variants, showing experimentally little or no activity), and a benign group (7 variants, showing a high allele frequency in the population of healthy males ($>$ 0.05)).

We obtained the 3D structural model of WT ARSL from the AlphaFold protein structure database and created 3D models of each variant based on the WT model. For the 25 pathogenic and 7 benign variants, the mutation energy and nSASA were calculated, along with SIFT, PolyPhen-2 and CADD scores (Table 1). We tested each group of continuous variables (mutation energy [pathogenic, benign], nSASA [pathogenic, benign]) for normal distribution by Shapiro-Wilk and Anderson-Darling tests. At a significance level of 0.05, mutation energy (benign) and nSASA (benign) were both consistent with a normal distribution, whereas mutation energy (pathogenic) and nSASA (pathogenic) were not consistent. Therefore, between-group (pathogenic, benign) differences in continuous variables were tested by Mann-Whitney *U* test (Fig. 1A and B). At a significance level of 0.05, both mutation energy and nSASA were significantly different between pathogenic and benign variants. These data demonstrated that mutation energy and nSASA may be good parameters by which to predict the severity of each variant.

To predict the given missense variants as pathogenic or benign,

Table 1
List of damaging and benign variants of ARSL identified by VarMeter and the corresponding scores and predictions from SIFT, PolyPhen-2, and CADD.

Variant (AA change)	Mutation energy (kcal/mol)	nSASA	VarMeter (Mutation energy and nSASA)	SIFT Score	SIFT	PolyPhen-2 Score	PolyPhen-2	CADD (PHRED)	CADD
[Pathogenic variants, $n = 25$]									
Ile40Ser	3.6	0.00	D	0.01	D	0.992	D	22.6	D
Asp47Asn	−2.6	0.00	PD	0.13	B	0.985	D	22.6	D
Gly57Ser	1.1	0.73	PD	0.00	D	0.990	D	23.2	D
Gly73Ser	15.2	0.00	D	0.00	D	0.986	D	22.4	D
His79Tyr	3.6	0.00	D	0.54	B	0.776	N	22.1	D
Ile80Asn	2.0	0.02	D	0.02	D	0.314	B	23.0	D
Ser89Gly	−0.1	0.00	PD	0.00	D	0.997	D	22.7	D
Arg90Gly	5.8	0.02	D	0.00	D	1.000	D	21.9	D
Thr95Met	1.0	0.00	D	0.00	D	0.997	D	22.9	D
Arg111Pro	9.2	0.09	D	0.27	B	0.427	B	16.7	B
Gly120Glu	13.3	0.01	D	0.01	D	0.805	N	22.6	D
Gly137Val	0.2	0.60	B	0.00	D	0.997	D	22.5	D
Gly137Ala	0.0	0.60	B	0.00	D	0.952	D	22.2	D
Gly149Cys	14.1	0.07	D	0.00	D	1.000	D	23.7	D
Gly245Arg	−1.4	0.00	PD	0.35	B	0.051	B	5.6	B
Thr306Ala	1.2	0.18	PD	0.70	B	0.264	B	15.5	B
Gly317Arg	−0.3	0.42	B	0.00	D	0.999	D	22.8	D
Gly355Ser	7.1	0.00	D	0.00	D	0.999	D	22.6	D
Gly377Glu	9.3	0.00	D	0.00	D	0.999	D	23.0	D
Gly391Arg	15.9	0.00	D	0.17	B	0.998	D	23.0	D
Thr409Met	1.5	0.00	D	0.04	D	0.997	D	23.3	D
Gly434Ser	12.1	0.08	D	0.00	D	0.897	N	23.1	D
Ala463Thr	10.2	0.00	D	0.21	B	0.554	N	17.6	B
Thr481Met	0.8	0.07	PD	0.00	D	0.994	D	23.1	D
Pro578Ser	1.2	0.55	PD	0.59	B	0.973	D	22.3	D
[Benign variants, $n = 7$]									
Ile53Val	0.7	0.00	PD	0.12	B	0.009	B	0.0	B
Ser156Asn	0.6	0.75	B	0.50	B	0.004	B	0.0	B
Arg183His	0.3	0.61	B	0.14	B	0.001	B	0.0	B
Val222Ile	−1.2	0.31	B	0.53	B	0.004	B	0.2	B
Gly424Ser	3.3	0.16	PD	0.53	B	0.490	N	15.8	B
Arg469Tyr	0.5	0.38	B	0.02	D	0.225	B	19.4	B
Gly490Ser	3.2	0.11	D	0.50	B	0.313	B	10.3	B

AA: amino acid.

[Footnote] For VarMeter, the mutation energy and normalized solvent-accessible surface area (nSASA) were interpreted as follows: D, damaging (mutation energy≥0.88 kcal/mol *and* nSASA≤0.11); PD, probably damaging (mutation energy≥0.88 kcal/mol *or* nSASA≤0.11); B, benign (mutation energy<0.88 kcal/mol *and* nSASA>0.11). For the other tools, the score was interpreted as follows: D, damaging/probably damaging (SIFT score ≤ 0.05, PolyPhen-2 score ≥ 0.9, CADD score ≥ 20); N, possibly damaging (PolyPhen-2 score > 0.45); B, benign (SIFT score > 0.05, PolyPhen-2 score ≤ 0.45, CADD score < 20). ARSL: NP_000038.2 (NM_000047.3).

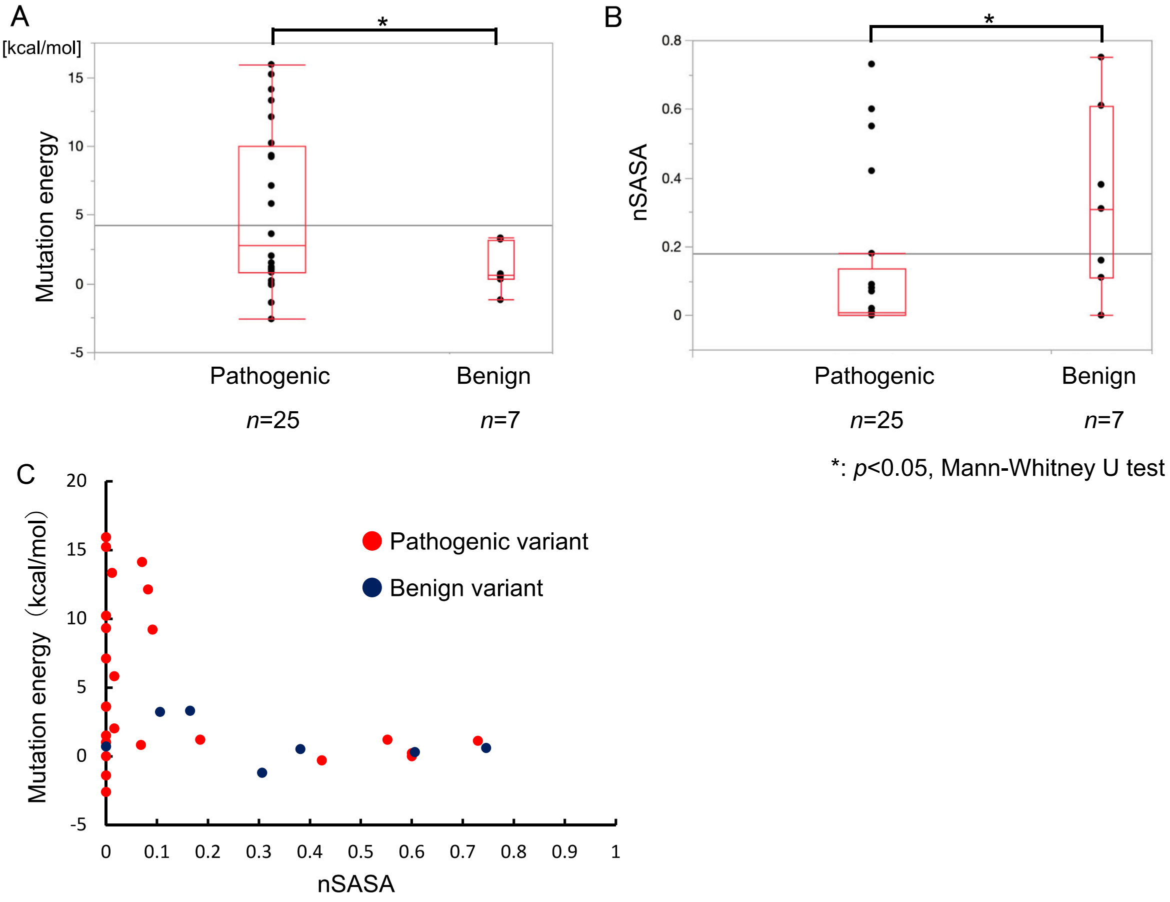

Fig. 1. Box- and scatterplots of calculated mutations energy and nSASA for pathogenic and benign variants of ARSL. (A) Mutation energy. (B) nSASA. In (A) and (B), the horizontal line in each box indicates the median and the box limits indicate the 25th and 75th percentiles. The whiskers indicate the range from minimum to maximum. * Significant at $p < 0.05$ by Mann-Whitney U test. (C) Scatter plot of mutation energy versus nSASA. Pathogenic variants are plotted as red circles, and benign variants are plotted as dark blue circles.

appropriate thresholds need to be determined for nSASA and mutation energy. We used receiver operating characteristic (ROC) plots of nSASA and mutation energy to find the best threshold for distinguishing pathogenic and benign variants. We initially aimed to determine the thresholds of nSASA and mutation energy using ARSL variant data, but we were concerned that the number of data was insufficient to determine the thresholds (pathogenic variants, $n = 25$; benign variants, $n = 7$). To increase the accuracy of prediction, therefore, variant data from two additional proteins (ALG1 and MPI), whose activities in patients had been experimentally measured in many studies, were added to the ARSL

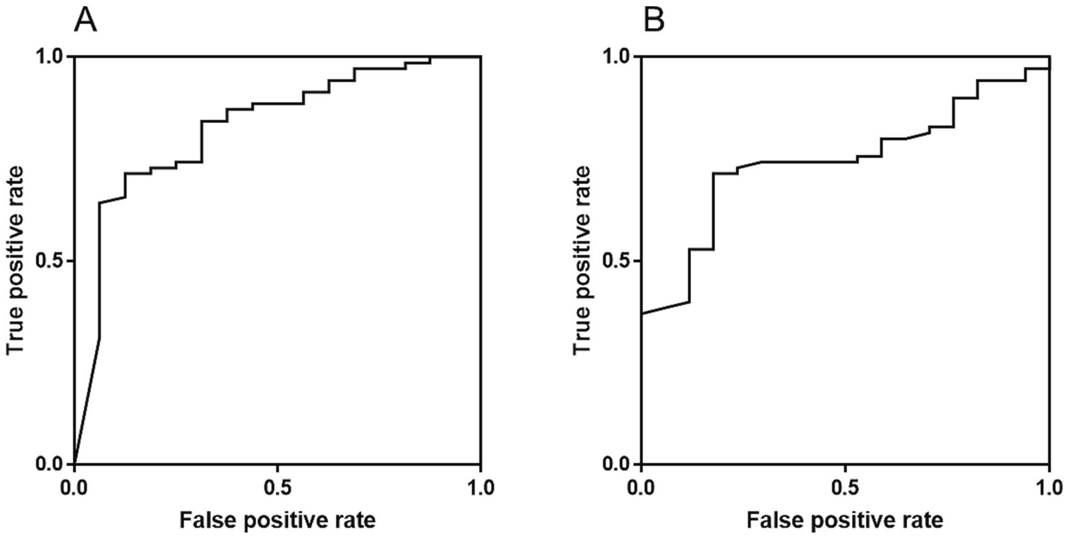

Fig. 2. ROC curve analysis for classification of missense variants using nSASA and the mutation energy. Missense data of three proteins (ARSL, ALG1 and MPI, pathogenic; $n = 70$, benign; $n = 16$) were used in the ROC analysis. ROC curves using nSASA (A) and mutation energy (B) are shown.

data, resulting in a total of 70 pathogenic and 16 benign variants (Supplementary Table 2 and 3). ROC curves were plotted for nSASA (Fig. 2A) and mutation energy (Fig. 2B). The utility of each parameter was quantified by the area under the curve (AUC), which corresponds to the probability of a randomly chosen pathogenic variant being assigned as pathogenic, not as benign. The AUC was 0.83 for nSASA, and 0.75 for mutation energy, suggesting that nSASA and mutation energy might be good parameters for prediction. According to the distance-to-corner metric, a threshold of 0.11 was set for nSASA and 0.88 kcal/mol for mutation energy. Of course, these thresholds can be updated by including additional variant data from other proteins, and we will continue to update them in the future. Using these thresholds, we here propose the following method, named VarMeter, to predict the effect of a missense variant:

Requirement A : $\Delta\Delta G\mathrm{mut} \geq +0.88$ kcal/mol.

Requirement B : nSASA ≤ 0.11.

(1) A variant satisfying both requirements A and B is predicted as "damaging".
(2) A variant satisfying requirement A or B is predicted as "possibly damaging".
(3) A variant not satisfying requirement A or B is predicted as "benign".

Based on these criteria, variants (1) and (2) are predicted as pathogenic, and variant (3) is predicted as benign. The percentage accuracy of this prediction was calculated as 81% for ARSL.

To compare VarMeter with other prediction methods, the effect of ARSL variants was also predicted using SIFT [34], PolyPhen-2 [35] and CADD [36] (Table 1). SIFT predicts a variant as "damaging" at a SIFT score of ≤0.05, and Polyphen-2 predicts "probably damaging" at a score of ≥0.9, while CADD (PHRED) predicts "damaging" at a score of ≥20. Based on these criteria, the percentage accuracy of prediction was estimated as 72%, 75% and 88% for SIFT, PolyPhen-2 and CADD, respectively. For ARSL variants ($n = 32$, pathogenic and benign), inaccurate prediction by VarMeter relative to other methods is summarized in Table 2. The percentage of accurate SIFT prediction/inaccurate VarMeter prediction was 19%, that of accurate PolyPhen-2 prediction/inaccurate VarMeter prediction was 16%, and that of accurate CADD prediction/inaccurate VarMeter prediction was 19%. In contrast, percentage of accurate VarMeter prediction/inaccurate SIFT prediction was 28%, that of accurate VarMeter prediction/inaccurate PolyPhen-2 prediction was 13%, and that of accurate VarMeter prediction/inaccurate CADD prediction was 13%. Thus, our prediction method using the mutation energy and nSASA has a prediction accuracy comparable to widely used prediction tools and may be a complementary approach. Because our method is simple and uses only two parameters, it can be applied to high-throughput screening and also coupled with other prediction methods to improve evaluation of variants.

Table 2
Percentage of VarMeter prediction accuracy and inaccuracy relative to other methods for ARSL variants ($n = 32$, pathogenic and benign).

	VarMeter vs SIFT	VarMeter vs PolyPhen-2	VarMeter vs CADD
Both accurate	53	69	69
VarMeter accurate and opponent inaccurate	28	13	13
VarMeter inaccurate and opponent accurate	19	16	19
Both inaccurate	0	3	0

3.2. Estimation of variants of ARSL found in undiagnosed individuals by VarMeter and identification of novel damaging variants

We applied our new prediction method VarMeter to *ARSL* variants in our in-house database of patients with rare or undiagnosed disease. In the whole-exome sequencing data of 900 patients with rare or undiagnosed diseases, single nucleotide substitutions in the *ARSL* gene were investigated. We identified four males who had a hemizygous variant, p. Ser89Asn, p.Arg111His, p.Gly117Arg or p.Arg469Trp, in ARSL (Fig. 3, Table 3), of which p.Ser89Asn and p.Arg469Trp have not been reported previously or registered in the public database. The mutation energies of these variants were calculated as +3.4 kcal/mol (Ser89Asn), +3.1 kcal/mol (Arg111His), −1.1 kcal/mol (Gly117Arg) and + 0.5 kcal/mol (Arg469Trp), while the nSASA values were 0 (Ser89Asn), 0.09 (Arg111His), 0 (Gly117Arg) and 0.38 (Arg469Trp). Based on our criteria defined above, Ser89Asn and Arg111His were predicted as "damaging", Gly117Arg as "possibly damaging", and Arg469Trp as "benign" variants.

3.3. Loss of arylsulfatase activity in missense ARSL variants predicted as "damaging" or "possibly damaging" by VarMeter

Next, we investigated the function and stability of the ARSL variants. Wild-type ARSL, the three ARSL variants predicted as "damaging" or "possibly damaging" by VarMeter (Ser89Asn, Arg111His and Gly117Arg) and the variant predicted as "benign" (Arg469Trp). The activity of ARSL Gly117Arg has been reported to be negligible in an activity assay performed on lysate from COS cells transfected with ARSL Gly117Arg expression plasmid [9]. Our activity assay used purified ARSL variants, which were expressed in HeLa cells with a FLAG-tag at the C-terminus and purified with anti-FLAG antibody-conjugated beads.

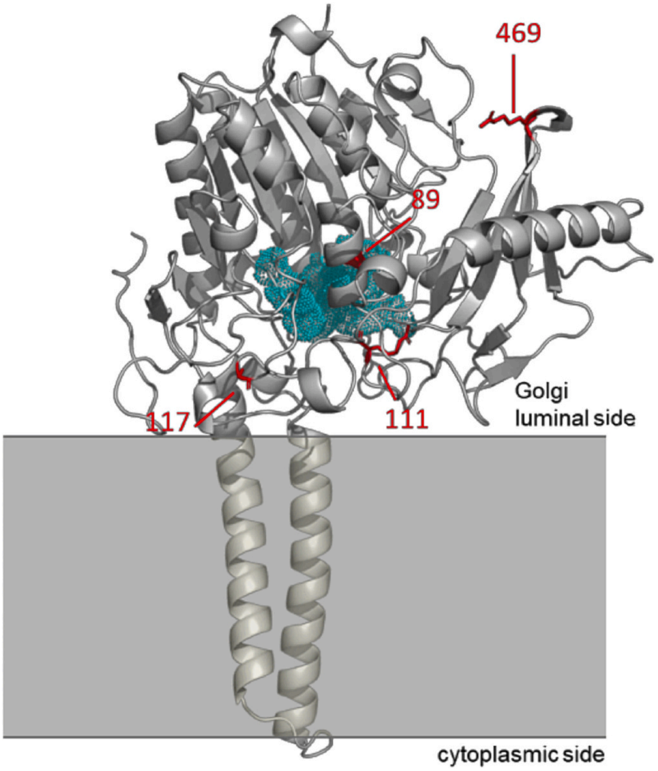

Fig. 3. The 3D structural AlphaFold model of WT ARSL. Side chains at the substitution sites (Ser89, Arg111, Gly117 and Arg469) are shown in stick representation in red. The putative substrate-binding region is shown in sky blue. The figure was prepared using PyMOL software. The signal sequence has been removed.

Table 3
Four hemizygous ARSL variants identified in 900 individuals with rare or undiagnosed disease in the in-house database.

Variant (AA change)	Mutation energy (kcal/mol)	nSASA	Prediction by VarMeter (Mutation energy and nSASA)	Novel or previously published	Sulfatase activity	Protease resistance	Clinical diagnosis
Ser89Asn	3.4	0	Damaging	Novel	Complete loss	Reduced	CDPX1
Arg469Trp	0.5	0.38	Benign	Novel	Comparable to wild-type	Not reduced	Not CDPX1
Arg111His	3.1	0.09	Damaging	Reported [37]	Complete loss	Reduced	CDPX1
Gly117Arg	−1.1	0	Possibly damaging	Reported [8]	Complete loss	Reduced	CDPX1

The C-terminal tag attached via the linker was sufficiently distant from the active site and substrate entrance based on the AlphaFold model of ARSL. These purified proteins were detected on SDS–PAGE as a band at approximately 70 kDa (Fig. 4A), consistent with the size of the mature *N*-glycosylated ARSL protein (68 kDa) [7]. To determine enzyme activity, 4-MU sulfate was added to the purified protein solution, and the amount of 4-MU produced after 4-h incubation was estimated from fluorescence measurements (Fig. 4B). Fluorescence from 4-MU was detected in the reaction mixture with WT ARSL or the Arg469Trp variant, demonstrating that WT ARSL and the Arg469Trp variant have arylsulfatase activity. In contrast, 4-MU was not detected in the reaction mixture with the Ser89Asn, Arg111His or Gly117Arg variant.

To confirm the lack of enzyme activity for the Ser89Asn, Arg111His and Gly117Arg variants, a time course experiment of arylsulfatase activity over a longer incubation time was conducted (Fig. 4C). The concentration of 4-MU increased as the reaction proceeded and reached a plateau at about 5 h for WT and the Arg469Trp variant. In contrast, no production of 4-MU was detected even after a reaction time of 24 h for the Ser89Asn, Arg111His and Gly117Arg variants. These results clearly showed that arylsulfatase activity was significantly impaired in the Ser89Asn, Arg111His and Gly117Arg ARSL variants predicted as “damaging” and “possibly damaging” by VarMeter, whereas the Arg469Trp variant predicted as “benign” had activity comparable to WT (Table 3). Thus, the activities of the variants were consistent with, and demonstrate the validity of, predictions made by VarMeter.

3.4. *Reduced protease resistance of ARSL variant proteins predicted as “damaging” or “possibly damaging” by VarMeter*

VarMeter predicts the pathogenicity of variants based on Gibbs free energy and nSASA, two parameters that are highly correlated to protein stability. Because incompletely folded proteins, such as destabilized proteins, are much more sensitive to protease digestion than a native protein [38], we analyzed the protease resistance of WT ARSL and the Ser89Asn, Arg111His, Gly117Arg and Arg469Trp variants as a measure of protein stability.

Purified WT and variant proteins were treated with trypsin and analyzed by SDS–PAGE (Fig. 5A). For all variants, a band corresponding to full-length ARSL was observed in untreated samples. For WT, a band corresponding to full-length ARSL was shifted to a slightly lower molecular weight after trypsin digestion (Fig. 5A and Supplementary Fig. 1). This slight change of molecular weight might be caused by trypsin digestion of the flexible linker and FLAG tag attached at the C-terminus, because the folded regions are resistant to digestion, resulting in a trypsin-resistant ARSL fragment (~70 kDa). This ~70 kDa trypsin-resistant was observed in the Arg469Trp variant predicted as “benign”, and the band intensity did not seem to differ from WT. In contrast, the ~70 kDa trypsin-resistant fragment was hardly observed in Ser89Asn and Gly117Arg, and slightly observed in Arg111His. These variants were predicted as “damaging” or “possibly damaging”. For each variant, we quantified the band intensity of the trypsin-resistant fragment as a proportion of the full-length ARSL band intensity in the untreated sample, and normalized the value to that of WT (Fig. 5B). As a result, there was no difference in the proportion of the remaining trypsin-resistant fragment between WT and Arg469Trp. In contrast, the amount of trypsin-resistant fragment remaining after digestion of Ser89Asn, Arg111His and Gly117Arg was significantly smaller ($p < 0.0001$), showing that the stability of these variants was considerably reduced as compared with WT. Thus, similar to the arylsulfatase activity of the ARSL variants observed above, the protease resistance of the variants was consistent with the predictions by VarMeter.

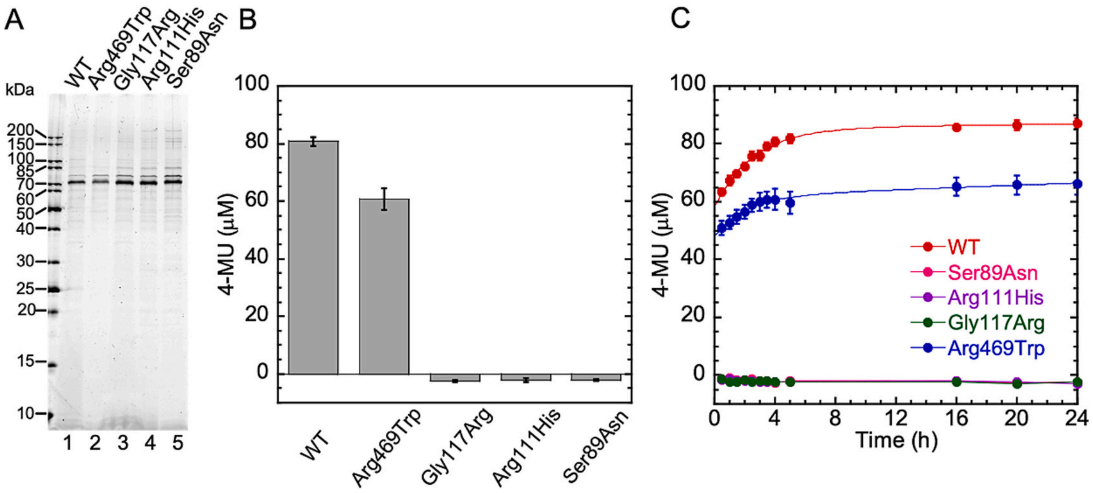

Fig. 4. Arylsulfatase activity of ARSL variant proteins. (A) SDS-PAGE analysis of purified WT and variants. (B) Production of 4-MU at 4 h after initiation of the enzyme reaction for WT and variants. (C) Time course of 4-MU production by WT and variants. Error bars represent the standard error; $n = 3$ (WT, Ser89Asn, Arg111His, Gly117Arg); $n = 8$ (Arg469Trp).

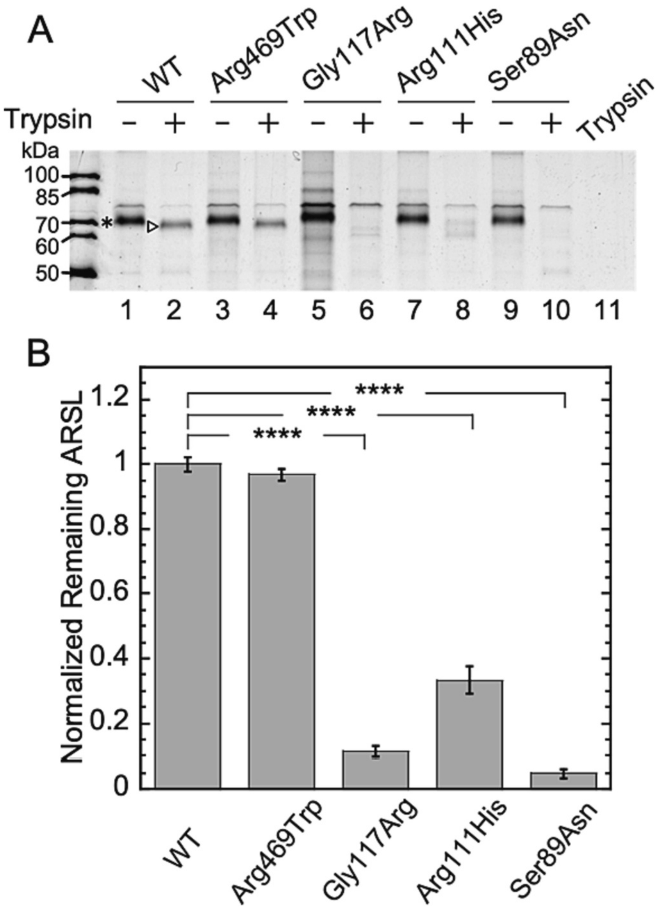

Fig. 5. Trypsin digestion of WT ARSL and variants. (A) SDS–PAGE analysis of ARSL proteins treated with trypsin. Lanes 1, 3, 5, 7 and 9 indicate ARSL proteins without trypsin. Lanes 2, 4, 6, 7 and 10 indicate ARSL proteins with trypsin. Lane 11 corresponds to a sample incubated with trypsin and without ARSL protein. Asterisk indicates full-length WT ARSL; arrowhead indicates the trypsin-resistant fragment of ARSL WT. (B) Proportion of trypsin-resistant fragment remaining after trypsin digestion. The remaining proportion was calculated as the band intensity of the trypsin digestion-resistant fragment to that of full-length ARSL in untreated samples; values have been normalized to WT. Error bars represent standard error of the mean ($n = 3$). Statistical comparisons were done by one-way ANOVA followed by Tukey HSD test. **** $p < 0.0001$.

3.5. Clinical manifestations of chondrodysplasia punctata for individuals carrying missense ARSL variants predicted as "damaging" or "possibly damaging" by VarMeter

As mentioned above, we identified four Ser89Asn, Arg111His, Gly117Arg and Arg469Trp variants of ARSL in the in-house whole-exome sequencing data of 900 patients with rare or undiagnosed diseases (Fig. 3, Table 3). The Ser89Asn and Arg469Trp ARSL variants have not been previously reported, while the Arg111His and Gly117Arg variants have been previously reported in patients with CDPX1 [8,9,37]. In this study, the two patients with the known Arg111His or Gly117Arg variant, respectively predicted as "damaging" or "possibly damaging" by VarMeter, showed clinical manifestations including hypoplasia of the anterior nasal spine, flattened nasal bridge, short stature and calcification in the ankle and distal phalanges in infancy, and had been clinically diagnosed with chondrodysplasia punctata.

The patient with the novel Ser89Asn variant of ARSL (Fig. 6A), also predicted as "damaging" by VarMeter, had facial features of chondrodysplasia punctata, such as depressed nasal bridge and short nasal septum, in addition to calcification in the ankle and paravertebral region (Fig. 6B), resulting in a diagnosis of chondrodysplasia punctata as well. The Ser89Asn variant was not registered in gnomAD, 1000 genome or the in-house database. The variant was confirmed in the patient, but not in his mother by Sanger sequencing (Fig. 6C), suggesting that Ser89Asn is a novel de novo variant of chondrodysplasia punctata. Lastly, the other patient harboring variant Arg469Trp, which was predicted as "benign", had no clinical manifestations of chondrodysplasia punctata, and was not diagnosed with this disease. Collectively, the clinical manifestations of patients with the ARSL variants also supported the validity of prediction by VarMeter.

4. Discussion

In this study, we established a new method for predicting deleterious missense variants based on Gibbs free energy and nSASA using previously reported variants of ARSL, ALG1, and MPI. We applied this method, named VarMeter, to the in-house whole exome sequencing data of 900 individuals in Japan with rare or undiagnosed disease, and found that three variants, predicted as "damaging" or "possibly damaging" by VarMeter, had reduced protein stability and lacked enzymatic activity (Table 3). Males with ARSL variants predicted as "damaging" or "possibly damaging" were found to exhibit clinical manifestations of chondrodysplasia punctata and were diagnosed with this disease. Lastly, VarMeter enabled us to identify a novel deleterious ARSL variant, Ser89Asn, which was functionally damaged and linked to clinical manifestations of chondrodysplasia punctata.

We found that a single amino acid substitution caused protein destabilization and resulted in loss of enzymatic activity in more cases than expected. Among known pathogenic variants of ARSL (25 variants), 18 had a mutation energy – corresponding to the difference in Gibbs free energy between WT and the variant – of >0.88 kcal/mol. This implies that loss or perturbation of ARSL activity is mainly due to protein instability caused by the variant. In contrast, benign variants of ARSL (7 variants) had smaller mutation energies, and only two showed a mutation energy of >0.88 kcal/mol. For our prediction method, we also considered nSASA at the missense variant; this parameter is also highly correlated to protein destabilization because variant of inner hydrophobic core will greatly affect protein stability. Among the 25 pathogenic variants, 19 showed an nSASA of <0.11, meaning that the substitution site is in the inner core of the protein. This also implies that loss or perturbation of ARSL activity is mainly due to protein instability caused by the variant. In contrast, the seven benign variants showed larger nSASA values and only two had an nSASA of less than or equal to 0.11.

Regarding the damaging variants of ARSL identified among individuals in the in-house database, Ser89 is located close to the substrate-binding site (Fig. 3) [39]; therefore, it is possible that the novel Ser-to-Asn substitution at residue 89 affects sulfatase activity due to the local conformational change, as well as protein instability. Gly117Arg showed a negative mutation energy (−1.1 kcal/mol) and was not predicted as a destabilizing substitution by mutation energy alone. However, the nSASA of Gly117 was 0, indicating that this residue is located inside the molecule (Fig. 3). Taken together, the Gly-to-Arg substitution at residue 117 is also predicted to induce instability of the protein, and in fact its protease resistance was greatly reduced (Fig. 5). Another possibility is that the side chain of the substituted Arg117 residue might not be able to interact with the membrane. Gly117 is located within or close to the transmembrane region (Fig. 3); thus Gly-to-Arg substitution at 117 may cause a steric clash with membrane.

We demonstrated that VarMeter can predict damaging variants of ARSL with an accuracy of 81%. There are many patterns of pathogenic variants that affect the biological activity and hence cause disease; for example, a protein may lose function if the substituted residue is located within the functional sites (e.g., protein–ligand interaction site, post-translational modification site, etc.) [40–42]. The prediction accuracy of our method indicates that damaging variants are also frequently

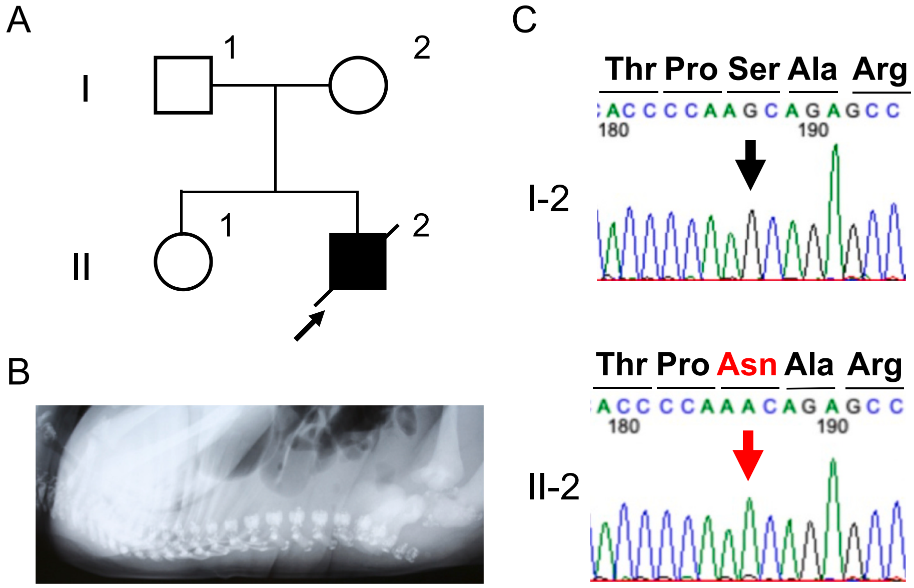

Fig. 6. A patient with chondrodysplasia punctata.
The patient was affected by polyhydramnios in utero, showed severe nasomaxillary hypoplasia on ultrasound, and died at 1 day of age due to severe thoracic hypoplasia and respiratory failure after birth. This patient had no pathogenic variants of other genes known to cause punctate calcifications, such as the *EBP* gene. (A) Family pedigree of the family. (B) X-ray lateral view of the spine in the patient. Diffuse stippling around the vertebras. (C) Electropherograms of Sanger sequencing in the patient and his mother, showing the substitution of G > A at c.266 [NM_000047.3:c.266G > A,p.(Ser89Asn)]. Red arrow indicates the substitution. Translated amino acids are indicated above the sequences.

related to protein stability. In other words, mutation energy and nSASA are sensitive parameters for predicting the severity of variants that induce protein instability. This is supported by the fact that three pathogenic variants (Arg111Pro, Gly245Arg, Thr306Ala) that our method assigned as damaging or possibly damaging were predicted as benign by other tools. In fact, the three variant proteins with these variants had all lost enzymatic activity [7,9].

So far, many approaches based on the 3D structure of mutant and WT proteins have been reported for predicting the severity of missense variants, such as missense3D [43]. Most methods rely on the calculation of Gibbs free energy and/or a combination of features affecting protein stability, such as hydrogen bond breakage. In terms of high-throughput screening and coupling with other methods, our simple and rapid method using minimum critical parameters is highly suitable for prediction. Furthermore, our method does not rely on a database of previously reported information on variants, which is the basis of SIFT, Polyphen-2 and CADD [34–36]. Regarding this point, our method is a good alternative to previously reported prediction tools and has an advantage for proteins with limited variant data for machine-learning. We are now considering a second version of the prediction tool to improve the prediction accuracy by analyzing other disease-related proteins.

The results of our predictive method, VarMeter, for ARSL variants, and clinical manifestations and clinical diagnosis of chondrodysplasia punctata were consistent among the four patients investigated in this study (Table 3). However, the results for the 32 previously reported variants of ARSL predicted by VarMeter in this study did not completely match their known severity (Table 1), and therefore the method cannot be said to be accurate for all variants. However, VarMeter may be an effective method for evaluating new variants in protein function, because it can accurately predict variants that other methods fail to assign adequately (Arg111Pro, Gly245Arg, Thr306Ala) (Table 1). Our results indicate that this new prediction method, based on a combination of mutation energy and nSASA, is supportive for the diagnosis of patients with rare or undiagnosed diseases, especially those with unknown missense variants.

5. Conclusion

Often the interpretation of variants, especially missense variants, can be problematic. Therefore, it is important to establish methods that allow the evaluation of missense variants with high accuracy. We have established and developed a new method for supporting diagnosis, VarMeter, which predicts missense variants leading to protein damage via Gibbs free energy calculation and nSASA of the corresponding 3D model protein. This method supported the diagnosis of a patient with a novel variant of ARSL (p.Ser89Asn) with chondrodysplasia punctata. In addition, two variants of ARSL previously reported as causative pathogenic variants of chondrodysplasia punctata, Arg111His and Gly117Arg, were predicted as damaging and possibly damaging by VarMeter. In principle, VarMeter could be applied to predict deleterious missense variants of other folded proteins. We propose that the prediction of a variant's severity from genomic data using VarMeter may be helpful to make diagnoses of rare or undiagnosed diseases.

Funding

This research was supported by AMED (Japan Agency for Medical Research and Development) under Grant Number JP22ek0109446.

CRediT authorship contribution statement

Eriko Aoki: Investigation, Writing – original draft, Visualization. **Noriyoshi Manabe:** Investigation, Methodology, Writing – original draft, Visualization. **Shiho Ohno:** Investigation, Formal analysis, Validation, Visualization. **Taiga Aoki:** Resources, Investigation. **Jun-Ichi Furukawa:** Investigation. **Akira Togayachi:** Methodology. **Kiyoko Aoki-Kinoshita:** Data curation. **Jin-Ichi Inokuchi:** Validation. **Kenji Kurosawa:** Resources. **Tadashi Kaname:** Resources, Writing – original draft, Writing – review & editing, Conceptualization, Project administration. **Yoshiki Yamaguchi:** Writing – original draft, Writing – review & editing, Conceptualization, Project administration. **Shoko Nishihara:** Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Project administration, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be available on reasonable request.

Acknowledgements

The authors thank Professor Kumiko Ui-Tei (The University of Tokyo) for providing the pCAGI-puro plasmid. They also thank Dr. Kazuyoshi Itoh (Soka University) for helpful comments on plasmid construction and cell culture.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymgmr.2023.101016.

References

[1] S. Nguengang Wakap, D.M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, D. Murphy, Y. Le Cam, A. Rath, Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database, Eur. J. Hum. Genet. 28 (2020) 165–173.

[2] A. Hamosh, J.S. Amberger, C. Bocchini, A.F. Scott, S.A. Rasmussen, Online Mendelian inheritance in man (OMIM(R)): victor McKusick's magnum opus, Am. J. Med. Genet. A 185 (2021) 3259–3265.

[3] Online Inheritance in Man: OMIM, 1966-2022, 2023.

[4] R.M. Kliegman, J.S. Geme, Nelson Textbook of Pediatrics, 21th ed., Elsevier, Philadelphia, Pennsylvania, 2020.

[5] P. Yue, Z. Li, J. Moult, Loss of protein structure stability as a major causative factor in monogenic disease, J. Mol. Biol. 353 (2005) 459–473.

[6] R.S. Holmes, Comparative and evolutionary studies of mammalian arylsulfatase and sterylsulfatase genes and proteins encoded on the X-chromosome, Comput. Biol. Chem. 68 (2017) 71–77.

[7] A. Daniele, G. Parenti, M. d'Addio, G. Andria, A. Ballabio, G. Meroni, Biochemical characterization of arylsulfatase E and functional analysis of mutations found in patients with X-linked chondrodysplasia punctata, Am. J. Hum. Genet. 62 (1998) 562–572.

[8] B. Franco, G. Meroni, G. Parenti, J. Levilliers, L. Bernard, M. Gebbia, L. Cox, P. Maroteaux, L. Sheffield, G.A. Rappold, G. Andria, C. Petit, A. Ballabio, A cluster of sulfatase genes on Xp22.3: mutations in chondrodysplasia punctata (CDPX) and implications for warfarin embryopathy, Cell 81 (1995) 15–25.

[9] C. Matos-Miranda, G. Nimmo, B. Williams, C. Tysoe, M. Owens, S. Bale, N. Braverman, A prospective study of brachytelephalangic chondrodysplasia punctata: identification of arylsulfatase E mutations, functional analysis of novel missense alleles, and determination of potential phenocopies, Genet. Med. 15 (2013) 650–657.

[10] N. Brunetti-Pierri, M.V. Andreucci, R. Tuzzi, G.R. Vega, G. Gray, C. McKeown, A. Ballabio, G. Andria, G. Meroni, G. Parenti, X-linked recessive chondrodysplasia punctata: spectrum of arylsulfatase E gene mutations and expanded clinical variability, Am. J. Med. Genet. A 117A (2003) 164–168.

[11] B.G. Ng, S.A. Shiryaev, D. Rymen, E.A. Eklund, K. Raymond, M. Kircher, J. E. Abdenur, F. Alehan, A.T. Midro, M.J. Bamshad, R. Barone, G.T. Berry, J. E. Brumbaugh, K.J. Buckingham, K. Clarkson, F.S. Cole, S. O'Connor, G.M. Cooper, R. Van Coster, L.A. Demmer, L. Diogo, A.J. Fay, C. Ficicioglu, A. Fiumara, W. A. Gahl, R. Ganetzky, H. Goel, L.A. Harshman, M. He, J. Jaeken, P.M. James, D. Katz, L. Keldermans, M. Kibaek, A.J. Kornberg, K. Lachlan, C. Lam, J. Yaplito-Lee, D.A. Nickerson, H.L. Peters, V. Race, L. Regal, J.S. Rush, S.L. Rutledge, J. Shendure, E. Souche, S.E. Sparks, P. Trapane, A. Sanchez-Valle, E. Vilain, A. Vollo, C.J. Waechter, R.Y. Wang, L.A. Wolfe, D.A. Wong, T. Wood, A.C. Yang, G. University of Washington Center for Mendelian, G. Matthijs, H.H. Freeze, ALG1-CDG: Clinical and Molecular Characterization of 39 Unreported Patients Hum Mutat 37, 2016, pp. 653–660.

[12] J. Jaeken, G. Matthijs, J.M. Saudubray, C. Dionisi-Vici, E. Bertini, P. de Lonlay, H. Henri, H. Carchon, E. Schollen, E. Van Schaftingen, Phosphomannose isomerase deficiency: a carbohydrate-deficient glycoprotein syndrome with hepatic-intestinal presentation, Am. J. Hum. Genet. 62 (1998) 1535–1539.

[13] D. Penel-Capelle, D. Dobbelaere, J. Jaeken, A. Klein, M. Cartigny, J. Weill, Congenital disorder of glycosylation Ib (CDG-Ib) without gastrointestinal symptoms, J. Inherit. Metab. Dis. 26 (2003) 83–85.

[14] E. Schollen, L. Dorland, T.J. de Koning, O.P. Van Diggelen, J.G. Huijmans, T. Marquardt, D. Babovic-Vuksanovic, M. Patterson, F. Imtiaz, B. Winchester, M. Adamowicz, E. Pronicka, H. Freeze, G. Matthijs, Genomic organization of the human phosphomannose isomerase (MPI) gene and mutation analysis in patients with congenital disorders of glycosylation type Ib (CDG-Ib), Hum. Mutat. 16 (2000) 247–252.

[15] V. Westphal, S. Kjaergaard, J.A. Davis, S.M. Peterson, F. Skovby, H.H. Freeze, Genetic and metabolic analysis of the first adult with congenital disorder of glycosylation type Ib: long-term outcome and effects of mannose supplementation, Mol. Genet. Metab. 73 (2001) 77–85.

[16] R. Niehues, M. Hasilik, G. Alton, C. Korner, M. Schiebe-Sukumar, H.G. Koch, K. P. Zimmer, R. Wu, E. Harms, K. Reiter, K. von Figura, H.H. Freeze, H.K. Harms, T. Marquardt, Carbohydrate-deficient glycoprotein syndrome type Ib. Phosphomannose isomerase deficiency and mannose therapy, J. Clin. Invest. 101 (1998) 1414–1420.

[17] P. de Lonlay, M. Cuer, S. Vuillaumier-Barrot, G. Beaune, P. Castelnau, M. Kretz, G. Durand, J.M. Saudubray, N. Seta, Hyperinsulinemic hypoglycemia as a presenting sign in phosphomannose isomerase deficiency: a new manifestation of carbohydrate-deficient glycoprotein syndrome treatable with mannose, J. Pediatr. 135 (1999) 379–383.

[18] S. Vuillaumier-Barrot, C. Le Bizec, P. de Lonlay, A. Barnier, G. Mitchell, V. Pelletier, C. Prevost, J.M. Saudubray, G. Durand, N. Seta, Protein losing enteropathy-hepatic fibrosis syndrome in Saguenay-lac St-Jean, Quebec is a congenital disorder of glycosylation type Ib, J. Med. Genet. 39 (2002) 849–851.

[19] D. Babovic-Vuksanovic, M.C. Patterson, W.F. Schwenk, J.F. O'Brien, J. Vockley, H. H. Freeze, D.P. Mehta, V.V. Michels, Severe hypoglycemia as a presenting symptom of carbohydrate-deficient glycoprotein syndrome, J. Pediatr. 135 (1999) 775–781.

[20] K. Yanagi, N. Morimoto, M. Iso, Y. Abe, K. Okamura, T. Nakamura, Y. Matsubara, T. Kaname, A novel missense variant of the GNAI3 gene and recognisable morphological characteristics of the mandibula in ARCND1, J. Hum. Genet. 66 (2021) 1029–1034.

[21] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589.

[22] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[23] B.R. Brooks, C.L. Brooks 3rd, A.D. Mackerell Jr., L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.

[24] D. Bashford, D.A. Case, Generalized born models of macromolecular solvation effects, Annu. Rev. Phys. Chem. 51 (2000) 129–152.

[25] A.V. Onufriev, D.A. Case, Generalized born implicit solvent models for biomolecules, Annu. Rev. Biophys. 48 (2019) 275–296.

[26] S. Miller, J. Janin, A.M. Lesk, C. Chothia, Interior and surface of monomeric proteins, J. Mol. Biol. 196 (1987) 641–656.

[27] S.S. Shapiro, M.B. Wilk, An analysis of variance test for normality (complete samples), Biometrika 52 (1965) 591–611.

[28] T.W. Anderson, D.A. Darling, Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes The Annals of Mathematical Statistics 23, 1952, pp. 193–212.

[29] M. Greiner, D. Pfeiffer, R.D. Smith, Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests, Prev. Vet. Med. 45 (2000) 23–41.

[30] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis, Nat. Methods 9 (2012) 671–675.

[31] L. Zhang, H. Hu, D. Liang, Z. Li, L. Wu, Prenatal diagnosis in a fetus with X-linked recessive chondrodysplasia punctata: identification and functional study of a novel missense mutation in ARSE, Front. Genet. 12 (2021), 722694.

[32] V.Z. Spassov, L. Yan, A pH-dependent computational approach to the effect of mutations on protein stability, J. Comput. Chem. 37 (2016) 2573–2587.

[33] S. Iqbal, E. Pérez-Palma, J.B. Jespersen, P. May, D. Hoksza, H.O. Heyne, S. S. Ahmed, Z.T. Rifat, M.S. Rahman, K. Lage, A. Palotie, J.R. Cottrell, F.F. Wagner, M.J. Daly, A.J. Campbell, D. Lal, Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants, Proc. Natl. Acad. Sci. 117 (2020) 28201–28211.

[34] P.C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions, Genome Res. 11 (2001) 863–874.

[35] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, Nat. Methods 7 (2010) 248–249.

[36] P. Rentzsch, M. Schubach, J. Shendure, M. Kircher, CADD-splice-improving genome-wide variant effect prediction using deep learning-derived splice scores, Genome Med. 13 (2021) 31.

[37] S. Meyer, G. Loffler, M. Gencik, P. Fries, P. Papanagiotou, B. Oehl-Jaschkowitz, L. Gortner, Brachytelephalangic chondrodysplasia punctata with a new hemizygous missense mutation in a neonate, Am. J. Med. Genet. A 161A (2013) 626–629.

[38] D. Vestweber, G. Schatz, Point mutations destabilizing a precursor protein enhance its post-translational import into mitochondria, EMBO J. 7 (1988) 1147–1151.

[39] A. Waldow, B. Schmidt, T. Dierks, R. von Bulow, K. von Figura, Amino acid residues forming the active site of arylsulfatase A. Role in catalytic activity and substrate binding, J. Biol. Chem. 274 (1999) 12284–12288.

[40] M. Nemethova, J. Radvanszky, L. Kadasi, D.B. Ascher, D.E. Pires, T.L. Blundell, B. Porfirio, A. Mannoni, A. Santucci, L. Milucci, S. Sestini, G. Biolcati, F. Sorge, C. Aurizi, R. Aquaron, M. Alsbou, C.M. Lourenco, K. Ramadevi, L.R. Ranganath, J. A. Gallagher, C. van Kan, A.K. Hall, B. Olsson, N. Sireau, H. Ayoob, O.G. Timmis, K. H. Sang, F. Genovese, R. Imrich, J. Rovensky, R. Srinivasaraghavan, S. K. Bharadwaj, R. Spiegel, A. Zatkova, Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy, Eur. J. Hum. Genet. 24 (2016) 66–72.

[41] P. Kim, J. Zhao, P. Lu, Z., Zhao, mutLBSgeneDB: mutated ligand binding site gene DataBase, Nucleic Acids Res. 45 (2017) D256–D263.

[42] M. Krassowski, M. Paczkowska, K. Cullion, T. Huang, I. Dzneladze, B.F.F. Ouellette, J.T. Yamada, A. Fradet-Turcotte, J. Reimand, ActiveDriverDB: human disease mutations and genome variation in post-translational modification sites of proteins, Nucleic Acids Res. 46 (2018) D901–D910.

[43] S. Ittisoponpisan, S.A. Islam, T. Khanna, E. Alhuzimi, A. David, M.J.E. Sternberg, Can predicted protein 3D structures provide reliable insights into whether missense variants are disease associated? J. Mol. Biol. 431 (2019) 2197–2212.