European Review for Medical and Pharmacological Sciences 2023; 27: 137-147

AKR1C1 and hormone metabolism in lipedema pathogenesis: a computational biology approach

J. KAFTALLI[1], G. BONETTI[2,3], G. MARCEDDU[1], K. DHULI[2], P.E. MALTESE[2], K. DONATO[1,4], K.L. HERBST[5], S. MICHELINI[6], P. CHIURAZZI[7,8], M. HILL[9], S. MICHELINI[10], S. MICHELINI[11], A. BERNINI[12], M. BERTELLI[1,2,4]

[1]MAGI EUREGIO, Bolzano, Italy
[2]MAGI LAB, Rovereto, Trento, Italy
[3]Department of Pharmaceutical Sciences, University of Perugia, Perugia, Italy
[4]MAGISNAT, Atlanta Tech Park, Technology Parkway, Peachtree Corners, GA, USA
[5]Total Lipedema Care, Beverly Hills, California and Tucson, Arizona, USA
[6]Vascular Diagnostics and Rehabilitation Service, Marino Hospital, ASL Roma 6, Marino, Rome, Italy
[7]UOC Genetica Medica, Fondazione Policlinico Universitario "A. Gemelli" IRCCS, Rome, Italy
[8]Istituto di Medicina Genomica, Università Cattolica del S. Cuore, Rome, Italy
[9]Department of Steroids and Proteofactors, Institute of Endocrinology, Prague, Czech Republic
[10]Unit of Physical Medicine, "Sapienza" University of Rome, Rome, Italy
[11]Neurosurgery, University of Tor Vergata, Rome, Italy
[12]Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena, Italy

J. Kaftalli and G. Bonetti contributed equally to this work

Abstract. – OBJECTIVE: Lipedema is an autosomal dominant genetic disease that mainly affects women. It is characterized by excess deposition of subcutaneous adipose tissue, pain, and anxiety. The genetic and environmental etiology of lipedema is still largely unknown. Although considered a rare disease, this pathology has been suggested to be underdiagnosed or misdiagnosed as obesity or lymphedema. Steroid hormones seem to be involved in the pathogenesis of lipedema. Indeed, aldo-keto reductase family 1 member C1 (*AKR1C1*), a gene coding for a protein involved in steroid hormones metabolism, was the first proposed to be correlated with lipedema.

PATIENTS AND METHODS: In this study, we employed a molecular dynamics approach to assess the pathogenicity of *AKR1C1* genetic variants found in patients with lipedema. Moreover, we combined information theory and structural bioinformatics to identify *AKR1C1* polymorphisms from the gnomAD database that could predispose to the development of lipedema.

RESULTS: Three genetic variants in *AKR1C1* found in patients with lipedema were disruptive to the protein's function. Furthermore, eight *AKR1C1* variants found in the general population could predispose to the development of lipedema.

CONCLUSIONS: The results of this study provide evidence that *AKR1C1* may be a key gene in lipedema pathogenesis, and that common polymorphisms could predispose to lipedema development.

Key Words:
Lipedema, Adipose tissue, *AKR1C1*, Molecular dynamics, Information theory, Genetic variants.

Introduction

Lipedema is an autosomal dominant genetic disease with differential sex-penetrance, with a prevalence of almost 10/100,000[1]. However, since it may be misdiagnosed with obesity or lymphedema[2-4], its true prevalence could be much higher. Lipedema is proposed to be caused by defects in the hormonal axes[5]. Indeed, its symptoms, such as subcutaneous fat accumulation, pain, and anxiety, could be caused by hormonal dysregulation, especially in the estrogen axis[1,6,7]. Estrogens seem to be of utmost importance in the pathogenesis of lipedema since women are more affected than males, and the symptoms often worsen with puberty[5]. Therapies of lipedema are essentially targeted on symptoms and include manual lymphatic drainage therapy, intermittent pneumatic compression pumps, and, in recalcitrant cases, liposuction[1,8].

The genetic basis of lipedema largely remains unknown: although several genes are involved in fat tissue deposition or estrogen metabolism, none of them has been clearly correlated with lipedema pathogenesis[1,9]. The first gene proposed to be

Corresponding Author: Jurgen Kaftalli, BSc; e-mail: jurgen.kaftalli@assomagi.org

correlated with lipedema is aldo-keto reductase family 1 member C1 (*AKR1C1*), an aldo-keto reductase involved in steroid hormone metabolism in adipose tissue[5]. Hydroxysteroid dehydrogenases, like *AKR1C1* and aldo-keto reductase family 1 member C2 (*AKR1C2*), regulate the synthesis and inactivation of steroid hormones in various tissue, like the testis, prostate, liver, and adipose tissues. *AKR1C1* and *AKR1C2* participate in the inactivation of progesterone and dihydrotestosterone with different mechanisms[10,11], and the inactivation of *AKR1C* enzymes increases adipogenesis and fat mass, especially in subcutaneous tissues[10]. A comprehensive review showing the role of *AKR1C1* in steroid hormones metabolism can be found in Kiani et al[10]. Indeed, *AKR1C* enzymes are of utmost importance in subcutaneous adipose tissue, where they participate in the androgens-glucocorticoid axis, regulating adipogenesis[10].

The aim of this paper is twofold: evaluate the effects of *AKR1C1* genetic variants found in patients diagnosed with lipedema and identify *AKR1C1* genetic variants in the general population that could have a detrimental effect on *AKR1C1* function, which, especially in the presence of specific environmental stimuli, could predispose to the development of lipedema. These variants can be useful in shedding light on the pathology and contribute to identifying the prevalence of lipedema in various populations.

An information theory approach to the sequence evolution allowed for critical positions to be defined. Protein multiple sequence alignments (MSA) are widely used to infer the conservation of amino acid residues within an evolutionarily related family[12]. Highly conserved residues tend to correlate with structural or functional importance, and accurate identification of such important residues aids in the experimental characterization of protein function. Multiple sequence alignments of protein sequences produce a matrix of amino acids; by elaborating the columns as vectors, entropy can be calculated according to Shannon[13], describing the amount of variability through a column in the alignment. The lower the value, the lower the variability accepted by the sequence position, implying high conservation throughout evolution. It has been shown that entropy scores that do not incorporate background amino acid frequencies are not theoretically optimal for calculating residue conservation[13], and the occurrence of each amino acid should be weighted by the preference of evolution towards it, rewarding rare amino acids for their presence. By looking at the original BLOSUM matrices and re-interpreting them as mutation probability matrices, it is possible to derive a distribution of background frequencies[14]. Thus, as suggested by Gao et al[15], Shannon Entropy was corrected by using BLOSUM62 matrix as background distribution (**Supplementary Table I**) and reported as Shannon Entropy/BLOSUM62. Moreover, BLOSUM62 was also used as the background distribution for computing the relative entropy[16,17]. In statistics, relative entropy (Kullback-Leibler divergence) is used to quantify the difference between two probability distributions. In the case of residue conservation, a high deviation from the background indicates stronger evolutionary constraints and a potentially functional position[16].

Patients and Methods

Patients Recruitment

We analyzed Italian and US patients with a clinical diagnosis of lipedema, according to Wold et al[18] criteria. All patients received genetic counseling to explain the risks and benefits of genetic testing. The patients were informed about the significance of genetic testing. All of them gave their written informed consent, in compliance with the Helsinki Declaration. Ethical approval and clearance were received from the Ethical Committee of Azienda Sanitaria dell'Alto Adige, Italy (Approval No. 132-2020). Blood or saliva samples were collected from each subject, and genomic DNA was extracted using a commercial kit (SAMAG BLOOD DNA Extraction Kit (Beckman, Milan, MI, Italy) according to the manufacturer's instruction, as previously reported[19,20].

Panel Design, Library Preparation and Sequencing

We developed a custom NGS-based panel of 305 genes, as described by Michelini et al[21]. The genes included in the panel were associated with lipedema and subcutaneous fat tissue accumulation on the basis of data reported in the Human Gene Mutation Database (HGMD Professional)[22], Online Mendelian Inheritance in Man (OMIM)[23], Orphanet[24], GeneReviews[25], and PubMed[26]. The custom DNA probes were designed using Twist Bioscience technology (available at: https://www.twistbioscience.com/, last accessed on 22 December 2021). Illumina NGS sequencing was carried

out by an IntegraGen Genomic Service (available at: www.integragen-genomics.com, last accessed on 22 December 2021). The NGS panel included genomic targets comprising coding exons and 15 bp flanking regions of each gene. The cumulative target length of the gene panel was 839,308 bp. Library preparation from genomic DNA samples and target enrichment was performed according to the manufacturer's protocol using the Twist Library Preparation with Enzymatic Fragmentation with Twist Universal Adapter (UDI) System and Twist Target Enrichment (Twist Bioscience) (TwistBioscience, South San Francisco, CA, USA). A MiSeq personal sequencer (Illumina, San Diego, CA, USA) was used to perform 150 bp paired-end read sequencing following the manufacturer's instructions.

Bioinformatics

Fastq (forward-reverse) files were obtained after sequencing. Bioinformatic analysis was performed as previously described[27,28]. Briefly, the sequencing reads were mapped to the reference genome (hg38/GRCh38) using Burrow-Wheeler Aligner (version 0.7.17-r1188; available at: https://github.com/lh3/bwa) software. Duplicates were removed using SAMBAMBA (version 0.6.7; available at: https://github.com/biod/sambamba) and MarkDuplicates GATK (version 4.0.0.0; https://gatk.broadinstitute.org/hc/en-us/articles/360037052812-MarkDuplicates-Picard-). The BAM alignment files generated were refined by local realignment and base quality score recalibration, using the RealignerTargetCreator (available at: https://github.com/broadinstitute/gatk-docs/blob/master/gatk3-tutorials/(howto)_Perform_local_realignment_around_indels.md) and IndelRealigner GATK tools (available at: https://github.com/broadinstitute/gatk-docs/blob/master/gatk3-tutorials/(howto)_Perform_local_realignment_around_indels.md).

Statistical Analysis

We aim to investigate conformational changes in a protein due to specific mutations, therefore there is no output that can be analyzed statistically.

Multiple Sequence Alignment and Conservation

We aligned 120 sequences from the AKR1C family (**Supplementary File 1**: Alignment File) to derive Shannon entropy[29] of each position with gaps occupying less than 50% of the sequences - referred in this paper as well represented positions.

The Shannon entropy score is computed as:

$$H = - \sum_i^{n_{aa}} p_i \log(p_i)$$

where p_i is the frequency of residue i in the aligned column, $\sum$

$$p_i = count(i) / N$$

and n_{aa} is the total number of residue types in the alignment. Highly conserved columns (positions) will have low variability and thus a low Shannon entropy score. The maximum value of 4.32 is reached when no conservation is present, while the minimum value of 0 is reached when there is no variability throghout the column.

As suggested by Gao et al[15] we define the new frequencies p^b, which account for the background distribution as:

$$p_i^b = (p_i / q_i) / (\sum_{(m=1)}^{20} p_m / q_m)$$

where q is the BLOSUM62 amino acid background distribution. Dividing by q, the frequencies of the disfavored amino acids are scaled up against the favored ones. Both Shannon entropy scores were normalized (division by the respective maximal values).

Using the same abbreviations, relative entropy was computed using the following formula:

$$S = \sum_i^{n_{aa}} p_i \log(p_i / q_i)$$

The background frequencies were retrieved from the BLOSUM62 background frequencies provided in Gao et al[15].

Unlike Shannon entropy, relative entropy is not inversely porportional to conservation. The higher the conservation, the higher the relative entropy. Maximum score is 7.16 when using BLOSUM62 as background, and the minimum score is 0.

Molecular Modeling and Molecular Dynamics Simulations

Molecular dynamic simulations of wildtype and mutant human *AKR1C1* in ternary complex with Nicotinammide adenina dinucleotide fosfato (NADP+) and 20α-hydroxyprogesterone (hPGS) were performed using the structure of Protein Data Bank (PDB, available at: https://www.rcsb.org/) entry 1MRQ[30] as the starting structure. In silico mutagenesis of each variant was performed using the highest-ranked rotamer provided by the

"mutagenesis" function of PyMOL v2.4.1 (Schrodinger LLC, NY, USA; available at: https://pymol.org/2/). At first, the mCSM web server (available at: https://biosig.lab.uq.edu.au/mcsm_ppi2/) was used to obtain the predicted Stability Changes (ΔΔG) of the variant, which evaluates the possible detrimental effect of aminoacidic mutations on protein folding[31]. Then, all molecular dynamics simulations were run using Gromacs v.2019.3 package (available at: https://manual.gromacs.org/current/release-notes/2019/2019.3.html)[32], with CHARMM36 as the force field, while parametrization of the ligands was obtained using CgenFF(available at: https://cgenff.silcsbio.com/)[33]. For each of the *AKR1C1* variants, the protein molecule was placed in a triclinic box with a minimum spacing of 1.2 nm on each side. The system was then solvated using TIP3P water molecules, neutralized with Na^+/Cl^-, and energy minimized *via* the gradient descent algorithm. The minimized systems were subjected to two subsequential equilibration steps of position-restrained molecular dynamics in the NVT and NPT ensembles, 100 ps each. Reference temperature and pressure of 300 K and 1 bar were imposed, respectively. Finally, a molecular dynamics production run was performed for 1,000 ns (1 μs) with a 2-fs integration step. The protein-hPGS and protein-NADP+ binding energies were calculated by using the MM/PBSA method implemented in GROMACS *via* the g_mmpbsa tool (available at: https://rashmikumari.github.io/g_mmpbsa/)[34].

Results

Identification of a Missense AKR1C1 Variant in Patients with Lipedema

We analyzed 106 Italian and US patients with a clinical diagnosis of lipedema. Among them, we identified 3 patients with missense genetic variants in the *AKR1C1* gene (**Supplementary Table II**). The effects of the identified genetic variants were studied with a molecular dynamics approach.

AKR1C1 Genetic Variants and Entropies

The resulting Shannon entropy/BLOSUM62 and relative entropy for the three variants found in patients with lipedema are reported in Table

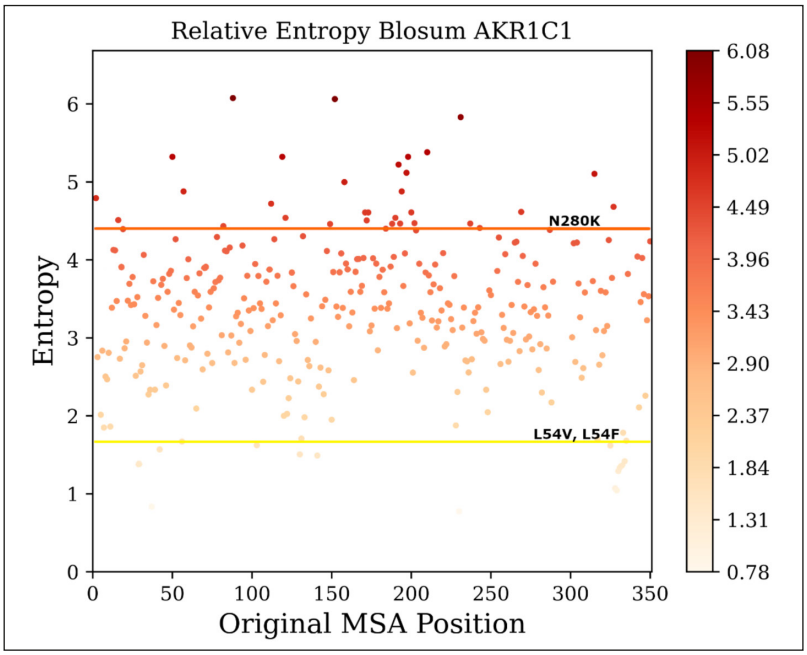

Figure 1. Relative Entropy heatmap of the multiple sequence alignments (MSA) positions. The scores of the positions of interest are highlighted with colored horizontal lines.

Table I. Entropies at the positions of identified variants. In the "Shannon Entropy BLOSUM62" both non-normalized and normalized values are shown. Multiple sequence alignments (MSA) positions differ from the Residue number because they account for gaps.

Residue Number	MSA position	Variants	Shannon Entropy BLOSUM62	Relative Entropy
54	56	L54F, L54V	2.52 / 0.6	1.67
280	287	N280K	0.15 / 0.03	4.38

I, while the entropies of all positions are reported in **Supplementary Table III**. Variants were found at positions 54 and 280. The latter, occupied by asparagine in the wildtype human *AKR1C1*, is the most conserved, with Shannon entropy/BLOSUM62 scoring very low and relative entropy scoring among the highest (Figure 1). On the other hand, for position 54, the Shannon entropy/BLOSUM62 score was way above 0, while the relative entropy score was among the lowest. This is reflected in various *AKR1C1* homologs, showing leucine 54 replaced by valine. Although position 54 scored poorly in terms of individual amino acid conservation, the physicochemical properties were preserved among the amino acids, and further structural studies provided insight into the importance of position 54 in substrate specificity.

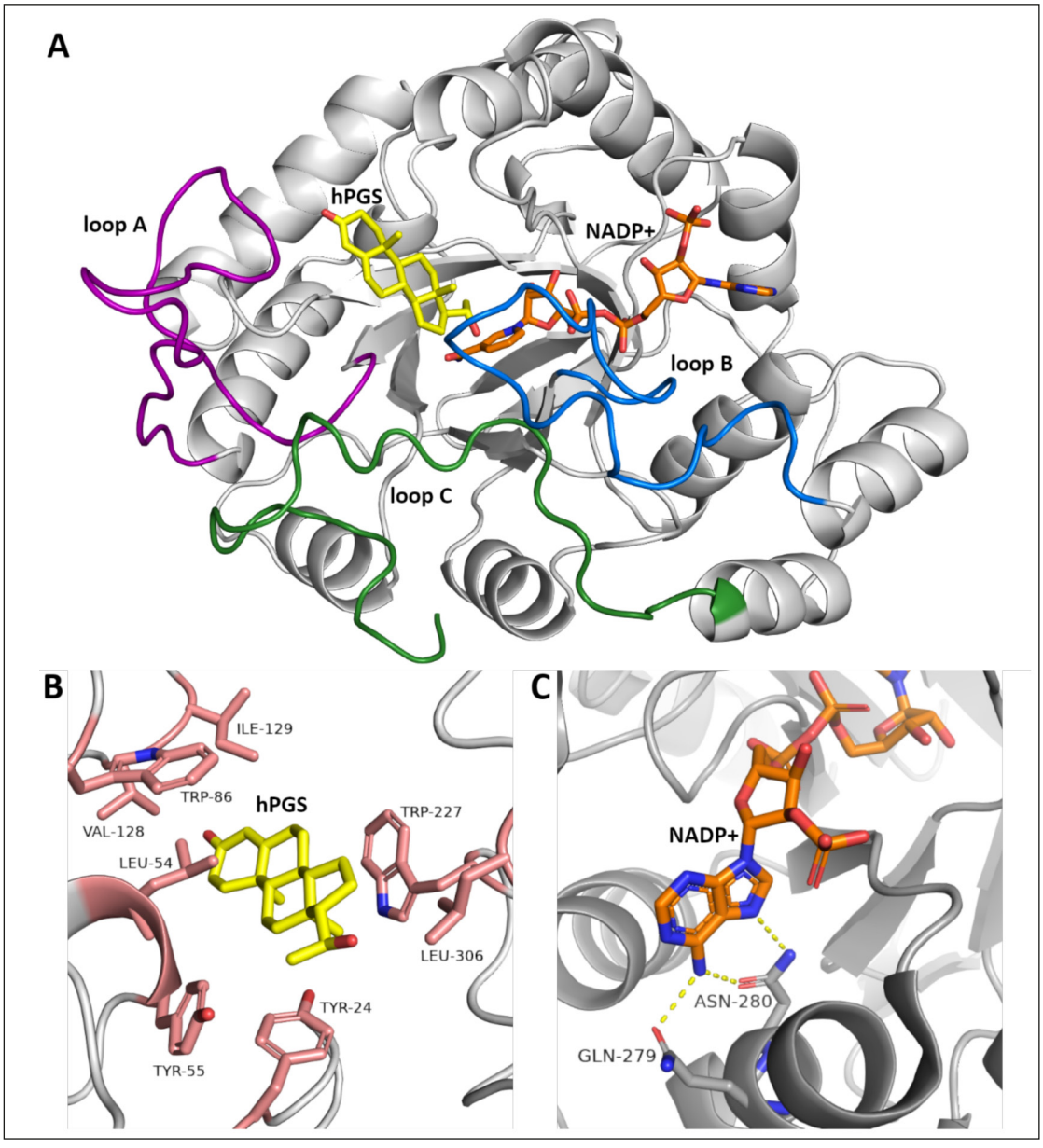

Figure 2. **A**, Ribbon representation of the structure of *AKR1C1* with ligands; NAPD+ and hPGS are colored in orange and yellow, respectively. The large loops contributing to the flexible part of the binding sites are also highlighted in colors, with loop A, loop B, and loop C colored in purple, blue, and green, respectively. **B**, Residues present in the steroid-binding pocket are shown as sticks colored in pink, with the steroid shown in yellow. **C**, The cofactor binding pocket and the residues involved in hydrogen bonding with the adenine moiety of NADP(H). Gln278 and Asn280 are shown as sticks; NADP(H) is shown in orange; hydrogen bonds are shown as dashed yellow lines.

Structural Analysis of Wild Type AKR1C1

Human *AKR1C1* three-dimensional structure shows an (αβ)8-barrel motif. Two more β-sheets, B1 (7-9) and B2 (15-17), and two more α-helices, H1(239-248) and H2 (290-298), not taking part in the core barrel structure, are found. Three large loops complete the structure: loop A is located at 117-143, loop B is located at 217-238, and loop C is located at 299-322 (Figure 2A).

Residues involved in substrate binding are Tyr24, Leu54, Phe118, Phe129, Thr226, Trp227, Asn306, and Tyr310 (Figure 2B), while those involved in catalysis are: Asp50, Tyr55, Lys84, and His117[35]. Regarding cofactor binding, the NADP(H)-binding residues are highly conserved and include Thr23, Asp50, Ser166, Asn167, Gln190, Tyr216, Leu219, Ser221, Arg270, Ser271, Phe272, Arg276, Glu279 and Asn280; the latter two amino acids contribute toward the binding affinity and specificity of the adenine moiety (Figure 2C).

Energy Landscape of Binding

To describe the interaction of the enzyme with cofactor and substrate in energetic terms, molecular dynamics simulations were carried out on the *AKR1C1*/hPGS/NADP+ ternary complex, and binding energy was calculated as well by using the MM/PBSA36 method and GROMACS molecular dynamic software[32]. The overall energies of binding for the two are reported in Table II.

The MM/PBSA method also allowed for quantification of the contribution to the binding of each amino acid, allowing the impact of an amino acidic missense substitution to be evaluated as follows. MM/PBSA profile of steroid (hPGS) binding shows that Tyr24, Leu54, and Trp227 account for 50% of the binding energy, with other significant contributions by Asp50, Tyr55, Trp86, Val128, Ile129, Leu306, showing an overall hydrophobic nature of the binding (Figure 3A).

MM/PBSA profile of NADP+ binding is dominated by charge pairs, giving prominent repulsion/attraction peaks between charged amino acids of the protein and the phosphate groups of the cofactor. The four most prominent negative binding energy

Table II. Overall binding energies of hPGS and NADP+ were calculated using MM/PBSA (molecular mechanics energies combined with the Poisson–Boltzmann or generalized Born and surface area continuum solvation) and GROMACS molecular dynamics software.

Ligand	Binding Energy (MM/PBSA)	Error Estimate
hPGS	-115.4 kJ/mol	+/- 10.6 kJ/mol
NADP+	-337.6 kJ/mol	+/- 53.9 kJ/mol

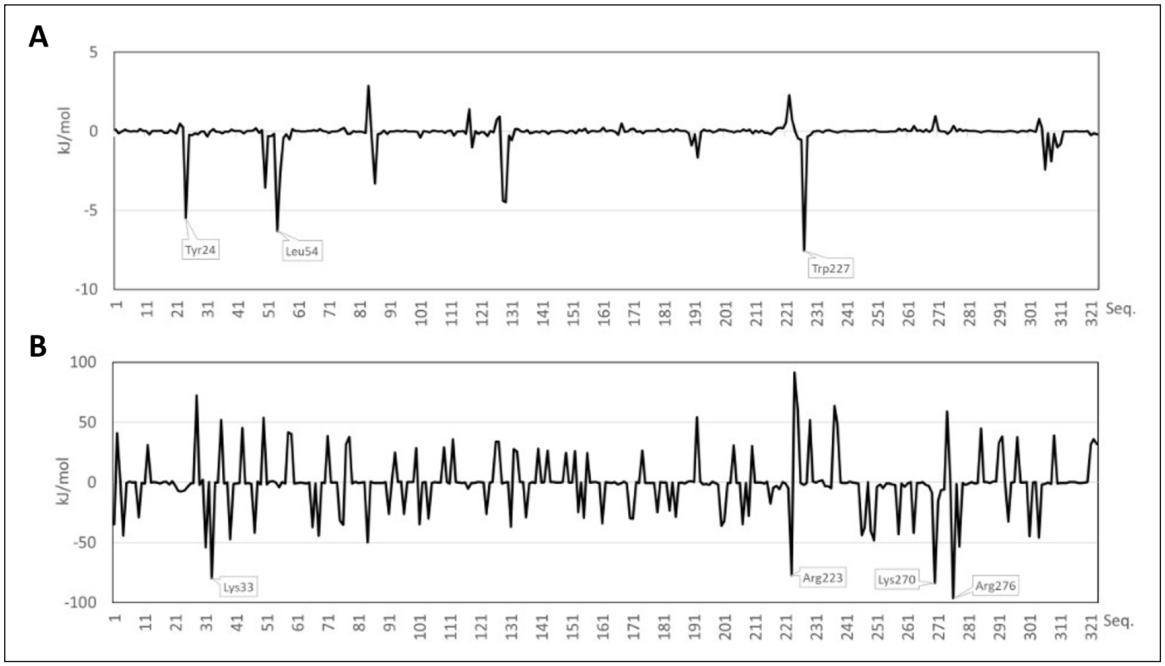

Figure 3. **A**, Average contribution of the residues to the protein-steroid binding energy in the wild type. **B**, Average contribution of the residues to the protein-cofactor binding energy in the wild type. The y-axis represents the contribution to the interaction energy, while the x-axis represents the protein sequence. Low energies mean favorable interaction, identifying possible binding hotspots.

peaks derive from Lys33, His222+Arg223, Lys270, and Arg276 (Figure 3B), all neighboring the phosphate group on the 2nd position of the ribose ring that carries the adenine moiety (Figure 2C). Such evidence accounts for the significant difference in affinity for NADP(H) *vs.* NAD(H) cofactors in binding AKR enzymes, with the former showing a mid-nanomolar value (100 nM), whereas the latter binds with mid-micromolar affinity (200 mM)[11].

Description of Structural Consequences of Variants Found in Lipedema Families

In the studied patients, three missense genetic variants were found, leading to Leu54Val, Leu54Phe and Asn280Lys. To evaluate the involvement of these variants in lipedema development, their effects on enzyme folding, stability, and biological activity were studied using a computational approach.

Variants Leu54Val and Leu54Phe

The role of Leu54 in substrate selectivity has been previously elucidated[11,37,38], and can be summarized as follows. Human *AKR1C1* and *AKR1C2* differ in that *AKR1C1* exhibits 20α-HSD activity, whereas *AKR1C2* exhibits 3α-HSD. The two enzymes differ for seven amino acids, and only one is located at the active site at position 54: leucine for *AKR1C1* and valine for *AKR1C2*[39]. The replacement of Leu54 by the less bulky valine (Leu-54Val) changes the 20α activity to 3α, disrupting the default *AKR1C1* function of reducing progesterone to 20α-hydroxyprogesterone. Meanwhile, according to Zhang et al[40], the Val54Leu variant in *AKR1C2* changes its 3α activity to 20α (Figure 4). Evidence that these enzymes act as reductases in mammalian cells was already given by Penning et al[11] and Zhang et al[40]. Similarly, the interaction between the steroid and the enzyme is disrupted by the replacement of the same Leu54 by phenylalanine, as shown by molecular modeling and dynamics. In the wild type, Leu54 and Trp227 play a significant role in binding the steroid by interacting with opposite faces of the polycyclic ring of the ligand and contributing as much as 33% of the overall binding energy. Mutation of Leu54 to Phe, although enhancing the hydrophobic nature of the interaction, introduces a second large, aromatic sidechain in place, hampering the ligand entrance in the site and conducive to binding disruption (Figure 5). Indeed, from the molecular dynamics simulations (**Video S1**), we observed Phe54 (rather than Trp227) to be pushed away from the steroid with a consequent increase in binding instability. Interestingly, phenylalanine is present at position 54 of the wild type, non-human *AKR1C8P*, but here, the steric hindrance with the opposite amino acid in position 227 is compensated by the replacement of tryptophan with the smaller asparagine. At the same time, the cumbersome tryptophan is 'shifted' to position 228. Nonetheless, 1C8 preserves the same 20α-HSD activity of *AKR1C1*.

Variant Asn280Lys

Asparagine 280 takes part in cofactor binding; together with Gln279, it is responsible for adenine moiety binding through a hydrogen bond to the amine group (Figure 2C). The molecular simulation showed how asparagine is the stronger binder of the two. Such finding is also confirmed by the molecular mechanics' energy contributions to the cofactor binding resulting from MM/PBSA, showing 6-fold higher interaction energy for Asn280 with respect to Gln279 (-17 kJ/mol *vs.* -3 kJ/mol).

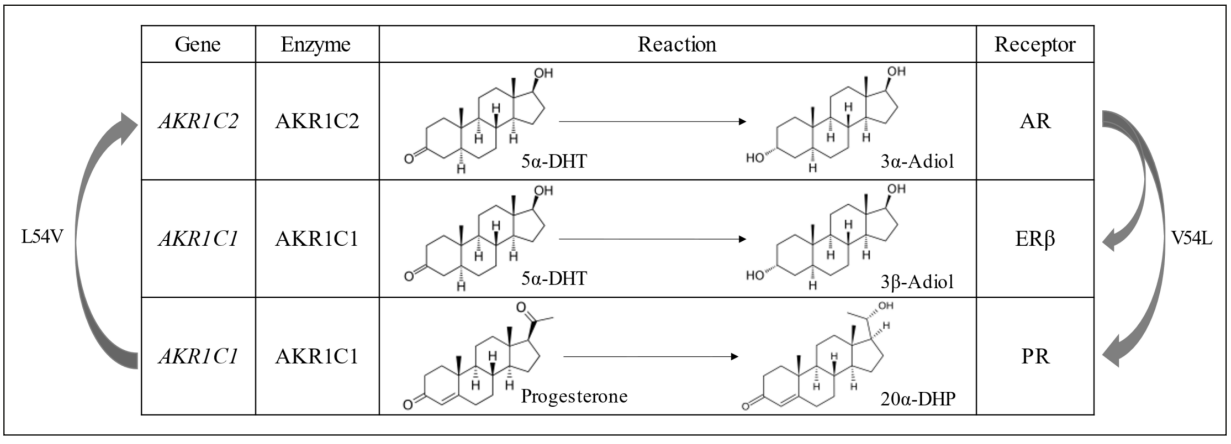

Figure 4. Metabolic role of *AKR1C1* and *AKR1C2*, with the L54V variant turning *AKR1C1* activity into that of *AKR1C2* as reported in Bertelli et al[38], and vice-versa.

Although this variant involved the replacement of a small sidechain with a bulky one, molecular modeling showed how the displacement of water molecules can easily accommodate the hydrophobic moiety of lysine. MD simulation analysis confirmed that a small effect is exerted on the protein structure, while the missing H-bond acceptor capability of lysine led to the loss of interaction with the adenine ring, resulting in the aromatic ring flipping away from its position, also because of the attraction of Lys280 to the phosphate group (**Video S2**). The optimal binding geometry is then disrupted rather than the folding.

Evaluation of AKR1C1 Variants in the General Population

The computational approaches used in this paper provided us with the ability to evaluate the effects of genetic *AKR1C1* variants present in the general population. We built a selection criterion based on two conditions:

1. if the residue was one of the binding hotspots residues, as evaluated by the contribution binding energies of each residue (Figure 3);
2. if the relative entropy of the position was greater than 4 OR the predicted ΔΔG of the variant was less than -3.

If both conditions 1 and 2 were satisfied, the variant was predicted as harmful to the protein and was further evaluated by molecular analyses. Among 265 *AKR1C1* variants (**Supplementary Table IV**) present in the gnomAD database (available at: https://gnomad.broadinstitute.org/)[41], we were able to select 8 of them (Table III), which, according to the criteria, were predicted to disrupt protein function. For each variant that meets the criteria, there are three types of disruption based on the specific residue that is mutated. The variant is said to disrupt binding, catalysis, or folding if the affected residue is involved in substrate binding, catalysis, or neither, respectively.

Discussion

The genetic etiology of lipedema is still largely unknown, and although it is considered a rare disease, it has been suggested[2-4] that this pathology is underdiagnosed or misdiagnosed, being confused with obesity or lymphedema, and therefore may be more common than that of a single rare gene mutation. Thus, finding the genetic causes of lipedema, and identifying how genetics and environmental factors predispose to this pathology is of utmost importance to improve the diagnosis and treatment of women living with lipedema.

AKR1C1, coding for a protein involved in steroid hormones metabolism, was the first gene proposed to be correlated to the clinical diagnosis of lipedema. Indeed, steroid hormones seem to be

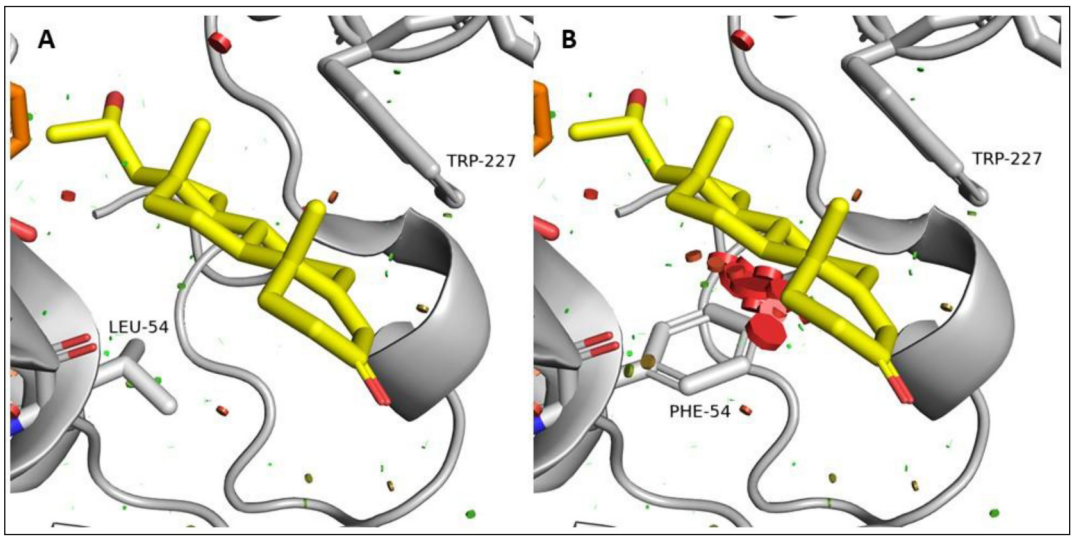

Figure 5. L54F steric hindrance, with the steroid shown in yellow, and protein residues shown in light gray. The mutagenesis of leucine to phenylalanine at position 54 generates clashes with the substrate (shown in red polygons), indicating the steric hindrance phenylalanine introduces to the steroid binding cavity. **A**, Structure of wildtype *AKR1C1* binding site with leucine at position 54. **B**, Structure of mutant *AKR1C1* binding site with phenylalanine at position 54.

Table III. Evaluation of aldo-keto reductase family 1 member C1 (*AKR1C1*) variants in the general population.

rsID	Amino Acid Variant	Relative Entropy	Predicted Stability Changes (ΔΔG) (kcal/mol)	Outcome	Allele Frequency
rs999611958	His48Arg	5.32	-1.52	Destabilizing (Disrupts folding)	7.95e-06
-	Tyr55His	5.00	-1	Destabilizing (Disrupts catalysis)	3.97e-06
rs1462840208	Trp86Ser	6.07	-3.769	Highly Destabilizing (Disrupts folding)	4.03e-06
rs754792432	His117Asp	5.32	-2.164	Highly Destabilizing (Disrupts catalysis)	3.98e-06
rs778903438	His117Pro	5.32	-1.26	Destabilizing (Disrupts catalysis)	2.83e-05
-	His117Arg	5.32	-1.112	Destabilizing (Disrupts catalysis)	6.57e-06
rs752532298	Pro119Thr	4.53	-2.639	Highly Destabilizing (Disrupts folding)	3.98e-06
rs782186892	Trp227Arg	5.83	-0.779	Destabilizing (Disrupts substrate binding)	1.59e-05

important for the pathogenesis of lipedema because they could drive adipogenesis, anxiety, and pain, three common features of lipedema.

In this work, we identified three genetic missense variants in *AKR1C1* in the genome of women with lipedema, which were detrimental to the protein's function, either preventing substrate activation or the cofactor binding site. Further studies on larger cohorts or that provide other laboratory data will be needed to corroborate the results of this investigation.

Moreover, our computational approach allowed us to predict pathogenic variants in the general population. From 265 *AKR1C1* variants taken from the gnomAD database[41], 8 of them proved to be detrimental to the function of the protein by the criteria developed in this study. This way, genetic variants found in the general population can be quickly screened for their pathogenicity, allowing for the identification of individuals at risk of developing lipedema.

Since lipedema is underdiagnosed and hormones influence its pathogenesis, we suggest that genetic variants in *AKR1C1* could predispose to this disease, most likely in the context of hormone disequilibrium. Hormone disequilibrium could be caused by the intake of hormones from the diet (meat, dairy foods, and vegetables all contain hormones or hormone-like molecules, like phytoestrogens[42-44]) or by contact with endocrine-disrupting compounds[45,46]. Indeed, research[45,46] supports the role of endocrine disruptors on adipose tissue metabolism and related pathologies, including obesity. Thus, the control of steroid hormone levels in patients with lipedema or in women with a family history of lipedema could be important to consider in the prevention of lipedema. This, coupled with the genomic test, especially considering genes involved in steroid metabolism, like *AKR1C1*, could be a new way forward to diagnose lipedema, or to define a predisposition to this pathology.

Conclusions

In this article, we have strengthened the link between *AKR1C1* and the onset of lipedema through the study of three missense variants that impede the enzyme's function. These variants either interfere with substrate binding or cofactor binding. We adopted an information theory approach to assess the conservation of residues of interest, further developing our analysis with molecular dynamics simulations on the enzyme's bound complex, which allowed us to assess the impact of these variants at the molecular level. Finally, by combining information theory and molecular mechanics, we were able to identify *AKR1C1* variants in the general population that may disrupt the enzyme's function and predispose to lipedema.

Authors' Contributions

Conceptualization: MB; methodology: JK and AB; software: JK and AB; investigation: KLH, SandroM, and PC; data curation: GB, PEM, GM, KDo and KDh; writing – original draft preparation: JK, GB, and AB; writing – review and editing: SerenaM, SilviaM and AB; supervision: AB and MB; project administration: AB and MB; funding acquisition: MB. These authors contributed equally to this work: JK and GB.

Funding

This research was funded by the Provincia Autonoma di Bolzano in the framework of LP 14/2006.

Ethics Approval

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethical Committee of Azienda Sanitaria dell'Alto Adige, Italy (Approval No. 132-2020).

Informed Consent
Informed consent was obtained from all subjects involved in the study, which includes consent to use the anonymized genetic results for research.

Data Availability
Data are contained within the supplementary material.

Acknowledgments
The authors wish to thank all the patients and the clinicians for their kind collaboration.

Conflicts of Interest
The authors declare no conflicts of interest.

References

1) Paolacci S, Precone V, Acquaviva F, Chiurazzi P, Fulcheri E, Pinelli M, Buffelli F, Michelini S, Herbst KL, Unfer V. Genetics of Lipedema: New Perspectives on Genetic Research and Molecular Diagnoses. Eur Rev Med Pharmacol Sci 2019; 23: 5581-5594.
2) Warren AG, Janz BA, Borud LJ, Slavin SA. Evaluation and Management of the Fat Leg Syndrome. Plast Reconstr Surg 2007; 119: 9e-15e.
3) Bonetti G, Herbst KL, Donato K, Dhuli K, Kiani AK, Aquilanti B, Velluti V, Matera G, Iaconelli A, Bertelli M. Dietary Supplements for Obesity. J Prev Med Hyg 2022; 63: E160-E168.
4) Forner-Corder I, Szolnoky G, Forner-Cordero A, Kemény L. Lipedema: An Overview of Its Clinical Manifestations, Diagnosis and Treatment of the Disproportional Fatty Deposition Syndrome - Systematic Review. Clin Obes 2012; 2: 86-95.
5) Michelini S, Chiurazzi P, Marino V, Dell'Orco D, Manara E, Baglivo M, Fiorentino A, Maltese PE, Pinelli M, Herbst KL. Aldo-Keto Reductase 1C1 (AKR1C1) as the First Mutated Gene in a Family with Nonsyndromic Primary Lipedema. Int J Mol Sci 2020; 21: 6264.
6) Precone V, Barati S, Paolacci S, Salgarello M, Visconti G, Gentileschi S, Guerri G, Gagliardi L, Aquilanti B, Matera G. Genetic Syndromes with Localized Subcutaneous Fat Tissue Accumulation. Acta Biomed 2019; 90: 90-92.
7) Michelini S, Herbst KL, Precone V, Manara E, Marceddu G, Dautaj A, Maltese PE, Paolacci S, Ceccarini MR, Beccari T. A Multi-Gene Panel to Identify Lipedema-Predisposing Genetic Variants by a Next-Generation Sequencing Strategy. J Pers Med 2022; 12: 268.
8) Bonetti G, Herbst KL, Dhuli K, Kiani AK, Michelini S, Michelini S, Ceccarini MR, Michelini S, Ricci M, Cestari M. Dietary Supplements for Lipedema. J Prev Med Hyg 2022; 63: E169-E173.
9) Camilleri G, Kiani AK, Herbst KL, Kaftalli J, Bernini A, Dhuli K, Manara E, Bonetti G, Stuppia L, Paolacci S. Genetics of Fat Deposition. Eur Rev Med Pharmacol Sci 2021; 25: 14-22.
10) Kiani AK, Mor M, Bernini A, Fulcheri E, Michelini S, Herbst KL, Buffelli F, Belgrado JP, Kaftalli J, Stuppia L. Steroid-Converting Enzymes in Human Adipose Tissues and Fat Deposition with a Focus on AKR1C Enzymes. Eur Rev Med Pharmacol Sci 2021; 25: 23-32.
11) Penning TM, Wangtrakuldee P, Auchus RJ. Structural and Functional Biology of Aldo-Keto Reductase Steroid-Transforming Enzymes. Endocr Rev 2019; 40: 447-475.
12) Gil N, Fiser A. Identifying functionally informative evolutionary sequence profiles. Bioinformatics 2018; 34: 1278-1286.
13) Shannon CE. Prediction and Entropy of Printed English. Bell Syst Tech J 1951; 30: 50-64.
14) Mihalek I, Reš I, Lichtarge O. Background Frequencies for Residue Variability Estimates: BLOSUM Revisited. BMC Bioinformatics 2007; 8: 488.
15) Gao H, Dou Y, Yang J, Wang J. New Methods to Measure Residues Coevolution in Proteins. BMC Bioinformatics 2011; 12: 206.
16) Wang K, Samudrala R. Incorporating Background Frequency Improves Entropy-Based Residue Conservation Measures. BMC Bioinform. 2006; 7: 385.
17) Dou Y, Zheng X, Wang J. Several Appropriate Background Distributions for Entropy-Based Protein Sequence Conservation Measures. J Theor Biol 2010; 262: 317-322.
18) Wold LE, Hines EA, Allen E. Lipedema of the Legs; a Syndrome Characterized by Fat Legs and Edema. Ann Intern Med 1951; 34: 1243-1250.
19) Bonetti G, Paolacci S, Samaja M, Maltese PE, Michelini S, Ricci M, Cestari M, Dautaj A. Low Efficacy of Genetic Tests for the Diagnosis of Primary Lymphedema Prompts Novel Insights into the Underlying Molecular Pathways. Int J Mol Sci 2022; 23: 7414.
20) Bonetti G, Dhuli K, Ceccarini MR, Kaftalli J, Samaja M, Precone V, Cecchin S, Maltese PE, Guerri G, Marceddu G. Next-Generation Sequencing of a Large Gene Panel for Outcome Prediction of Bariatric Surgery in Patients with Severe Obesity. J Clin Med 2022; 11: 7531.
21) Michelini S, Herbst KL, Precone V, Manara E, Marceddu G, Dautaj A, Maltese PE, Paolacci S, Ceccarini MR, Beccari T. A Multi-Gene Panel to Identify Lipedema-Predisposing Genetic Variants by a Next-Generation Sequencing Strategy. J Pers Med 2022; 12: 268.
22) Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 2014; 133: 1-9.

23) Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Available at: https://omim.org/.

24) Orphanet: an online database of rare diseases and orphan drugs. Copyright, INSERM 1997. Available at: http:www.orpha.net.

25) Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Amemiya A. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2023.

26) Bethesda (MD): National Library of Medicine (US), NCBI; 1988. Available at: https://pubmed.ncbi.nlm.nih.gov/20301295/https://www.ncbi.nlm.nih.gov/.

27) Mattassi R, Manara E, Colombo PG, Manara S, Porcella A, Bruno G, Bruson A, Bertelli M. Variant Discovery in Patients with Mendelian Vascular Anomalies by Next-Generation Sequencing and Their Use in Patient Clinical Management. J Vasc Surg 2018; 67: 922-932.e11.

28) Marceddu G, Dallavilla T, Guerri G, Zulian A, Marinelli C, Bertelli M. Analysis of Machine Learning Algorithms as Integrative Tools for Validation of next Generation Sequencing Data. Eur Rev Med Pharmacol Sci 2019; 23: 8139-8147.

29) Shannon CE. A Mathematical Theory of Communication. Bell Syst tech j 1948; 27: 379-423.

30) Couture JF, Legrand P, Cantin L, Luu-The V, Labrie F, Breton R. Human 20α-Hydroxysteroid Dehydrogenase: Crystallographic and Site-Directed Mutagenesis Studies Lead to the Identification of an Alternative Binding Site for C21-Steroids. J Mol Biol 2003; 331: 593-604.

31) Pires DE, Ascher DB, Blundell TL. MCSM: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. Bioinformatics 2014; 30: 335-342.

32) Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl EG: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX 2015; 1-2: 19-25.

33) Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. J Comput Chem 2010; 31: 671-690.

34) Kumari R, Kumar R, Lynn A. A GROMACS Tool for High-Throughput MM-PBSA Calculations. J Chem Inf Model 2014; 54: 1951-1962.

35) Jin Y, Penning TM. Molecular Docking Simulations of Steroid Substrates into Human Cytosolic Hydroxysteroid Dehydrogenases (AKR1C1 and AKR1C2): Insights into Positional and Stereochemical Preferences. Steroids 2006; 71: 380-391.

36) Genheden S, Ryde U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. Expert Opin Drug Discov 2015; 1: 449-461.

37) Hara A, Matsuura K, Tamada Y, Sato K, Miyabe Y, Deyashiki Y, Ishida N. Relationship of Human Liver Dihydrodiol Dehydrogenases to Hepatic Bile-Acid-Binding Protein and an Oxidoreductase of Human Colon Cells. Biochem J 1996; 313: 373-376.

38) Bertelli M, Manara E, Maltese PE, Michelini S, Michelini S, Ceccarini MR, Chiurazzi P, Herbst KL. Methods of Treating Lipedema Including Akr1c1 as a Therapeutic Target. Patent 2022. Available at: https://patents.google.com/patent/US20220362260A1/en?q=(bertelli+matteo)&oq=(bertelli+matteo).

39) Matsuura K, Deyashiki Y, Sato K, Ishida N, Miwa G, Hara A. Identification of Amino Acid Residues Responsible for Differences in Substrate Specificity and Inhibitor Sensitivity between Two Human Liver Dihydrodiol Dehydrogenase Isoenzymes by Site-Directed Mutagenesis. Biochem J 1997; 323: 61-64.

40) Zhang B, Zhu DW, Hu XJ, Zhou M, Shang P, Lin SX. Human 3-Alpha Hydroxysteroid Dehydrogenase Type 3 (3α-HSD3): The V54L Mutation Restricting the Steroid Alternative Binding and Enhancing the 20α-HSD Activity. J Steroid Biochem Mol Biol 2014; 141: 135-143.

41) Chen S, Francioli LC, Goodrich JK, Collins RL, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, Poterba T, Wilson MW, Tarasova Y, Phu W, Yohannes MT, Koenig Z, Farjoun Y, Banks E, Donnelly S, Gabriel S, Gupta N, Ferriera S, Tolonen C, Novod S, Bergelson L, Roazen D, Ruano-Rubio V, Covarrubias M, Llanwarne C, Petrillo N, Wade G, Jeandet T, Munshi R, Tibbetts K. gnomAD Project Consortium, O'Donnell-Luria A, Solomonson M, Seed C, Martin AR, Talkowski ME, Rehm HL, Daly MJ, Tiao G, Neale BM, MacArthur DG, Karczewski KJ. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. bioRxiv 2022. Available at: https://www.biorxiv.org/content/10.1101/2022.03.20.485034v1.

42) Stephany RW. Hormones in Meat: Different Approaches in the EU and in the USA. APMIS Suppl 2001; S357-363, S363-364.

43) Kuryłowicz A, Cąkała-Jakimowicz M, Puzianowska-Kuźnicka M. Targeting Abdominal Obesity and Its Complications with Dietary Phytoestrogens. Nutrients 2020; 12: 582.

44) Malekinejad H, Rezabakhsh A. Hormones in Dairy Foods and Their Impact on Public Health - A Narrative Review Article. Iran J Public Health 2015; 44: 742-758.

45) Le Magueresse-Battistoni B. Adipose Tissue and Endocrine-Disrupting Chemicals: Does Sex Matter? Int J Environ Res Public Health 2020; 17: 9403.

46) Darbre PD. Endocrine Disruptors and Obesity. Curr Obes Rep 2017; 6: 18-27.