Research

A novel 9-bp insertion detected in steroid 21-hydroxylase gene (CYP21A2): prediction of its structural and functional implications by computational methods

Sudhisha Dubey*[1], Susan Idicula-Thomas[2], Mohammad Anwaruddin[2], Chinnaraj Saravanan[1], R Raveendra Varma[3] and Anurupa Maitra[1]

Address: [1]Department of Molecular Endocrinology, National Institute for Research in Reproductive Health, Indian Council of Medical Research, J M Street, Parel, Mumbai, Maharashtra, India, [2]Biomedical Informatics Centre of Indian Council of Medical Research, National Institute for Research in Reproductive Health, J M Street, Parel, Mumbai, Maharashtra, India and [3]Department of Pediatrics and Neonatology, Mother's Hospital Trissur, Kerala, India

Email: Sudhisha Dubey* - dsudhisha@yahoo.com; Susan Idicula-Thomas - susansherry@gmail.com; Mohammad Anwaruddin - anwar.m1@gmail.com; Chinnaraj Saravanan - saravanankpd@rediffmail.com; R Raveendra Varma - rrvarma@hotmail.com; Anurupa Maitra - anurupamaitra@yahoo.co.in

* Corresponding author

Abstract

Background: Steroid 21-hydroxylase deficiency is the most common cause of congenital adrenal hyperplasia (CAH). Detection of underlying mutations in *CYP21A2* gene encoding steroid 21-hydroxylase enzyme is helpful both for confirmation of diagnosis and management of CAH patients. Here we report a novel 9-bp insertion in *CYP21A2* gene and its structural and functional consequences on P450c21 protein by molecular modeling and molecular dynamics simulations methods.

Methods: A 30-day-old child was referred to our laboratory for molecular diagnosis of CAH. Sequencing of the entire *CYP21A2* gene revealed a novel insertion (duplication) of 9-bp in exon 2 of one allele and a well-known mutation I172N in exon 4 of other allele. Molecular modeling and simulation studies were carried out to understand the plausible structural and functional implications caused by the novel mutation.

Results: Insertion of the nine bases in exon 2 resulted in addition of three valine residues at codon 71 of the P450c21 protein. Molecular dynamics simulations revealed that the mutant exhibits a faster unfolding kinetics and an overall destabilization of the structure due to the triple valine insertion was also observed.

Conclusion: The novel 9-bp insertion in exon 2 of *CYP21A2* genesignificantly lowers the structural stability of P450c21 thereby leading to the probable loss of its function.

Background

Congenital adrenal hyperplasia (CAH; OMIM# 201910) is an autosomal recessive disorder caused by deficiency of one of the five steroidogenic enzymes involved in cortisol biosynthesis. Steroid 21-hydroxylase deficiency accounts for about 90–95% of all CAH cases [1]. Deficiency of cor-

tisol results in excessive production of androgens leading to prenatal virilization in females and rapid somatic growth in both sexes [2]. CAH has been traditionally divided into three forms; severe salt wasting (SW), less severe simple virilizing (SV) and asymptomatic non-classic (NC) form. The SW form is common and found in 75% of all CAH patients. In addition to decreased cortisol, aldosterone biosynthesis is also impaired in these patients resulting in severe renal salt loss and hypotonic shock, unless treated during the neonatal period [2]. Patients with SV form do not have aldosterone deficiency and thus salt loss is not present. Symptoms like variable degree of ambiguous genitalia in females, growth acceleration and pseudoprecocious puberty are seen in SV forms. The milder non-classical form is asymptomatic at birth and presents with various degrees of late onset features of hyperandrogenism.

The wide spectrum of clinical manifestations seen in this disorder is due to varied degrees of enzyme activity caused by different mutations in the *CYP21A2* gene encoding the steroid 21-hydroxylase enzyme. Although, both large and point mutations have been seen in different populations [3-7], point mutations constitute a larger proportion. The *CYP21A2* gene is part of a complicated structure, referred to as the RCCX-module located in human leukocyte antigen class III locus on chromosome 6p21.3 [8]. Approximately 30 kb upstream of *CYP21A2*, a pseudogene (*CYP21A1P*) is present; which is 98% identical to *CYP21A2*. *CYP21A1P* cannot synthesize the functional protein due to presence of numerous deleterious mutations.

About 95% of the mutated alleles in patients with steroid 21-hydroxylase deficiency are generated by transfer of DNA sequences from *CYP21A1P* to *CYP21A2* by gene conversion events [9]. The remaining 5% alleles have new/rare mutations due to random events [2]. The number of rare mutations identified has increased dramatically in the last few years [10]. Most of these mutations are unique to individual families but some are population specific [11-13]. Functional characterization of most of the mutants using site directed mutagenesis followed by *in vitro* expression analyses have been helpful in correlating clinical severity to the degree of loss of enzyme function caused by the mutations [10]. However such studies are difficult to perform in routine laboratory set up with limited resources. Alternatively, these experimental assays could be complemented with computational studies; wherein the structural and functional perturbations of a protein by virtue of mutation/s could be predicted. Consequence of single mutation (W62G) on structural stability of lysozyme [14] and structural perturbations caused by three different activating mutations of the *hLHR* gene found in four unrelated Brazilian boys with male-limited preco-

cious puberty have been studied by MD simulations [15]. Computational scientists have also used MD simulations to independently explore the role of mutations on protein stability and activity [16-22]. In the present study, molecular modeling and MD simulations were carried out to understand the structural and probable functional implications of the triple valine insertion of human P450c21.

Materials and methods
Clinical history of proband
The proband presented on day 30 with ambiguous genitalia and failure to thrive. She was the first child born of non-consanguineous marriage with an uneventful antenatal history. She was born at full term by normal delivery. Her clinical examination showed prominent phallus with clitoral hypertrophy and partial fusion of labia (Prader stage IV). Gonads were non-palpable. Genitogram and panendoscopy revealed normal bladder, common urogenital sinus, vagina, and uterine cervix. Her weight was initially 2.9 kg (well below the 0.4th centile). Later at 12 weeks her length was 54.7 cm (2nd centile) and weight 4.3 kg (2nd centile). Hormonal examination revealed elevated 17-hydroxyprogesterone (17-OHP) (8800 ng/dl; normal value < 630 ng/dl), testosterone (1.5 ng/ml; normal range 0–0.28 ng/ml) and low cortisol (5.6 ug/dl; normal range 11–18 ug/dl). Sodium and potassium levels were 138 mmol/l and 6.1 mmol/l respectively. Her chromosomal analysis revealed 46, XX normal karyotype. She was diagnosed as simple virilising CAH and put on a relatively low dose of 5 mg/m^2 of hydrocortisone and 50 µg of fludrocortisone daily. In second year, she showed normal development with 17-OHP level of 560 ng/dl. Her weight was 8 kg, height – 74 cm and head circumference – 45.3 cm (all below 0.4th centile). Her bone age was normal. She was continued with the same dose till her third year. Thereafter her dose had to be increased to 9 mg of hydrocortisone and 100 µg of fludrocortisone daily due to increased level of 17-OHP (7200 ng/dl). She is now 5 yrs old and developing normally.

Molecular analysis
Informed consent was obtained from both parents and the study was approved by the Ethical Committee of our institution. About 2 ml of whole blood was collected in EDTA vacutainers (Becton-Dickinson, USA) for *CYP21A2* gene analysis and DNA was extracted using Qiagen kit (QIAmp DNA Blood Kit, QIAGEN GmbH, Hilden). 200 ng of genomic DNA was subjected to selective amplification of *CYP21A2* gene in two different fragments of 1.2 and 2.3 kb respectively using previously described primers [23]. PCR was carried out in 2720 thermocycler (Applied Biosystem) in a reaction volume of 50 µl containing 1.5 mM MgCl$_2$, 0.2 mM dNTPs, 200 mM (NH4)$_2$SO$_4$, 750 mM Tris-HCl (pH 8.8), 0.1% Tween 20, 2 units of Taq polymerase (Fermentas, Life Sciences) and 0.5 µM of

primers. PCR was performed at 96°C for 3 min for initial denaturation followed by 30 cycles of 95°C for 1 min, 55.5°C for 25 sec, and 72°C for 3 min and final extension at 72°C for 6 min. PCR products were resolved on 1% agarose gel stained with ethidium bromide, visualized under UV transilluminator.

The PCR products were purified using QIAGEN kit (QIAmp Gel Purification Kit, QIAGEN GmbH, Hilden). Purified products were subjected to direct sequencing with ten different primers (Table 1) based on dideoxynucleotide terminator methodology using the BigDye Terminator Cycle Sequencing Ready Reaction kit (PE, Applied Biosystems, Perkin Elmer Corp., Foster City, CA). The analysis was carried out in the ABI PRISM 3130 Genetic Analyser (PE, Applied Biosystems). The numbering of the nucleotides was according to Higashi et al [24].

Web-based tools
The web-based tools were used for the purpose of secondary structure prediction, template recognition, model evaluation and sequence analysis (Table 2).

Secondary structure prediction
GeneSilico metaserver, JPRED and SYMPRED servers [25-27] were used to predict the secondary structures of human P450c21 (Table 2). The output of each of these secondary structure prediction servers is the consensus secondary structure predicted for the protein based on a collection of independent algorithms. The SYMPRED result is based on the consensus formed by the predictions of HD, PROF, SSPRO, YASPIN, JNET and PSIPRED programs while GeneSilico result is based on PSIPRED and JNET.

Identification of suitable template
FUGUE (v2.s.07; Table 2) was used for identification of template for modeling the structure of human P450c21 [28]. The best template is identified based on their Z-scores. The templates which have a Z-score of 6 or more are labeled as "*certain*" with 99% confidence.

Assessment of the theoretical structure available for human P450c21
The theoretical structure of human P450c21 (PDB ID: 2GEG) has been elucidated using rabbit P4502c5 (PDB ID: 1N6B) as template [29]. The quality of this theoretical structure was assessed in two ways. Firstly, using the alignment of the target-template provided by FUGUE, the agreement between the predicted consensus secondary structure of human P450c21 with that of rabbit P4502c5 secondary structure was compared. Secondly, the quality of the theoretical model was evaluated using Verify3D [30] and Colorado3D server [31] (Table 2).

Generation of structure for the mutant
The structure of mutant (MT) was modeled using the theoretical structure of wild-type (WT; 2GEG) as the template. Discovery Studio v 1.7 (Accelrys) was used for building the homology-based model.

MD simulations
MD simulations were performed using GROMACS 3.3.1 [32]. United atom representation was used except for polar and aromatic ring hydrogen atoms. GROMOS96 forcefield was used for energy calculations. Van der Waals interactions were calculated with a distance cut-off of 0.9 nm. Electrostatic interactions were treated using the cut-off method [33]. Neutralizing counter ions were added when charged residues were present. The models were solvated with SPC water molecules [34] and simulated in a triclinic box [35,36] with periodic boundary conditions. The simulations were performed in the canonical NVT ensemble. The models were first energy minimized using steepest descent algorithm with a tolerance of 1 J mol^{-1} nm^{-1}; this was followed by position restrained MD simulations for 10 ps. Initial velocities were generated conforming to Maxwell velocity distribution. A time step of integration of 2 fs was used. LINCS algorithm was used to

Table 1: List of primers used for sequencing the entire *CYP21A2*

Primer	Primer sequence	Nucleotide Number	Size	Location
1F	5' aca gtc tac aca gca gga gg3'	-78–(-59)	20	Promoter
1R	5' gtg agg gcc aga gcg aga t 3'	214–232	19	Intron 1
2F	5' gag gac cat tga tga agc 3'	322–339	18	Exon 2
2R	5' ctc aca gaa ctc ctg ggt ca 3'	806–825	20	Exon 3
4F	5' cct gtc ctt ggg aga cta ct 3'	700–719	20	Exon 3
4R	5' gtc cac aat ttg gat gga cca 3'	1187–1207	21	Exon 5
6R	5' agc aat gct gag gcc ggt ag 3'	1460–1479	20	Intron 6
7F	5' att gct atg agg cgg gtt c3'	1474–1492	19	Intron 6
8F	5' ctc act ggg ttg ctg agg gag 3'	1901–1921	21	Intron 7
9F	5' ctt cag cat ctc cgg cta c 3'	2238–2256	19	Intron 8
10R	5' ctg tgt tta cag ggg gga 3'	2778-2796	18	Intron 10

Table 2: Description of the web-based tools used in the study

Tool	Description	URL	Reference
WebLogo	Sequence Logo	http://weblogo.berkeley.edu/logo.cgi	41
FUGUE	Template identification	http://tardis.nibio.go.jp/fugue/prfsearch.html	28
JPRED	Secondary structure prediction	http://www.compbio.dundee.ac.uk/~www-jpred/	26
GeneSilico metaserver	Fold prediction metaserver	https://genesilico.pl/meta2/	25
SYMPRED	Secondary structure prediction	http://zeus.cs.vu.nl/programs/sympredwww/	27
Verify3D	Structure evaluation	http://nihserver.mbi.ucla.edu/Verify_3D/	30
COLORADO3D	Structure evaluation	http://asia.genesilico.pl/colorado3d/	31
PSI-BLAST	Homolog identification	http://www.ncbi.nlm.nih.gov/blast/Blast.cgi	42

constrain the bonds [37]. The productive run was initiated for 10 ns; this duration was sufficient to compare the stabilities of the protein structures. The simulations were performed at 300K.

Analysis of MD trajectories

With an objective of understanding the structural and functional implications of the triple valine insertion, trajectories of WT and MT were analyzed for the following structural properties as a function of time: a) Potential energy, b) The root mean square deviation (RMSD) of the Cα atoms with respect to the starting conformation, c) RMSD of the hydrophobic residues postulated to be involved in membrane binding [29] with respect to the starting conformation d) Distance between the Cα atoms of residues S108 and D287; both of which are suggested to be involved in heme and substrate-binding [29], e) Structural perturbations near the insertion site. The trajectories of the simulations were plotted using XMGRACE [38]. DSSP [39] program was used for secondary structure assignment and MolScript [40] was used to create and render the molecular images.

Sequence logo

For understanding the conservation of residues in the immediate vicinity of the insertion, a sequence logo was constructed using the program WebLogo [41]. Proteins that share more than 30% similarity with human P450c21 and are deposited in the SwissProt database were identified using the PSI-BLAST algorithm [42]. The 50 identified homologs were all cytochromes and shared similarity throughout the length of the protein. These proteins were then subjected to multiple sequence alignment using CLUSTALW [43] and a sequence logo was constructed.

Results

Mutational analysis

Clinical and biochemical findings in our patient were consistent with her simple virilizing phenotype. Entire sequencing of *CYP21A2* gene (including 10 exons, 9 introns and < 400 base pairs upstream to the transcription initiation codon) was performed in order to detect known and unknown mutations. Electropherogram showed a known I172N mutation in exon 4 of one allele and an in-frame insertion of 9-bp TGTGGTGGT at nucleotide 306 in exon 2 of other allele (Fig. 1A). This 9-bp insertion resulted in triple valine insertion between V70 and L71 of the wild type P450c21 (Fig. 1B). Re-sequencing of Exon 2 using reverse primer 2R (Table 1) showed a frameshift 9-bp away from the frameshift seen in the forward strand. Using SeqScape v2.1.1, forward and reverse sequences were aligned against the reference sequence (see Additional file 1 part A) that confirmed the insertion as highlighted sequences (see Additional file 1 part B). The shift in the insertion site was clearly observed in the forward (see Additional file 1 part C) and reverse (see Additional file 1 part D) strand that further confirmed this insertion as duplication of TGTGGTGGT base pairs in exon 2, at nucleotide position 306 of *CYP21A2* gene (see Additional file 1 part E). This novel insertion mutation was inherited from the father, while the I172N was inherited from the mother, but both parents were heterozygous for their respective *CYP21A2* mutation and, therefore, clinically healthy. This insertion/duplication in exon 2 is novel (GenBank accession no. EF661662) and has not been reported previously in patients with 21-hydroxylase deficiency. This mutation was not found in 100 control subjects.

In silico studies were performed to understand the effect of the insertion of triple valine on the structure and function of P450c21.

Effect on sequence conservation

The first step towards elucidating the effect of the insertion on the structure-function characteristics of the protein was to investigate the occurrence of conserved residues in the vicinity of the mutation. The sequence logo revealed the presence of two highly conserved, charged residues viz. E79 and R91 that occur downstream of the insertion site. In addition to these, hydrophobic residues occurring at 48, 51, 58, 61 63, 80, 81 and 82 positions and glycine residues occurring just before the insertion site at 56 and 64th position are also conserved (Fig. 2). Hence,
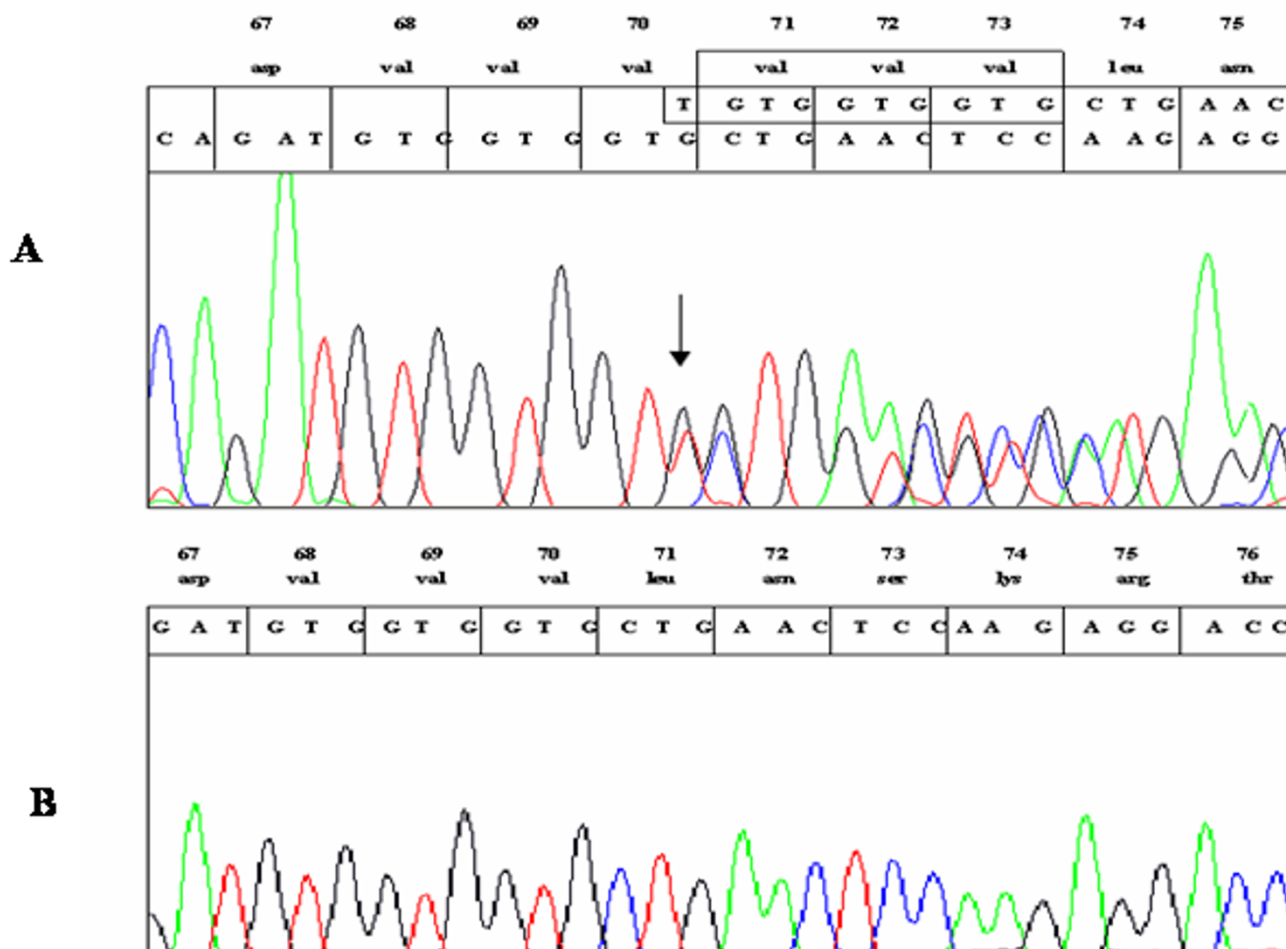
Figure 1
Direct DNA sequencing of exon 2 of CYP21A2 gene. (A) Sequence of mutated allele showing heterozygous insertion in exon 2 resulting in additional three valine residues after codon 70. The inserted sequence and the corresponding valine residues are boxed. The site of insertion is indicated by an arrow. (B) The lower sequence corresponds to the wild-type allele showing three valine residues.

the mutation could cause a drastic change in the position of these conserved residues and thus lead to loss of structure and function of the protein.

Secondary structure prediction
The secondary structures for human P450c21 were predicted using various online softwares (Table 2). Except for a few regions, there was good agreement in the predictions from the various servers (see Additional file 2). There does not seem to be any discordance with the secondary structures predicted at the site of mutation (see Additional file 2).

Identification of suitable template
Cytochrome P4502c5 of rabbit (PDB ID: 1DT6) was identified as the best template, for modeling the structure of human P450c21, by FUGUE (Z-score of 33.12; "*certain*" with 99% confidence). As compared to the other templates, it also showed full-length sequence alignment with least number of gaps. The consensus predicted secondary structural elements of human P450c21 also aligned fairly well (see Additional file 3). There exists another experimentally elucidated structure for rabbit P4502c5 [1N6B; [44]] with a better resolution as compared to 1DT6. The secondary structure compositions of both these PDB entries are identical and the RMSD between them is 0.8 Å. Hence, 1N6B was considered as the ideal template for modeling of human P450c21. A theoretical structure of human P450c21 (PDB ID: 2GEG) has been elucidated using rabbit P4502c5 (PDB ID: 1N6B) as template [29]. This structure was thus used to represent the WT and also serve as the template for modeling the MT.
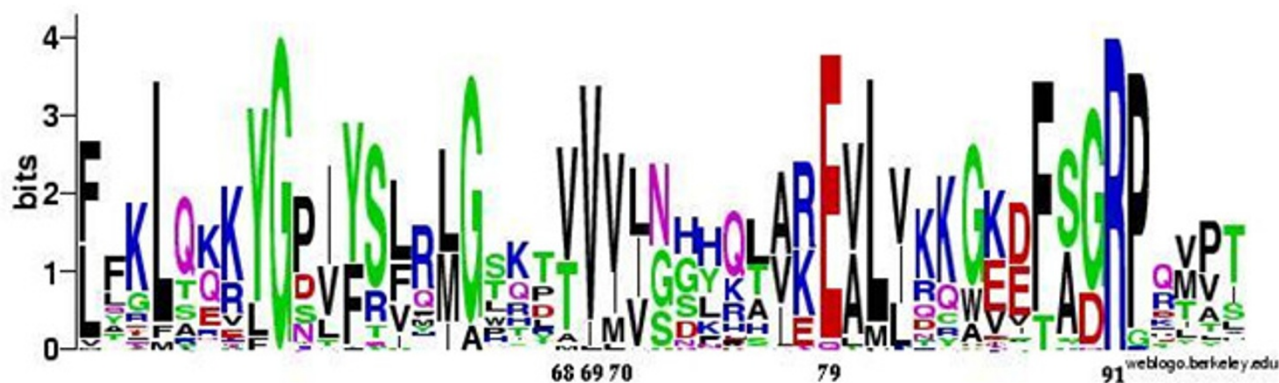
Figure 2
Sequence logo depicting the conservation of residues near the insertion site. The triple valine insertion occurs at position 71. The insertion site is flanked by many conserved residues.

Validation of the theoretical structures

The Verify3D analyses of theoretical models of both WT and MT revealed that most of the residues had a positive score. For low-scoring regions of the model, it was observed that the corresponding regions of the template (1N6B) too shared a low score. The structural anomalies could thus have been passed down from the template dur-ing modeling (see Additional file 4). The structures were also colour-rendered based on the Verify3D scores gener-ated by the Colorado3D server. The major part of the structures, in all the cases, had a good score (see Addi-tional file 5).
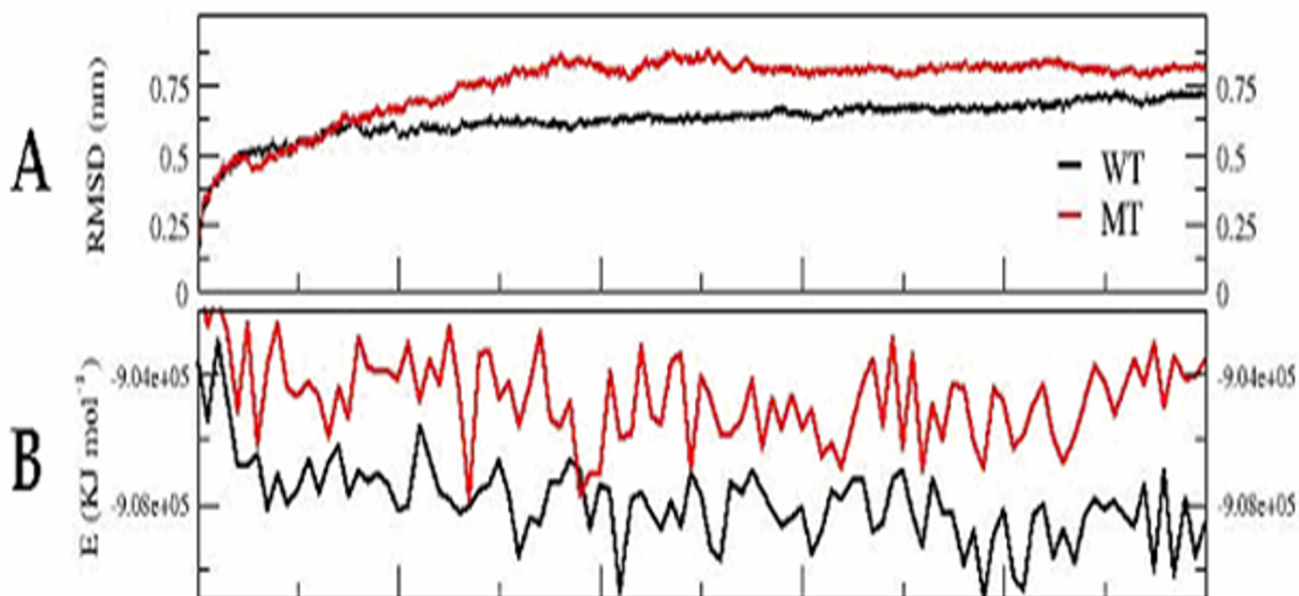

Figure 3
Cα-RMSD plot of the WT and MT P450c21 structures with respect to the starting conformation during the course of the simulation (A). The change in the potential energy of the WT and MT structures as a function of time. (B). WT and MT plots are rendered in black and red respectively. The figure was prepared using XMGRACE (38).

Structural implications of the insertion
Loss of structural stability

The WT and MT structures were subjected to MD simulations. The RMSD analysis of the trajectory revealed that the WT dynamics attained equilibration at around 3 ns with an RMSD of around 0.5 nm, while the MT attained equilibrium at around 6 ns with a higher RMSD (0.75 nm; Fig. 3A). In concurrence to the above observation, the potential energy of the MT remains higher through out the dynamics as compared to the WT (Fig. 3B). The loss of the secondary structures at the region near the insertion site during the course of the simulation clearly reveals a faster unfolding of the MT as compared to the WT (Fig. 4). All these observations suggest that the triple valine insertion causes an overall destabilization of the protein.

Loss of H-bond interactions at the vicinity of insertion

Further examination of the atomic-level interactions at the vicinity of the insertion present in human P450c21 (WT) revealed that residues N72 and E79 participate in hydrogen bonding with distantly placed residues viz. T52, N387 and S374 (Fig. 5). Trajectory analysis revealed that the hydrogen bond interactions between E79 and S374 seem to be destroyed in the early stages of simulation for MT whereas these were retained in WT (see Additional file 6 part A). The hydrogen bonds between N72-N387 and N72-T52 were also lost rapidly for the MT (see Additional file 6 parts B and C).

Functional implications of the insertion
Heme and substrate-binding

Docking and homology studies of cytochrome P450s have helped in identification of putative heme and substrate-binding residues. The residues S108 and D287 have thus been implicated in heme and substrate-binding [29]. These residues were seen to be interacting with each other in P450c21. The change in distance between the Cα atoms of these residues was monitored for the trajectories as a function of time for both the WT and MT structures (Fig. 6A). The distance between these two residues in the native WT and MT structures is 0.6 nm. It has to be noted here that the starting structures for simulations of WT and MT were identical except for the triple valine insertion and the residues S108 and D287 were distantly placed from the insertion site. In WT dynamics the distance between these two residues was maintained at around 0.6 nm, while for the MT the distance initially increased to around 2.25 nm and later stabilizes at around 1.25 nm (Fig. 6A).

Hydrophobic interactions

Based on Optimal Docking Area method, three segments corresponding to residues 30–42, 63–66 and 211–219 of human P450c21 were found to form a surface exposed
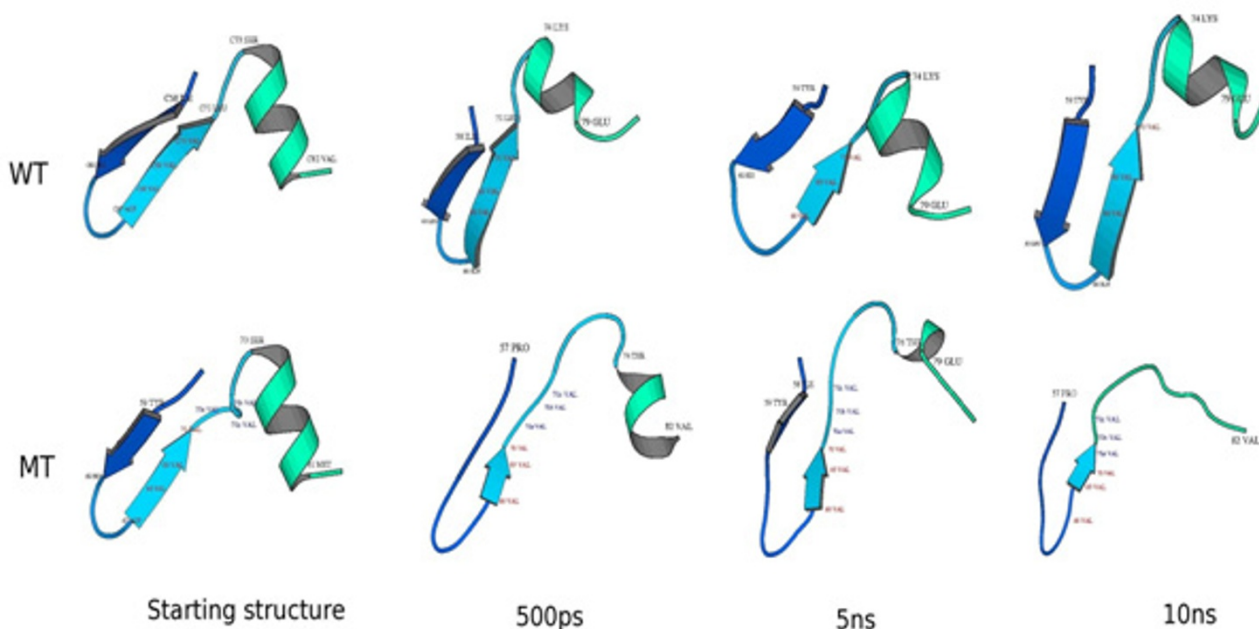

Figure 4
Snapshots of the conformation of human P450c21, near the insertion site (residues 57–82), during the course of simulation. The top and bottom panels represent the WT and MT structures respectively as a function of time. The structures have been rendered using MolScript (40) and DSSP program (39) was used for acquiring the secondary structure information.
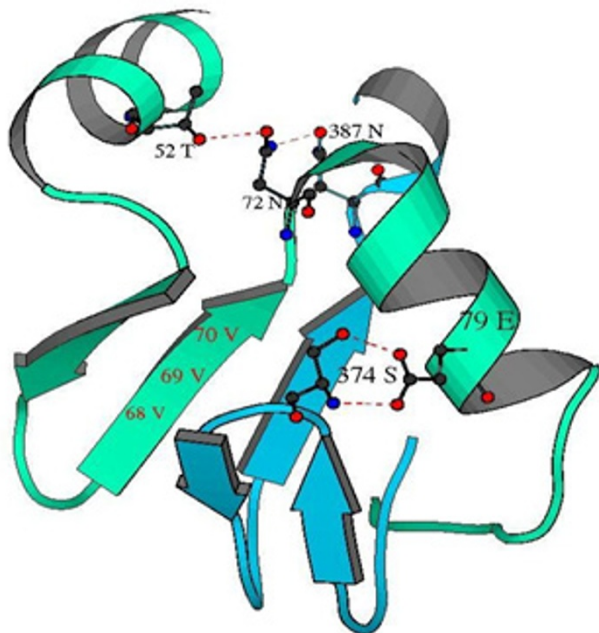
Figure 5
Partial structure of human P450c21 depicting the residues whose H-bond (dotted line) interactions get affected due to mutation. The structures in green and blue represent regions from 46–94 residues and 365–390 residues respectively. The figure has been rendered using MolScript (40).

hydrophobic patch [29]. It was hypothesized that these residues would have to participate in hydrophobic interactions to minimize the destabilizing effect caused by the surface exposed hydrophobic region. Additionally, they are positioned adjacent to the N-terminal transmembrane region of the protein. Hence these hydrophobic residues were suggested to be probably involved in binding with the ER membrane [29]. These residues were considered as a cluster and the RMS deviation of the Cα atoms of this cluster with respect to the starting conformation was compared for the WT and MT structures during the course of simulation. It was observed that the RMS deviation increased as a function of time for the MT as compared to the WT (Fig. 6B).

Discussion

CAH is found in wide range of clinical severity ranging from subtle hormone imbalance in adults to severe life threatening salt wastage in newborns. Detection of underlying mutations in *CYP21A2* gene encoding steroid 21-hydroxylase enzyme is helpful both for confirmation of diagnosis and management of CAH patients. Different kinds of mutations result in different degrees of enzymatic impairment of P450c21, which result in varied pheno-

types of CAH patients. Although, a large number of novel mutations have been reported in *CYP21A2* gene over the past few years and these mutations have continued expanding worldwide, large insertion/duplication mutations are uncommon. A duplication of 16 bases (CCT-GGATGACACGGTC) at codons 393–397 of exon 9 [45] and a large duplication of 111 bases from codons 21 to 57 inserted at codon 58 in exon 1 of the *CYP21A2* gene have been reported [46]. Once a novel mutation is encountered, it is necessary to carry out the functional studies to establish genotype – phenotype correlation so that the prognostic evaluation can be made for the proper management of the patient.

A large number of mutations detected in *CYP21A2* gene have been characterized to prove their clinical relevance and impact on P450c21 protein. For characterization of mutations, functional studies have been carried out by site directed mutagenesis followed by *in vitro* expression of the mutant protein in transiently transfected mammalian cells. The residual enzyme activity is then measured towards both natural substrates (17-OHP and progesterone) and compared with the WT protein. Percentage of enzyme activity is correlated with the clinical phenotype and accordingly the mutation is classified as SV, SW or NC type [10,47-50]. However such studies are laborious, time consuming and in some mutant proteins, the biochemical and biophysical evaluation is not possible by *in vitro* studies. In such cases, *in silico* studies can provide additional clinically useful information that could not be possible by examining the patients [29]. Molecular modeling has been used to study the putative effects of steroid 21-hydroxylase gene mutations. Structural features deduced from the models were in good correlation with clinical severity of P450c21 mutants, which shows the applicability of a modeling approach in assessment of new P450c21 mutations [29,51-54].

In this particular case, the 9-bp insertion/duplication did not result in frameshift as expected from insertions/duplications in general; instead it led to insertion of three valine residues between V70 and L71. Since this insertion mutation does not exist in pseudogene, gene conversion may not be the cause of this mutation. This duplication might have been generated by intergenic recombination during meiosis. Theoretically, this insertion could lead to absence of residual 21-hydroxylase activity resulting in the severe SW phenotype. Given the fact that CAH is an autosomal recessive disorder, the clinical severity reflects the milder mutation present in the patient. Hence SV phenotype of our patient that reflected the mild mutation I172N present on one allele, could not give additional information about the possible structural perturbation caused by the novel mutation present on the other allele. In such compound heterozygous cases with a mild known muta-
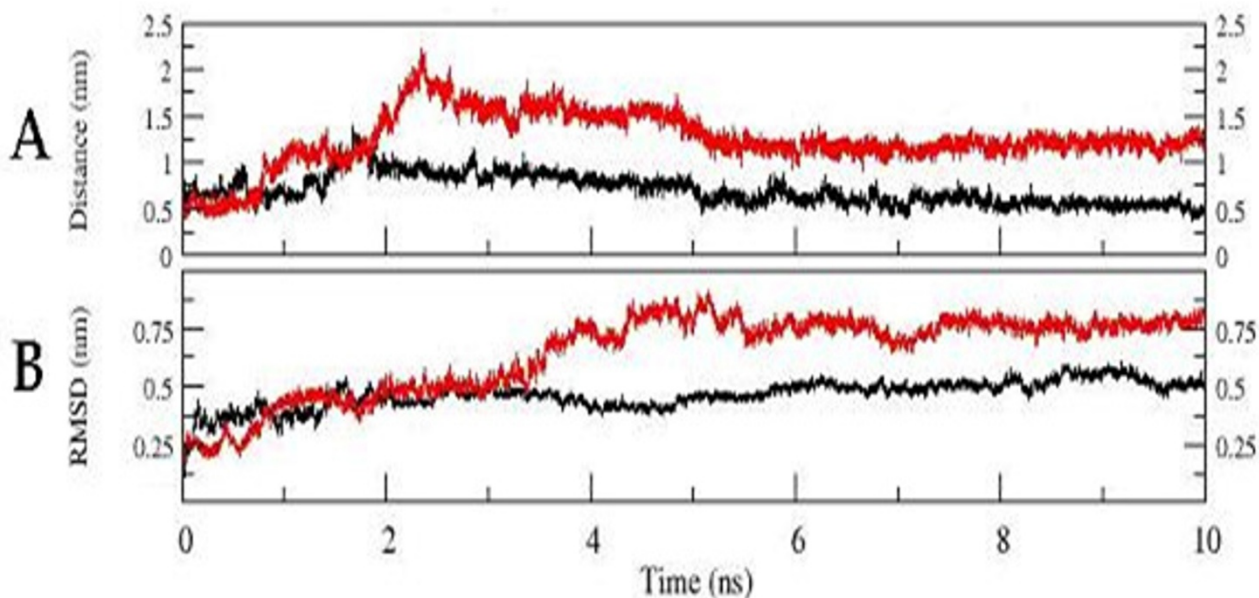
Figure 6
The change in Cα distances between the residues S108 and D287, both of which are postulated to be involved in heme and substrate binding, as a function of time. (A). Change in the CαRMSD plot of the hydrophobic patches with respect to the starting structure for WT and MT during the course of the simulation (B). WT and MT plots are rendered in black and red respectively. The figure was prepared using XMGRACE (38).

tion on one allele and a novel mutation on other allele, it is difficult to classify the mutation on the basis of clinical phenotype.

In the present study, molecular modeling and MD simulations were carried out to analyze the structural consequences of this insertion in P450c21 and to better understand the molecular pathology of CAH in the proband.

The structure of human P450c21 has not been experimentally elucidated. However, a theoretical model for it is available in the Protein Data Bank (PDB ID: 2GEG; [29]). The accuracy of this theoretical model was ascertained using structure prediction and evaluation tools (see Additional files 2, 3, 4, 5). Subsequently, using the WT as the template, the structure of the MT was modeled.

Both the WT and MT structures were subjected to MD simulations for comparison of their structural stabilities. The trajectory analyses were carried out to understand i) the effect on the structural stability and ii) plausible functional implications of the insertion.

Studies on the unfolding dynamics of the WT and MT revealed that, although the two shared an identical start-

ing structure – except for the site of insertion, the MT exhibits faster unfolding and is less stable (Figs. 3A–B). It was also observed that the secondary structures are gradually lost at the site of insertion during the course of simulation in the MT (Fig. 4). The observations indicate that the insertion seems to have a profound effect on reducing the intrinsic stability of human P450c21.

In the absence of experimental data, the interactions and stability of the regions of the protein predicted to be harboring functional importance were selected for trajectory analyses. Based on docking and homology studies of cytochrome P450s, residues S108 and D287 have been implicated in both heme and substrate-binding [29]. Docking and structural studies have also helped in identification of hydrophobic patches in human P450c21, which are suggested to be involved in ER membrane binding [29]. The results of the present simulation studies revealed that the Cα distance between the residues S108 and D287 increase drastically in the MT while it remains the same in the WT (Fig. 3C). These residues are placed far away from the site of insertion and yet their interactions seem to be affected in the mutant. This indicates that the insertion of triple valine seems to have caused not only a local structural perturbation but also a drastic disturbance in the structural stability of protein. Similarly, in the case of the hydropho-

bic regions implicated in ER membrane binding, the RMSD observed during the course of simulation suggested that this region is highly unstable in the MT as compared to the WT (Fig. 3D).

The sequence logo reveals that the insertion occurs in a region flanked by the presence of many conserved residues (Fig. 2). This observation, along with the results of the trajectory analyses, strongly suggests that the mutation could lead to a loss of the structure and function of human P450c21.

Conclusion

Analysis of *CYP21A2* gene in an Indian child with classical CAH, revealed a novel 9 base pair TGTGGTGGT insertion at nucleotide position 306 in exon 2. This insertion resulted in a triple valine insertion between V70 and L71 of P450c21. MD simulations revealed that this insertion seems to cause an overall destabilization of the structure. Although the insertion does not occur in the immediate vicinity of the postulated heme and substrate binding residues; trajectory analyses reveal that their interactions seem to get disrupted in the MT. These observations indicate that the insertion could result in SW phenotype had it been present in homozygous state. We emphasise this mutation should be added to the panel of mutations to be screened in Indian population.

Abbreviations

17-OHP: 17 hydroxyprogesterone; ER: Endoplasmic reticulum; GROMACS: Groningen Machine for Chemical Simulation; *hLHR*: Human Luteinizing Hormone Receptor; LINCS: Linear Constraint Solver.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SD conceived and designed the study, did molecular analysis, interpreted the data and drafted the manuscript. SIT did bioinformatics studies, interpreted the data and participated in drafting the manuscript. MA participated in the *in silico* analyses and their interpretations. RRV was involved with the management of the patient. CS performed the DNA sequencing. AM participated in the design of the study and critically reviewed the draft. All authors read and approved the final manuscript.

Additional material

Additional file 1

Fig. S1. SeqScape view. (A) Reference sequence of CYP21A2 gene. (Higashi et al 1986). (B) Consensus sequence generated from forward and reverse strand. Highlighted region in grey shows insertion of 9 base pair. The vertical bar indicates the site of insertion. (C) Forward sequence of exon 2 of CYP21A2 gene. (D) Reverse sequence of exon 2 of CYP21A2 gene. (E) Results of SeqScape showing heterozygous insertion of 9 bases at position 245 of the aligned sequence. Missense substitution at position 53 and 73 are detected by the software due to noisy peaks.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1423-0127-16-3-S1.tiff]

Additional file 2

Fig. S2. Alignment of the secondary structures of human P450c21 model with the secondary structures predicted using SYMPRED, JPRED and GeneSilico metaserver. The secondary structures are represented as black and blue highlights for helices (H) and strands (E) respectively. The arrow indicates the site of mutation. The red box highlights the discordant predictions in the secondary structure.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1423-0127-16-3-S2.tiff]

Additional file 3

Fig. S3. Sequence alignment of P4502c5 of rabbit (1DT6/1N6B) with human P450c21 generated by FUGUE server, along with their secondary structure information. The human P450c21 secondary structure is predicted using the online servers (Table 2). If two of the three servers have agreement for a secondary structure for a residue, then that secondary structure is taken as the consensus for that residue. The arrow indicates the site of mutation. The consensus predicted secondary structure for human P450c21 is represented in the last row. The helical and strand residues are depicted in red and blue respectively.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1423-0127-16-3-S3.tiff]

Additional file 4

Fig. S4. Verify3D plot of the theoretical models of human P450c21 (WT and MT) and template structure (rabbit P4502c5; 1N6B). Figure containing Verify3D plot of the theoretical models of human P450c21 (WT and MT) and template structure (rabbit P4502c5; 1N6B).
Click here for file
[http://www.biomedcentral.com/content/supplementary/1423-0127-16-3-S4.tiff]

Additional file 5

Fig. S5. Structures rendered based on the Verify3D scores generated by Colorado3D. The region of triple valine insertion is highlighted by a box and the site of insertion is depicted by a pink star in MT. Blue and red colours indicate the best and worst scoring regions respectively. WT represents theoretical structure for human P450c21 (PDB ID: 2GEG) and the template denotes rabbit P4502c5 (PDB ID: 1N6B).
Click here for file
[http://www.biomedcentral.com/content/supplementary/1423-0127-16-3-S5.tiff]

Additional file 6

Figure S6. The number of H-bonds between residues (A) E79-S374 (B) N72-N387 and (C) N72-T52, in WT and MT structures of human P450c21 as a function of time. The above interactions are absent in the MT throughout the course of simulation. The figure was prepared using XMGRACE (38).

Click here for file

[http://www.biomedcentral.com/content/supplementary/1423-0127-16-3-S6.tiff]

References

1. Morel Y, Miller WL: **Clinical and molecular genetics of congenital adrenal hyperplasia due to 21-hydroxylase deficiency.** *Adv Hum Genet* 1991, **20:**1-68.
2. White PC, Speiser PW: **Congenital adrenal hyperplasia due to 21-hydroxylase deficiency.** *Endocr Rev* 2000, **21:**245-291.
3. Mornet E, Crété P, Kuttenn F, Raux-Demay MC, Boue J, White PC, Boue A: **Distribution of deletions and seven point mutations on CYP21B genes in three clinical forms of steroid 21-hydroxylase deficiency.** *Am J Hum Genet* 1991, **48:**79-88.
4. Speiser PW, Dupont J, Zhu D, Serrat J, Buegeleisen M, Tusie-Luna MT, Lesser M, New MI, White PC: **Disease expression and molecular genotype in congenital adrenal hyperplasia due to 21-hydroxylase deficiency.** *J Clin Invest* 1992, **90:**584-595.
5. Wedell A, Thilén A, Ritzén EM, Stengler B, Luthman H: **Mutational spectrum of the steroid 21-hydroxylase gene in Sweden: implications for genetic diagnosis and association with disease manifestation.** *J Clin Endocrinol Metab* 1994, **78:**1145-1152.
6. Wilson RC, Wei JQ, Cheng KC, Mercado AB, New MI: **Rapid deoxyribonucleic acid analysis by allele-specific polymerase chain reaction for detection of mutations in the steroid 21-hydroxylase gene.** *J Clin Endocrinol Metab* 1995, **80:**1635-1640.
7. Bachega TAAS, Billerbeck AEC, Madureira G, Marcondes JAM, Longui CA, Leite MV, Arnhold IJP, Mendonca BB: **Molecular genotyping in Brazilian patients with the classical and nonclassical forms of 21-hydroxylase deficiency.** *J Clin Endocrinol Metab* 1998, **83:**4416-4419.
8. Yang Z, Mendoza AR, Welch TR, Zipf WB, Yu CY: **Modular variations of the human major histocompatibility complex class III genes for serine/threonine kinase RP, complement component C4, steroid 21-hydroxylase CYP21, and tenascin TNX (the RCCX module). A mechanism for gene deletions and disease associations.** *J Biol Chem* 1999, **274:**12147-12156.
9. Higashi Y, Tanae A, Inoue H, Fujii-Kuriyama Y: **Evidence for frequent gene conversion in the steroid 21-hydroxylase (P450c21) gene: Implications for steroid 21-hydroxylase deficiency.** *Am J Hum Genet* 1998, **42:**17-25.
10. **Database of CYP21A2 by human Cytochrome P450 (CYP) Allele Nomenclature Committee** [http://www.imm.ki.se/CYPalleles/cyp21.htm]
11. Wedell A, Luthman H: **Steroid 21-hydroxylase (P450c21): a new allele and spread of mutations through the pseudogene.** *Hum Genet* 1993, **91:**236-240.
12. Billerbeck AE, Bachega TA, Frazatto ET, Nishi MY, Goldberg AC, Marin ML, Madureira G, Monte O, Arnhold IJ, Mendonca BB: **A novel missense mutation, G424S, in Brazilian patients with 21-hydroxylase deficiency.** *J Clin Endocrinol Metab* 1999, **84:**2870-2872.
13. Barbaro M, Lajic S, Baldazzi L, Balsamo A, Pirazzoli P, Cicognani A, Wedell A, Cacciari E: **Functional analysis of two recurrent aminoacid substitution in the CYP21 gene from Italian patients with congenital adrenal hyperplasia.** *J Clin Endocrinol Metab* 2004, **89:**2402-2407.
14. Zhou R, Eleftheriou M, Royyuru AK, Berne BJ: **Destruction of long-range interactions by a single mutation in lysozyme.** *Proc Natl Acad Sci USA* 2007, **104:**5824-5829.
15. Latronico AC, Shinozaki H, Guerra G, Pereira MA, Lemos Marini SH, Baptista MT, Arnhold IJ, Fanelli F, Mendonca BB, Segaloff DL: **Gonadotropin-independent precocious puberty due to luteinizing hormone receptor mutations in Brazilian boys: a novel constitutively activating mutation in the first transmembrane helix.** *J Clin Endocrinol Metab* 2000, **85:**4799-4805.
16. Seibold SA, Cukier RI: **A molecular dynamics study comparing a wild-type with a multiple drug resistant HIV protease: differences in flap and aspartate 25 cavity dimensions.** *Proteins* 2007, **69:**551-565.
17. Cuesta-Lopez S, Falo F, Sancho J: **Computational diagnosis of protein conformational diseases: short molecular dynamics simulations reveal a fast unfolding of r-LDL mutants that cause familial hypercholesterolemia.** *Proteins* 2007, **66:**87-95.
18. Hamza A, Cho H, Tai HH, Zhan CG: **Molecular dynamics simulation of cocaine binding with human butyrylcholinesterase and its mutants.** *J Phys Chem B* 2005, **109:**4776-4782.
19. Cheng Q, Benson DR, Rivera M, Kuczera K: **Influence of point mutations on the flexibility of cytochrome b_5: molecular dynamics simulations of holoproteins.** *Biopolymers* 2006, **83:**297-312.
20. Achary MS, Reddy AB, Chakrabarti S, Panicker SG, Mandal AK, Ahmed N, Balasubramanian D, Hasnain SE, Nagarajaram HA: **Disease-causing mutations in proteins: structural analysis of the CYP1b1 mutations causing primary congenital glaucoma in humans.** *Biophys J* 2006, **91:**4329-4339.
21. Ceruso MA, Periole X, Weinstein H: **Molecular dynamics simulations of transducin: interdomain and front to back communication in activation and nucleotide exchange.** *J Mol Biol* 2004, **338:**469-481.
22. Menchise V, Corbier C, Didierjean C, Saviano M, Benedetti E, Jacquot JP, Aubry A: **Crystal structure of the wild-type and D30A mutant thioredoxin h of Chlamydomonas reinhardtii and implications for the catalytic mechanism.** *Biochem J* 2001, **359:**65-75.
23. Wedell A, Luthman H: **Steroid 21-hydroxylase deficiency: Two additional mutations in salt-wasting disease and rapid screening of disease causing mutations.** *Hum Mol Genet* 1993, **2:**499-509.
24. Higashi Y, Yoshioka H, Yamane M, Gotoh O, Fujii-Kuriyama Y: **Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene.** *Proc Natl Acad Sci USA* 1986, **83:**2841-2845.
25. Kurowski MA, Bujnicki JM: **GeneSilico protein structure prediction meta-server.** *Nucleic Acids Res* 2003, **31:**3305-3307.
26. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ: **Jpred: A Consensus Secondary Structure Prediction Server.** *Bioinformatics* 1998, **14:**892-893.
27. Simossis VA, Heringa J: **Optimally segmented consensus secondary structure prediction.** *Bioinformatics* 2004 in press.
28. Shi J, Blundell TL, Mizuguchi K: **FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties.** *J Mol Biol* 2001, **310:**243-257.
29. Robins T, Carlsson J, Sunnerhagen M, Wedell A, Persson B: **Molecular model of human CYP21 based on mammalian CYP2C5: Structural features correlate with clinical severity of mutations causing congenital adrenal hyperplasia.** *Mol Endocrinol* 2006, **20:**2946-2964.
30. Lüthy R, Bowie JU, Eisenberg D: **Assessment of protein models with three-dimensional profiles.** *Nature* 1992, **356:**83-85.
31. Sasin JM, Bujnicki JM: **COLORADO3D, a web server for the visual analysis of protein structures.** *Nucleic Acids Res* 2004, **32:**W586-W589.
32. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC: **GROMACS: Fast, Flexible and Free.** *J Comput Chem* 2005, **26:**1701-1718.

33. Darden T, York D, Pedersen L: **Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems.** *J Chem Phys* 1993, **98:**10089-10092.

34. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J: **Inter-molecular Forces.** In *Interaction models for water in relation to protein hydration* Edited by: Pullman B. D Reidel Publishing Company; 1981:331-342.

35. Lin FH, Graham LA, Campbell RL, Davies PL: **Structural modeling of snow flea antifreeze protein.** *Biophy J* 2007, **92:**1717-1723.

36. Louie TM, Yang H, Karnchanaphanurach P, Xie XS, Xun L: **FAD is a preferred substrate and an inhibitor of Escherichia coli general NAD(P)H: flavin oxidoreductase.** *J Biol Chem* 2002, **277:**39450-39455.

37. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM: **LINCS: A linear constraint solver for molecular simulations.** *J Comput Chem* 1997, **18:**1463-1472.

38. Turner PJ: **XMGRACE, Version 5.1.19.** Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR; 2005.

39. Kabsch W, Sander C: **Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.** *Biopolymers* 1983, **22:**2577-2637.

40. Kraulis PJ: *MOLSCRIPT*: **a program to produce both detailed and schematic plots of protein structures.** *J Appl Crystallogr* 1991, **24:**946-950.

41. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: A sequence logo generator.** *Genome Res* 2004, **14:**1188-1190.

42. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.

43. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.

44. Wester MR, Johnson EF, Marques-Soares C, Dansette PM, Mansuy D, Stout CD: **Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 A resolution: Evidence for multiple substrate binding modes.** *Biochemistry* 2003, **42:**6370-6379.

45. Lee HH, Chao HD, Lee YJ, Shu SG, Chao MC, Kuo JM, Chung BC: **Identification of four novel mutations in the CYP21 gene in congenital adrenal hyperplasia in the Chinese.** *Hum Genet* 1998, **103:**304-310.

46. Lee HH, Chang SF, Lo FS, Chao HT, Lin CY: **Duplication of 111 bases in exon 1 of the CYP21 gene is combined with deletion of CYP21P-C4B genes in steroid 21 hydroxylase deficiency.** *Mol Genet Metab* 2003, **79:**214-220.

47. Robins T, Bellanne-Chantelot C, Barbaro M, Cabrol S, Wedell A, Lajic S: **Characterization of novel missense mutations in CYP21 causing congenital adrenal Hyperplasia.** *J Mol Med* 2007, **85:**247-255.

48. Menassa R, Tardy V, Despert F, Bouvattier-Morel C, Brossier JP, Cartigny M, Morel Y: **p.H62L, a rare mutation of the CYP21 gene identified in two forms of 21-hydroxylase deficiency.** *J Clin Endocrinol Metab* 2008, **93:**1901-1908.

49. Soardi FC, Barbaro M, Lau IF, Lemos-Marini SHV, Baptista MTM, Guerra-Junior G, Wedell A, Lajic S, de Mello MP: **Inhibition of CYP21A2 Enzyme Activity Caused by Novel Missense Mutations Identified in Brazilian and Scandinavian Patients.** *J Clin Endocrinol Metab* 2008, **93:**2416-2420.

50. Riepe FG, Hiort O, Grötzinger J, Sippell WG, Krone N, Holterhus PM: **Functional and structural consequences of a novel point mutation in the CYP21A2 gene causing congenital adrenal hyperplasia: potential relevance of helix C for P450 oxidoreductase-21-hydroxylase interaction.** *J Clin Endocrinol Metab* 2008, **93:**2891-2895.

51. Krone N, Riepe FG, Grotzinger J, Partsch CJ, Sippell WG: **Functional characterization of two novel point mutations in the CYP21 gene causing simple virilizing forms of congenital adrenal hyperplasia due to 21-hydroxylase deficiency.** *J Clin Endocrinol Metab* 2005, **90:**445-454.

52. Janner M, Pandey AV, Mullis PE, Flück CE: **Clinical and biochemical description of a novel CYP21A2 gene mutation 962_963insA using a new 3D model for the P450c21 protein.** *Eur J Endocrinol* 2006, **155:**143-151.

53. Grischuk Y, Rubtsov P, Riepe FG, Grotzinger J, Beljelarskaia S, Prassolov V, Kalintchenko N, Semitcheva T, Peterkova V, Tiulpakov A, Sippell WG, Krone N: **Four Novel Missense Mutations in the CYP21A2 Gene Detected in Russian Patients Suffering from the Classical Form of Congenital Adrenal Hyperplasia: Identification, Functional Characterization, and Structural Analysis.** *J Clin Endocrinol Metab* 2006, **91:**4976-4980.

54. Baradaran-Heravi A, Vakili R, Robins T, Carlsson J, Ghaemi N, A'rabi A, Abbaszadegan MR: **Three novel CYP21A2 mutations and their protein modelling in patients with classical 21-hydroxylase deficiency from northeastern Iran.** *Clin Endocrinol* 2007, **67:**335-341.