Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism

Neda Gharani1,*, Dawn M. Waterworth1, Sari Batty2, Davinia White2, Carole Gilling-Smith2, Gerard S. Conway4, Mark McCarthy1,3, Stephen Franks2 and Robert Williamson1,5

1Department of Molecular Genetics, 2Department of Obstetrics and Gynaecology and 3Unit of Metabolic Medicine, Imperial College School of Medicine at St Mary’s, London W2 1PG, UK, 4Department of Endocrinology, The Middlesex Hospital, Mortimer Street, London W1N 8AA, UK and 5Murdoch Institute, Royal Children’s Hospital, Melbourne, Australia

Received September 16, 1996; Revised and Accepted January 6, 1997

Biochemical data implicate an underlying disorder of androgen biosynthesis and/or metabolism in the aetiology of polycystic ovary syndrome (PCOS). We have examined the segregation of the genes coding for two key enzymes in the synthesis and metabolism of androgens, cholesterol side chain cleavage (CYP11a) and aromatase (CYP19), with PCOS in 20 multiply-affected families. All analyses excluded CYP19 co-segregation with PCOS, demonstrating that this locus is not a major determinant of risk for the syndrome. However, our results provide evidence for linkage to the CYP11a locus (NPL score = 3.03, p = 0.003). Parametric analysis using a dominant model suggests genetic heterogeneity, generating a maximum HLOD score of 2.7 (\(\alpha = 0.63\)). An association study of 97 consecutively identified Europids with PCOS and matched controls demonstrates significant allelic association of a CYP11a 5’ UTR pentanucleotide repeat polymorphism with hirsute PCOS subjects (p = 0.03). A strong association was also found between alleles of this polymorphism and total serum testosterone levels in both affected and unaffected individuals (p = 0.002). Our data demonstrate that variation in CYP11a may play an important role in the aetiology of hyperandrogenaemia which is a common characteristic of polycystic ovary syndrome.

INTRODUCTION

Polycystic ovary syndrome (PCOS) is a highly prevalent endocrine disorder which is characterised by hyperandrogenaemia and represents the most common cause of anovulatory infertility and hirsutism (1–3). PCOS has been estimated to have a population prevalence of between 5–10%. The characteristic polycystic ovarian morphology, however, may be found in up to 22% of the normal population, with >90% of these women having at least one mild symptom that may be considered a clinical marker of PCOS (4).

Although it is clear that PCOS is a familial disorder, with a risk to siblings of ~50%, the precise mode of inheritance has not been agreed (5–11). A review of the various family studies carried out (12) has highlighted problems such as a lack of consensus regarding clinical phenotypes for PCOS, as well as a lack of agreement on the male phenotype. A dominant mode of inheritance has been suggested and there is evidence that premature male pattern baldness (MPB) may be the male phenotype (6,9,11). Rather than representing a simple, single gene disorder, it is more likely that PCOS has a complex genetic basis, where the interaction of multiple genetic and environmental factors determine the development of the syndrome. This would explain the observation that sisters with polycystic ovaries may present with different clinical symptoms.

Hyperandrogenaemia is seen both in women with PCOS and men with premature male pattern baldness suggesting an underlying disorder of androgen biosynthesis or metabolism (13–15). Androgens are synthesised by the adrenals, the theca cell layer of the developing ovarian follicle and the testicular Leydig cells. Both scalp hair loss and hirsutism are known to be mediated by androgens (16–18). The sensitivity of the hair follicle to androgens is dependent on a number of factors, such as serum concentrations of bioavailable androgens and the presence and number of androgen receptors (19,20).

It has been shown that theca cells from polycystic ovaries show a significant increase in both androstenedione and progesterone production in vitro when compared to normal theca (21). This suggests that the putative defect in steroidogenesis occurs at or above the level of progesterone biosynthesis. The enzyme cholesterol side chain cleavage (cytochrome P450scS) catalyzes the conversion of cholesterol to pregnenolone, the first, rate-limiting step at the start of the steroid hormone biosynthesis pathway (22,23). We hypothesise that upregulation of this enzyme could lead to an increase in androgen production.

A second steroidogenic enzyme, aromatase (cytochrome P450arom), catalyzes the conversion of the C19 steroids (androgens) to the C18 oestrogens. Evidence for altered regulation
of this enzyme in PCOS comes from the observation that granulosa cells from anovulatory polycystic ovaries are hyperresponsive to FSH in vitro, displaying significantly greater oestradiol production than granulosa cells from normal ovaries (24). Conversely, two other studies have demonstrated cases in which aromatase deficiency was associated with hyperandrogenaemia (25,26). These studies suggest that an altered regulation of this enzyme may be involved in PCOS.

The aim of the present study was to investigate P450scc and P450arom, as potential candidates for involvement in PCOS/MPB. The genes encoding P450 scc (CYP11a) and P450 arom (CYP19) are both located on the long arm of chromosome 15 at positions 15q24 and 15q21.1 respectively. Given uncertainties regarding the segregation model for PCOS, both parametric and non-parametric linkage analyses were conducted in 20 PCOS/MPB pedigrees. Mutation screening of the promoter region of CYP11a was also carried out. Finally, using a CYP11a intragenic marker, we carried out an association study in a case-control data set of consecutively identified Europid women with PCOS and matched controls.

RESULTS

Both parametric and non-parametric linkage analyses were carried out for each of the candidate loci using the GENEHUNTER program (27). The position and chromosomal order of the microsatellite markers used in the analyses are shown in Figure 1. The non-parametric results for the CYP11a locus have provided evidence for excess allele sharing, generating a maximum NPL score of 3.03 (p = 0.003) at the CYP11a locus (Table 1). Under the parametric analysis (using a dominant model with 95% penetrance and 10% phenocopy rate), the overall multipoint LOD score obtained at the CYP11a (ac)n/D15S169 locus was 1.74; a homogeneity test on these data suggested that ∼60% of the pedigrees may be linked to this locus [maximum HLOD score of 2.7, α = 0.63, computed at the CYP11a (ac)n/D15S169 locus]. It has been suggested that a p-value of <0.001 (for a non-parametric analysis) and a LOD score of >1.9 (for a parametric analysis) may be taken as suggestive evidence for linkage for a dense, complete genome scan (28). These criteria take into account the multiple testing aspect of a genome scan.

Table 1. Multipoint parametric and non-parametric analysis in the CYP11a region

<table>
<thead>
<tr>
<th>Marker</th>
<th>(cM)</th>
<th>LOD score</th>
<th>(alpha, HLOD)</th>
<th>NPL score</th>
<th>p-value</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>D15S155</td>
<td>0.00</td>
<td>-3.52</td>
<td>(0.23, 0.35)</td>
<td>0.625</td>
<td>0.250</td>
<td>0.820</td>
</tr>
<tr>
<td>D15S153</td>
<td>10.00</td>
<td>-2.98</td>
<td>(0.25, 0.44)</td>
<td>1.093</td>
<td>0.130</td>
<td>0.934</td>
</tr>
<tr>
<td>D15S125</td>
<td>12.50</td>
<td>-1.60</td>
<td>(0.28, 0.48)</td>
<td>1.243</td>
<td>0.102</td>
<td>0.943</td>
</tr>
<tr>
<td>CYP11a (ac)n/D15S169</td>
<td>21.50</td>
<td><strong>1.74</strong></td>
<td><em>(0.63, 2.67)</em></td>
<td><strong>3.034</strong></td>
<td><strong>0.003</strong></td>
<td>0.958</td>
</tr>
<tr>
<td>D15S211</td>
<td>26.30</td>
<td>-0.82</td>
<td>(0.41, 0.96)</td>
<td>2.093</td>
<td>0.021</td>
<td>0.915</td>
</tr>
</tbody>
</table>

Since we studied only a small number of relevant candidate genes, these criteria are likely to be overly stringent for our study.

The results for the CYP19 locus are summarised in Table 2. Clear exclusion of the CYP19 locus was obtained; there was no evidence for excess allele sharing across the whole region and the maximum negative LOD score generated at the CYP19 locus is −7.45.
Table 2. Multipoint parametric and non-parametric analysis in the CYP19 region

<table>
<thead>
<tr>
<th>Marker</th>
<th>(cM)</th>
<th>LOD score</th>
<th>NPL score</th>
<th>p-value</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>D15S118</td>
<td>0.00</td>
<td>−5.04</td>
<td>0.189</td>
<td>0.400</td>
<td>0.849</td>
</tr>
<tr>
<td>D15S123</td>
<td>10.00</td>
<td>−7.18</td>
<td>−0.421</td>
<td>0.664</td>
<td>0.954</td>
</tr>
<tr>
<td>CYP19[(ttta)ₙ]/D15S103</td>
<td>11.70</td>
<td>−7.45</td>
<td>−0.279</td>
<td>0.604</td>
<td>0.975</td>
</tr>
<tr>
<td>D15S126</td>
<td>15.30</td>
<td>−5.18</td>
<td>0.386</td>
<td>0.321</td>
<td>0.941</td>
</tr>
</tbody>
</table>

Interlocus scores of the multipoint data are available on request. Maximum negative parametric LOD score is shown in bold. Non-parametric analysis provides no significant evidence for excess allele sharing across the whole region. There is zero genetic distance between the D15S103 and CYP19[(ttta)ₙ] markers. For abbreviations refer to legend of Table 1.

Table 3. Comparison of genotype distributions between different sub-groups in the case-control data set

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>p-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>216+</td>
<td>0.015</td>
<td>(total group)</td>
</tr>
<tr>
<td>216−</td>
<td></td>
<td>(excluding normal controls)</td>
</tr>
</tbody>
</table>

Analysis by quartiles of serum testosterone:

Top quartile 19 15 0.015
50th-75th percentile 24 7 0.03
25th-50th percentile 33 7
Bottom quartile 27 4

Analysis by clinical diagnosis:

PCOS 72 25 0.01
aPCO/Normal controls 95 15 0.03

Analysis by hirsutism:

Hirsute 25 14 0.01
Non-hirsute 51 11

Analysis by ovulatory status:

Anovulatory 66 22 0.08
Ovulatory 101 18 0.23

The ‘Top quartile’ includes women with serum testosterone levels of 2.90 nmol/l and greater, the ‘50th-75th percentile’ includes levels between 2.2 to 2.9 nmol/l, the ‘25th-50th percentile’ includes levels between 1.6 to 2.2 nmol/l and the ‘Bottom quartile’ includes those with levels of 1.60 nmol/l and lower, for the entire data-set (affecteds and controls). ‘216+’ genotype includes individuals with at least one 216 allele and ‘216−’ includes those with no 216 allele. χ² contingency tables were used to compare genotype distributions between the different groups. Genotype comparisons for the ‘Analysis by hirsutism’ and ‘Analysis by ovulatory status’ were carried out for the total group (genotype distributions given above), as well as after the exclusion of normal controls (p-values for both comparisons have been reported). For the ‘Analysis by hirsutism’, only individuals with a clear assignment as either hirsute or non-hirsute were included.

Any pathogenic mutations in CYP11a are expected to lead to upregulation; therefore we focused our mutation screening on the promoter region of CYP11a. This region is believed to contain multiple cAMP-regulated elements, responsible for increasing basal transcriptional activity, and elements involved in repression of expression (29). PCR amplification, followed by direct sequencing analysis, was used to screen 1.85 kb of sequence in ten affected and eight unaffected members of the pedigrees. No alterations to the published sequence were found in any individual apart from variation in the number of repeat units of the two microsatellite polymorphisms present in this region [a dinucleotide (ac)ₙ repeat, and a pentanucleotide (tttta)ₙ repeat, at positions −1376 and −528 respectively from the ATG start of translation site].

Individuals in the case-control study were typed for the CYP11a (tttta)ₙ marker. Four alleles were observed in our data set, alleles 216, 226, 236 and 241 with observed frequencies of 0.59, 0.28, 0.04 and 0.09 respectively. In order to reduce multiple allele comparisons for infrequent alleles, individuals were allocated to two groups according to their genotype, those having at least one copy of the common, four-repeat units allele (designated 216+) and those with no 216 allele (designated 216−). An association between this polymorphism and serum testosterone levels was found in the total data set (Mann-Whitney, p = 0.002) (Fig. 2). We also compared the genotype distributions of each quartile of serum testosterone level within the total data set. The proportion of individuals with 216− genotypes in each quartile were (from highest to lowest) 0.44, 0.23, 0.18 and 0.13 (p = 0.015; Table 3).

On subgroup analysis, the association between CYP11a genotype and testosterone was only evident in the women with PCOS (Mann-Whitney, p = 0.009; Fig. 2), there being no significant difference between the mean testosterone levels of 216− and 216+ individuals for either the normal controls or the aPCO groups. However, the numbers in these groups were small, raising the possibility of a type 2 error.

We next compared the CYP11a genotype distributions between subject groups (Table 3). Since the association between serum testosterone levels and CYP11a genotype was only seen in the PCOS group, taking the entire data-set (including those without testosterone levels available), we compared PCOS women with a combined normal control/aPCO group. There was a significant
Figure 2. Association of CYP11a with total serum testosterone levels in the case control data-set. The Mann-Whitney test was used to compare total serum testosterone levels between the two CYP11a genotype groups. P-values obtained for each test are given above. The ‘Total group’ included all the women in the data set. The ‘PCOS’ group are the women with polycystic ovaries and symptoms of anovulation and/or hirsutism. The ‘aPCO’ group are the asymptomatic women with polycystic ovaries and the ‘normal control’ group are the control subjects with normal ovarian morphology. ‘216+’ sub-group are individuals with at least one 216 allele and ‘216–’ are those with no 216 allele. ‘a’, ‘b’, ‘c’, ‘d’ and ‘e’ are the 10th, 25th, 50th, 75th and 90th percentiles of serum testosterone levels respectively. The median, mean and standard deviation (±SD) of serum testosterone levels for each group are also given above.

difference in the distribution of genotypes between these groups (p = 0.03).
When women were classified according to hirsutism (hirsute vs non-hirsute) we also found a marked difference in genotype distributions (p = 0.01 for the total group and p = 0.03 for women with polycystic ovaries). No such difference was seen when subjects were reclassified according to ovulatory status (p = 0.08 for the total group and p = 0.23 for women with polycystic ovaries).

DISCUSSION
Our data show that the gene coding for aromatase (CYP19) can be excluded as a major determinant of risk for PCOS/MPB. In contrast, the linkage results for CYP11a (encoding side chain cleavage) provide support for the involvement of this gene in the aetiology of PCOS/MPB. The association data demonstrate that allelic variants of CYP11a mediate the development of hyperandrogenaemia, which is in turn associated with PCOS and hirsutism. The association of this gene with hirsutism and the lack of association with ovulatory status indicates that CYP11a predominantly has a role in the development of hirsutism in PCOS.
We have shown a significant association between alleles of the CYP11a (tttta)n marker and serum testosterone levels. It is feasible that allelic variants of this rate-limiting enzyme could influence the level of androgen production. This association, however, can only be demonstrated in the PCOS group. The failure to demonstrate association in the aPCO and normal groups may reflect the small sample sizes. Alternatively it may be that this locus has a modifying effect, interacting with other genetic and/or environmental factors controlling the development of the ovarian changes and thereby influencing the clinical presentation of PCOS.
CYP11a (tttta)n is present in the promoter region of CYP11a and sequencing analysis of the surrounding regions does not reveal any other mutations likely to have direct pathogenic effects. However only 1.85 kb of sequence was screened and the putative mutation could be located outside this region. Recently a number of gene-associated VNTRs (variable number tandem repeats) have been implicated in the regulation of the genes, such as the insulin gene VNTR (30). Also, direct association between alleles of a (tttta)n repeat polymorphism, present in the 5′ flanking region of the apolipoprotein(a) gene, and plasma levels of Lp(a) has been demonstrated (31). Expression studies have shown a 5-fold higher transcriptional activity in constructs containing eight copies of the apolipoprotein(a) (tttta) repeat compared to those with nine copies (32).
As yet no regulatory function has been assigned to the CYP11a (tttta)n polymorphism and further investigation of the promoter region of CYP11a is required to determine the relationship between the (tttta) repeat polymorphism and the regulation of this gene.
We propose that allelic variants of CYP11a have a role in the aetiology of hyperandrogenaemia. This in turn may be sufficient to cause the altered ovarian morphology or may act in conjunction with other genetic factors to modify the phenotypic expression of PCOS.
Table 4. Summary of information on the PCR fragments S1, S2 and S3

<table>
<thead>
<tr>
<th>PCR fragment</th>
<th>Primer-set sequence</th>
<th>Size of product (bp)</th>
<th>Distance from ATG codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>cttctgagggaggaatgtgg</td>
<td>700 bp</td>
<td>–1174 to –1874</td>
</tr>
<tr>
<td>S2</td>
<td>ggtgcattagagctagggg</td>
<td>660 bp</td>
<td>–551 to –1211</td>
</tr>
<tr>
<td>S3</td>
<td>cttaactgetccacctgc</td>
<td>581 bp</td>
<td>–21 to –602</td>
</tr>
</tbody>
</table>

MATERIALS AND METHODS

Subjects

Twenty pedigrees (145 members) were selected from the Reproductive Medicine and Endocrinology clinics at the Samaritan and St. Mary’s Hospital as described previously (11). Probands presented with either menstrual disturbances and/or hirsutism and all had bilateral polycystic ovaries on ultrasound scan. Other causes of anovulation and hirsutism such as Cushing’s syndrome and late-onset congenital adrenal hyperplasia were excluded. All family members underwent a full screening procedure. Assignment of affection status in all women of reproductive age was based solely on the presence of bilateral polycystic ovaries on ultrasound scan. The only exception to this strict diagnostic scheme was the assignment of affection status in post menopausal women in whom ovarian imaging is not informative given reduced ovarian activity. In such women we assigned individuals as affected on the basis of a clear history of previous menstrual dysfunction and/or hirsutism: by this criteria, six (6) post menopausal women were assigned as affected. Pre-menarchal women and post-menopausal women, with a negative history, were assigned as unknown. Men were assigned as affected if they demonstrated significant premature male pattern baldness (MPB), defined as greater than a revised Hamilton IIa score (33), before the age of 40. If MPB was found to be present, age of onset was estimated using both historical photographic evidence and subjective assessment. Our previous definition of onset before the age of 30 years (11) has not been used because during the collection of further families we have observed that within certain families the age of onset of hair loss is between 30 and 40 years. This observation is generally conserved within a single pedigree. Thus to avoid the loss of power in those pedigrees we have used the Ferriman–Gallwey index (34). Women with oligomenorrhea (intermenstrual interval >6 weeks) and amenorrhea (intermenstrual interval >6 months) were designated anovulatory.

Sequencing analysis

Three overlapping PCR fragments were generated using primers designed from the published sequence of the 5’ region of CYP11a (35), spanning 1.85 kb of sequence upstream of the ATG start of translation site. The three fragments were designated S1, S2 and S3 (Table 4).

All polymerase chain reactions (PCR) of genomic DNA were carried out in a Hybaid OmniGene thermocycler. For each PCR reaction the 100 µl amplification mixture contained 50 ng of genomic DNA, 125 pmol of each primer, 1 x PCR reaction buffer (5 mM KCl, 1 mM Tris–HCl (pH 9), 0.01% Triton X-100 and 1 mM MgCl₂ for the S2 fragment amplification) or 1.5 mM MgCl₂ (for the S1 and S3 fragments), 100 µl each of dTTP/dCTP/dGTP/dATP and 1.5 U AmpliTaq DNA polymerase (Cambio, UK). The reaction conditions were: 35 cycles of 1 min denaturation at 94°C, 1 min annealing at 60°C (for S1 and S2), 63°C (for S3) and 1 min extension at 72°C. An initial denaturation step of 5 min at 94°C and a final extension of 10 min at 72°C were employed.

In order to obtain single stranded DNA for sequencing in both the forward and reverse directions, two PCR reactions (one with only the forward strand biotinylated and the other with the reverse strand biotinylated) were carried out for each of the three fragments. Each PCR product was purified using the Promega Wizard PCR purification kit (Promega, UK). Biotinylated single stranded DNA was obtained using the Dynabeads M-280 Streptavidin system (Dynal Ltd, UK). Approximately 5 ng of each single stranded product was sequenced using the Sequanase version 2.0 DNA sequencing kit (USB Corporation, USA).

DNA samples from ten affected and eight unaffected members of the pedigrees, showing positive evidence for linkage, were screened.

Marker typing

All members of the pedigrees were typed for the markers D15S118, D15S123, D15S125, CYP19 (tta₉₉, D15S126, D15S155, D15S153, D15S125, CYP11a (ac₉₉, D15S169 and D15S211, described elsewhere (36–38). Genotype analysis was carried out following PCR amplification. All members of the case control study were typed for the CYP11a (ttta₉₉) microsatellite marker, using the following primers designed from the published sequence: forward gggtgaaactgtgccattgc and reverse gtttgggggaaatgaggggc, using standard PCR conditions (39).
Data analysis

Multipoint parametric and non-parametric analyses were performed using the GENEHUNTER (27) programme. The non-parametric linkage score (NPL score) is a Z-score. For imperfect data (small number of pedigrees with different pedigree structures) the distribution of the Z is asymptotically normally distributed as the number of pedigrees is increased and as the information content approaches 100%. The significance level (p-value) assigned is conservative and becomes increasingly accurate as information content increases. This approach is referred to as the ‘perfect data approximation’.

Model parameters for the parametric analysis were: disease frequency 20%, disease gene frequency 10%, penetrance 95%, phenocopy rate 10%, phenocopy risk 2.5%. \( \chi^2 \) contingency table analyses were used to compare the CYP11a (tttta)\(_g\) genotype distributions in the case-control study. Mann-Whitney test was used to compare serum testosterone levels of the different groups in the case-control study.

ACKNOWLEDGEMENTS

We thank Dr Adam Carey for contribution to the family collection, Rosanne Joseph-Horne for assistance in collecting the data. We thank Dr Adam Carey for contribution to the family collection. Work at the Middlesex Hospital was funded by the British Heart Foundation, grant number PG/92138.

REFERENCES