Spring 2018 – Systems Biology of Reproduction
Discussion Outline – Assisted Reproduction/Contraception
Michael K. Skinner – Biol 475/575
CUE 418, 10:35-11:50 am, Tuesday & Thursday
April 26, 2018
Week 16

Assisted Reproduction/Contraception

Primary Papers:

1. Ayoub, et al. (2017) Andrology 5(2):278-285
2. Danshina, et al. (2016) Molec Hum Reprod doi:10.1093/molehr/gaw016
3. Winand, et al. (2014) Human Reproduction 29:842-851.

Discussion

Student 19: Reference 1 above

- What is the potential target for a male contraceptive?
- What reduction in hormones was observed and will this be a good contraceptive?
- What future studies are needed and anticipated limitations?

Student 20: Reference 2 above

- What technology and approach was used to examine develop a contraceptive?
- What contraceptive target was used and cell type?
- What limitation exists in the approach proposed?

Student 28: Reference 3 above

- What was the approach, technology and database resource?
- What genetic mutations were observed?
- What conclusions were made on the use of next generation sequencing on pre-implantation embryos for IVF?

ANDROLOGY

ISSN: 2047-2919

ANDROLOGY

ORIGINAL ARTICLE

Correspondence:
Stephanie T. Page, Box 356426, 1959 NE Pacific Street, Seattle WA 98195.
E-mail: page@u.washington.edu

*These authors contributed equally to this work.Clinical Trial Number: NCT01382069

Keywords:
androgen, dimethandrolone, male contraception, pharmacokinetics, suppression of gonadotropins

Received: 17-Jun-2016
Revised: 20-Sep-2016
Accepted: 20-Oct-2016

doi: 10.1111/andr.12303

Comparison of the single dose pharmacokinetics, pharmacodynamics, and safety of two novel oral formulations of dimethandrolone undecanoate (DMAU): a potential oral, male contraceptive

[1,*]R. Ayoub, [2,*]S. T. Page, [1]R. S. Swerdloff, [1]P. Y. Liu, [2]J. K. Amory, [1]A. Leung, [1]L. Hull, [3]D. Blithe, [3]A. Christy, [2]J. H. Chao, [2]W. J. Bremner and [1]C. Wang

[1]Department of Medicine, Division of Endocrinology, Harbor-UCLA Medical Center and Los Angeles Biomedical Research Institute, Torrance, CA, USA, [2]Department of Medicine, University of Washington, Seattle, WA, USA, and [3]Contraception Research Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

SUMMARY

Dimethandrolone (DMA, 7α,11β-dimethyl-19-nortestosterone) has both androgenic and progestational activities, ideal properties for a male hormonal contraceptive. In vivo, dimethandrolone undecanoate (DMAU) is hydrolyzed to DMA. We showed previously that single oral doses of DMAU powder in capsule taken with food are well tolerated and effective at suppressing both LH and testosterone (T), but absorption was low. We compared the pharmacokinetics and pharmacodynamics of two new formulations of DMAU, in castor oil and in self-emulsifying drug delivery systems (SEDDS), with the previously tested powder formulation. DMAU was dosed orally in healthy adult male volunteers at two academic medical centers. For each formulation tested in this double-blind, placebo-controlled study, 10 men received single, escalating, oral doses of DMAU (100, 200, and 400 mg) and two subjects received placebo. All doses were evaluated for both fasting and with a high fat meal. All three formulations were well tolerated without clinically significant changes in vital signs, blood counts, or serum chemistries. For all formulations, DMA and DMAU showed higher maximum ($p < 0.007$) and average concentrations ($p < 0.002$) at the 400 mg dose, compared with the 200 mg dose. The powder formulation resulted in a lower conversion of DMAU to DMA ($p = 0.027$) compared with both castor oil and SEDDS formulations. DMAU in SEDDS given fasting resulted in higher serum DMA and DMAU concentrations compared to the other two formulations. Serum LH and sex hormone concentrations were suppressed by all formulations of 200 and 400 mg DMAU when administered with food, but only the SEDDS formulation was effectively suppressed serum T when given fasting. We conclude that while all three formulations of oral DMAU are effective and well tolerated when administered with food, DMAU in oil and SEDDS increased conversion to DMA, and SEDDS may have some effectiveness when given fasting. These properties might be advantageous for the application of DMAU as a male contraceptive.

INTRODUCTION

Current methods of male contraception include condoms and vasectomy, both of which have drawbacks. While condoms are reversible and widely available, they have a high user failure rate. Vasectomies are efficacious, but invasive and not readily reversible. Therefore, efforts are underway to develop alternative male contraceptives with agents of known mechanisms of action. Male hormonal contraception uses exogenous sex steroids to suppress gonadotropin secretion and spermatogenesis, is reversible, and might provide additional health benefits for men

© 2016 American Society of Andrology and European Academy of Andrology

if optimally designed (Liu *et al.*, 2006; Page *et al.*, 2008; Nieschlag, 2010; Piotrowska *et al.*, 2016; Wang *et al.*, 2016). Contraceptive efficacy studies in men with weekly intramuscular (IM) injections of testosterone enanthate or monthly injections of testosterone undecanoate have been encouraging, with high efficacy rates and few side effects (World Health Organization Task Force on Methods for the Regulation of Male Fertility, 1990; World Health Organization Task Force on the Regulation of Male Fertility, 1996; Gu *et al.*, 2009). However, suppression of spermatogenesis with exogenous testosterone alone is not uniform across all ethnic groups and requires supraphysiologic dosing. By combining testosterone with a progestin, suppression of spermatogenesis is enhanced (Meriggiola & Bremner, 1997; Liu *et al.*, 2008; Page *et al.*, 2008; Wang & Swerdloff, 2010). For example, the combination of testosterone implants and long-acting injections of the progestin depo-medroxyprogesterone acetate has excellent contraceptive efficacy (Turner *et al.*, 2003). However, injections and implants may be less desirable for some men than an oral medication such as DMAU. Oral contraceptives that are user-controlled, easy to administer, and have shorter 'on and off' rates, might be desirable for many couples (Liu *et al.*, 2006; Glasier, 2010).

Dimethandrolone (DMA) is a novel derivative of 19-nortestosterone that binds to both the androgen and the progesterone receptors, making it an attractive candidate as a single-agent male contraceptive (Attardi *et al.*, 2006). DMA undecanoate (DMAU), includes a long carbon chain ester at the C-17 position (Cook & Kepler, 2005). DMAU has been shown to be an effective, reversible contraceptive when dosed orally in pre-clinical studies. DMAU is hydrolyzed to DMA in vivo where it effectively decreases fertility in rabbits without appreciable toxicity, and is similarly non-toxic in rats and monkeys (Hild *et al.*, 2010; Attardi *et al.*, 2011a,b). We previously reported the pharmacokinetics, safety, and tolerability of DMAU when given as a single oral dose as a powder in capsule to healthy male volunteers. However, we noted that this formulation resulted in low serum concentrations of DMA, likely because of the poor conversion of DMAU to DMA (about 3%) and that concomitant administration with food was required for both appreciable absorption and for conversion to DMA (Surampudi *et al.*, 2014). In an effort to increase the bioavailability of DMA, we reformulated oral DMAU into capsules in either castor oil or self-emulsifying drug delivery systems (SEDDS) and assessed their pharmacokinetics and pharmacodynamics when administered orally to healthy men. We hypothesized that utilization of these lipophilic drug delivery entities would enhance absorption and hydrolysis of DMAU, might negate the need for concomitant administration with food, and would optimize DMA serum concentrations for future evaluations of oral DMAU as a male hormonal contraceptive.

RESEARCH PARTICIPANTS AND METHODS

Research participants

Healthy men, age 18–50 years, with no significant medical history or illnesses, and normal physical examination, blood count, clinical chemistries, hepatitis panel, liver function tests, prostate-specific antigen (PSA) levels, electrocardiogram, and BMI <33 kg/m^2, were included in the study. Men were excluded if they had participated in a clinical trial involving an investigational drug within 30 days, had used hormonal therapy within the last 3 months, had a disorder of the hypothalamus/pituitary/testis, desired fertility within a year, or had a pregnant partner, or had clinically significant abnormal physical or laboratory findings, or an elevated PSA. Participants were recruited and enrolled at the Harbor-UCLA Medical Center/Los Angeles Biomedical Research Institute in Torrance, California and the University of Washington in Seattle, Washington. The study protocol was approved by the institutional review boards for both participating institutions. All participants provided written informed consent prior to any study procedures. The medical monitor and the investigators reviewed adverse events and safety data weekly with the provision that an external independent data safety monitoring board be notified if and when a grade 3 adverse event occurred.

Study medications

DMAU is manufactured by Evestra Inc (San Antonio, TX, USA). For the powder in capsule formulation, DMAU was micronized by Micron Technologies, Inc. (Malvern, PA, USA) and encapsulated as 25 or 200 mg DMAU powder and placebo in capsules (Pharmtek Laboratories, Inc., San Diego, CA, USA) in a cGMP environment. SRI International (Menlo Park, CA, USA) manufactured 100 mg DMAU in castor oil/benzyl benzoate (70 : 30 volume : volume) and 50 mg DMAU in SEDDS capsules and corresponding placebo under Good Manufacturing Practice standards.

Study design

All participants were assessed by a study physician to ensure that all inclusion and exclusion criteria were met. For all three formulations, at each dose level evaluated, 10 men received active drug and two received placebo in a double-blind fashion. For the castor oil and SEDDS formulations, 10 participants received 100, 200, and 400 mg DMAU (for) both fasting and after a high fat meal, and two men received identical placebo capsules. One subject randomized to the DMAU in SEDDS group had no measurable DMA levels on two pharmacokinetics sampling days and his data were not included in the study (despite approval for additional procedures by the UCLA-IRB, he declined re-dosing). For the DMAU powder in capsules, participants were administered 100, 200, and 400 mg DMAU or identical placebo with a high fat meal as previously described (Surampudi *et al.*, 2014).

All participants were admitted to the clinical research unit within the Clinical and Translational Science Institute at each site and were observed for 24 h with hourly vital signs monitoring following dosing. An electrocardiogram was performed 4–6 and 24 h after drug administration. Safety laboratory tests (clinical chemistry panel, liver function tests) were measured at baseline and 24 h after each DMAU dose. Fasting lipids and complete blood counts were quantified at baseline, and approximately, 7 days following each dose. Serum hormones [T, free T, estradiol, dihydrotestosterone (DHT), LH, FSH, and sex hormone-binding globulin (SHBG)], were measured before drug administration (time zero) and either every 4 (castor oil and SEDDS formulations) or 12 (powder formulation) h post-administration. In all cases, DMA and DMAU concentrations were quantified in the blood drawn -0.5 before, 0, and 0.5, 1, 2, 4, 6, 8, 12, 18, and 24 h after oral administration of DMAU. The participants returned to the clinic at least 7 days after the dose of

© 2016 American Society of Andrology and European Academy of Andrology

DMAU for safety laboratory tests, adverse event reporting, vital signs, and hormone evaluation. Halting parameters for predefined safety criteria and adverse events were included in the protocol, but these parameters were never reached.

Analytical methods

Safety laboratory tests and lipid panels were performed at each institution's respective clinical laboratory. All hormones were quantified by the licensed Endocrine and Metabolic Research Laboratory at Harbor-UCLA/LA Biomed using validated methods. Serum T, DHT, estradiol, DMAU, and DMA were measured by liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Shiraishi *et al.*, 2008; Rothman *et al.*, 2011; Surampudi *et al.*, 2014) and free T was calculated using a standard formula (Vermeulen *et al.*, 1999). Serum LH, FSH, and SHBG were measured using sensitive fluoroimmunometric assays as previously described (Swerdloff *et al.*, 2000).

Statistical analyses

The primary endpoints of the trial were safety and tolerability of the three formulations of DMAU. Secondary endpoints included the 24-h PK of DMAU and DMA after oral dosing fasting and with food, as well as suppression of gonadotropins and testosterone production.

The number of participants for this Phase 1 clinical study was powered to provide at least a 0.80 probability to exclude at least 20% of participants developing Grade 3 adverse events assuming a two-sided 95% confidence interval with each dose and formulation given. The PK parameters for each full sampling day for DMAU were determined by non-compartmental methods and were primarily assessed using the area under the curve from 0–24 h (AUC_{0-24}) of serum DMAU/DMA levels generated by the 10 blood sampling times over 24 h for each dose of DMAU and computed using the trapezoid method. Other PK parameters assessed included C_{avg} (average concentration over 24 h), C_{max} (maximum concentration over 24 h), C_{min} (minimum concentration over 24 h), and T_{max} (time to reach C_{max}). The elimination half-life, $T_{1/2}$, was calculated assuming exponential decay when there were at least three measurable concentrations after C_{max}.

As a dose effect is not anticipated for zero dose, the zero dose was removed from the analysis. Also, as we anticipated a marked effect of food, we planned separate analyses under the fed and fasting conditions a priori, assuming this was verified. Mixed models incorporated repeated measurements within subjects using a compound symmetrical covariance structure were constructed to examine the effect of dose (0, 200, and 400 mg), formulation (powder, SEDDS, castor oil), and the interaction on AUC, C_{avg}, C_{max}, T_{max}, and $T_{1/2}$ for serum DMA, serum DMAU and the ratio of DMA to DMAU. Post hoc Bonferroni-adjusted testing was performed only when a significant main effect was detected. All analyses were performed using SAS version 9.3 (SAS Institute Inc., Cary, NC, USA), with two-tailed $p < 0.05$ construing statistical significance. Data are presented as mean ± SEM.

The effect of oral DMAU treatment on serum T, DHT, estradiol (E2), LH, FSH, and calculated free T was analyzed by mixed model analogously for C_{avg} and C_{min}. Models here were constructed separately using three levels of formulation (powder, SEDDS, castor oil) as well as two levels of formulation (SEDDS and castor oil), as serum concentrations for hormones were measured less frequently when the powder formulation was administered than when SEDDS and castor oil formulations were dosed. Analyses under both models yielded congruent findings, and hence, we present analyses that include all three formulations, unless otherwise indicated.

RESULTS

Research participants' demographics, disposition, and safety

There were a total of 44 participants between the two study sites, 19 for powder in capsule, 12 for castor oil, and 13 for SEDDS. A total of eight men discontinued during the dosing periods and were replaced to ensure that for all doses, 12 participants were evaluated for both fasting and fed; seven discontinued dosing during the evaluation of powder in capsule and one from SEDDS dosing. In all cases, discontinuation was because of scheduling or personal reasons and not because of adverse effects. Demographics for the participants in each phase of the study are shown in Table 1. Across all three groups, men had an average age of 33 years and BMI of 25.

All three formulations were well tolerated and there were no serious adverse events. One participant had an AST >twofold the upper limit of normal thought to be related to binge alcohol intake; this elevation resolved without treatment. There were no significant changes in chemistry or lipid panels and hematocrit was not significantly different between baseline and the end of the study. EKGs and QTc intervals in all participants were not significantly different from baseline and there were no clinically significant changes in vital signs in any of the participants (data not shown).

Food effects

There were marked food effects on DMAU and DMA pharmacokinetics in all three formulations. When DMAU was administered with a high fat meal (50% calories as fat), DMAU absorption and DMA serum concentrations were increased leading to a significantly higher AUC, C_{avg}, and C_{max} for all formulations compared to administration fasting (Fig. 1, $p < 0.001$ in all cases). Hence, the data were analyzed separately for the fasting and fed conditions.

Pharmacokinetics of DMA and DMAU when DMAU is given with food

DMA and DMAU showed higher AUC, C_{avg}, and C_{max} ($p \leq 0.001$ for each) and DMA/DMAU AUC ratio ($p = 0.044$) at the 400 mg dose, compared with the 200 mg dose (2). This dose effect was true for all three formulations. Comparing the three formulations, the powder in capsule resulted in higher AUC and C_{max} for DMAU than dosing in castor oil ($p = 0.007$ and $p = 0.029$, respectively). Despite these higher serum concentrations of DMAU, administration in powder did not increase serum DMA concentrations compared to the other two formulations. In fact, powder in capsule resulted in the lowest proportion of DMAU conversion to DMA (DMA/DMAU AUC ratio) compared with either of the other two formulations ($p < 0.02$ for powder in capsule compared to both castor oil and SEDDS, Table 2).

There was no dose effect on the time to maximum concentration, $T_{max,,}$ for either DMAU or DMA. However, T_{max} for both

 © 2016 American Society of Andrology and European Academy of Andrology

Table 1 Characteristics of the participants (±standard error of the mean)

	Total (44)	Powder (19)	Castor oil (12)	SEDDS (13)
Race				
American Indian or Alaska Native	2 (4.5%)	2 (10.5%)	0	0
Asian	1 (2.3%)	0	0	1 (7.7%)
Native Hawaiian or other Pacific Islander	1 (2.3%)	1 (5.3%)	0	0
Black or African American	5 (11.4%)	2 (10.5%)	3 (25%)	0
White	33 (75%)	12 (63.2%)	9 (75%)	12 (92.3%)
Other	2 (4.5%)	2 (10.5%)	0	0
BMI (kg/m^2)	25.4 ± 0.5	25.9 ± 0.71	25.2 ± 1.1	24.9 ± 1.1
Age (years)	33.2 ± 1.4	31.7 ± 2.3	36.4 ± 2.6	32.4 ± 2.4
Screening T (ng/dL)	546 ± 29	546 ± 43	481 ± 63	610 ± 45

Figure 1 Serum DMA (upper panel) and DMAU (lower panel) concentrations after oral administration of DMAU in three formulations as a single dose 0, 100, 200, or 400 mg after fasting overnight (left panels) or a high fat meal (50% fat) (right panels). Note, *y*-axis is log scale. (Conversion DMA 1 ng/mL = 3.29 nmol/L and DMAU 1 ng/mL = 2.12 nmol/L).

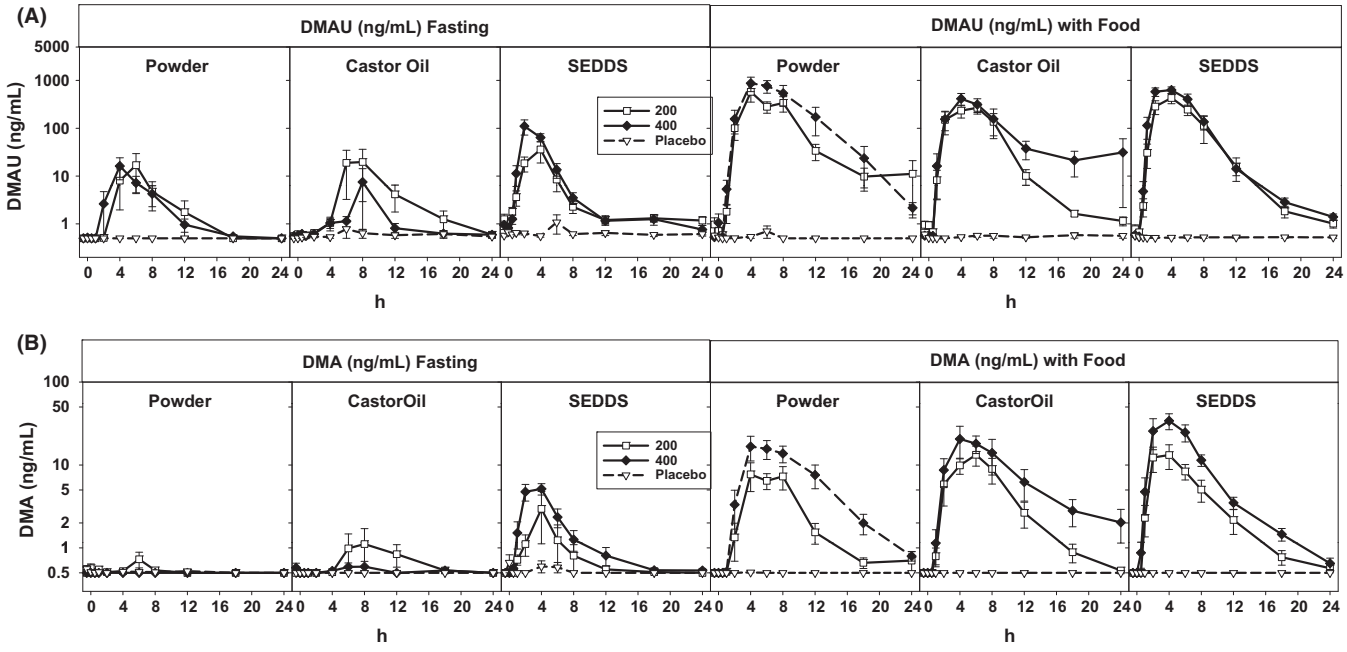


Table 2 Comparison of PK parameters after single oral dose of DMAU in three formulations (±standard error of the mean).

Fed DMAU	200 mg SEDDS (n = 10)	Castor oil (n = 10)	Powder in capsule (n = 10)	400 mg SEDDS (n = 9)	Castor oil (n = 10)	Powder in capsule (n = 10)
AUC (ng/mL/24 h)	2233 ± 340	1706 ± 247	3149 ± 470	3534 ± 289	2583 ± 400	6153 ± 1143
C_{avg} (ng/mL)	93.0 ± 14.1	71.1 ± 10.3	131.2 ± 19.6	147.3 ± 12.1	107.6 ± 16.7	256.4 ± 47.6
C_{max} (ng/mL)	599 ± 90	485 ± 59	854 ± 194	889 ± 50	665 ± 109	1410 ± 283
T_{max} (h)	3.8 ± 0.6	4.6 ± 0.7	5.8 ± 0.6	3.3 ± 0.6	6.4 ± 2.0	6.2 ± 1.1
HalfLife(h)	2.0 ± 0.63	1.94 ± 0.87	1.72 ± 0.57	2.13 ± 0.71	2.30 ± 0.76	1.83 ± 0.61
DMA						
	(n = 10)	(n = 10)	(n = 10)	(n = 9)	(n = 10)	(n = 10)
AUC (ng/mL/24 h)	96.0 ± 16.3	103.1 ± 16.3	66.7 ± 12.0	222.5 ± 28.1	187.6 ± 35.3	163.9 ± 23.2
C_{avg} (ng/mL)	4.0 ± 0.7	4.3 ± 0.7	2.8 ± 0.5	9.3 ± 1.2	7.8 ± 1.5	6.8 ± 1.0
C_{max} (ng/mL)	19.2 ± 4.7	19.2 ± 3.3	12.5 ± 2.8	47.9 ± 8.1	36.1 ± 8.2	27.7 ± 4.3
T_{max} (h)	3.8 ± 0.8	4.8 ± 0.5	5.8 ± 0.5	3.8 ± 0.5	5.6 ± 0.9	6.4 ± 1.0
HalfLife(h)	2.84 ± 1.0	2.68 ± 0.85	2.44 ± 0.92	3.23 ± 1.07	3.10 ± 1.03	3.31 ± 1.10
AUC ratio						
DMA/DMAU AUC	0.050 ± 0.009	0.063 ± 0.007	0.028 ± 0.006	0.112 ± 0.051	0.075 ± 0.008	0.031 ± 0.004

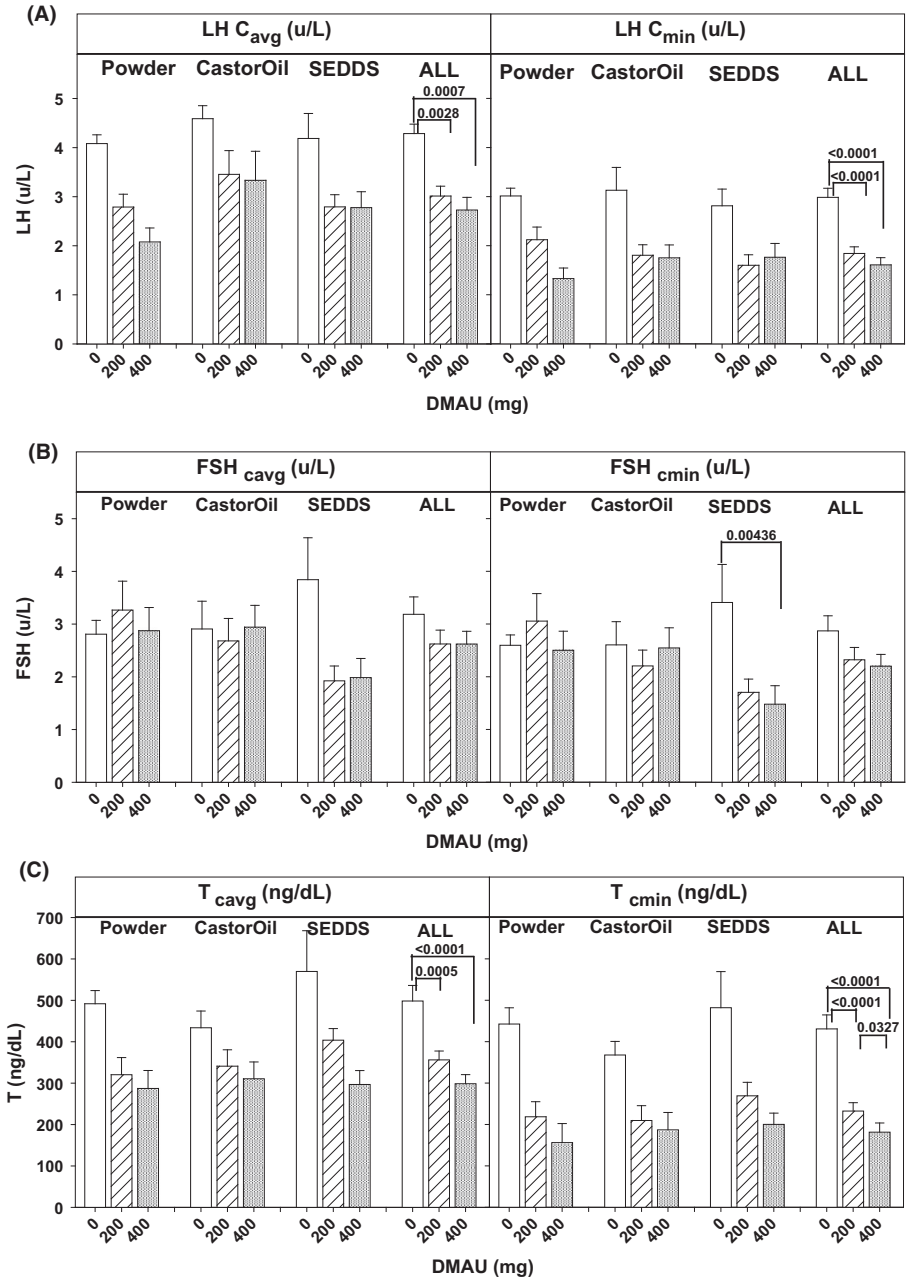
© 2016 American Society of Andrology and European Academy of Andrology

serum concentrations of DMA and DMAU were significantly delayed with powder in capsule compared with SEDDS ($p < 0.05$ for both DMAU and DMA). In contrast, there was an effect of dose on the elimination half-lives of DMA and DMAU ($p < 0.05$ for both, Table 2); importantly, however, the elimination half-lives for both DMA and DMAU were not significantly affected by formulation.

Pharmacodynamic effects of DMAU given with food

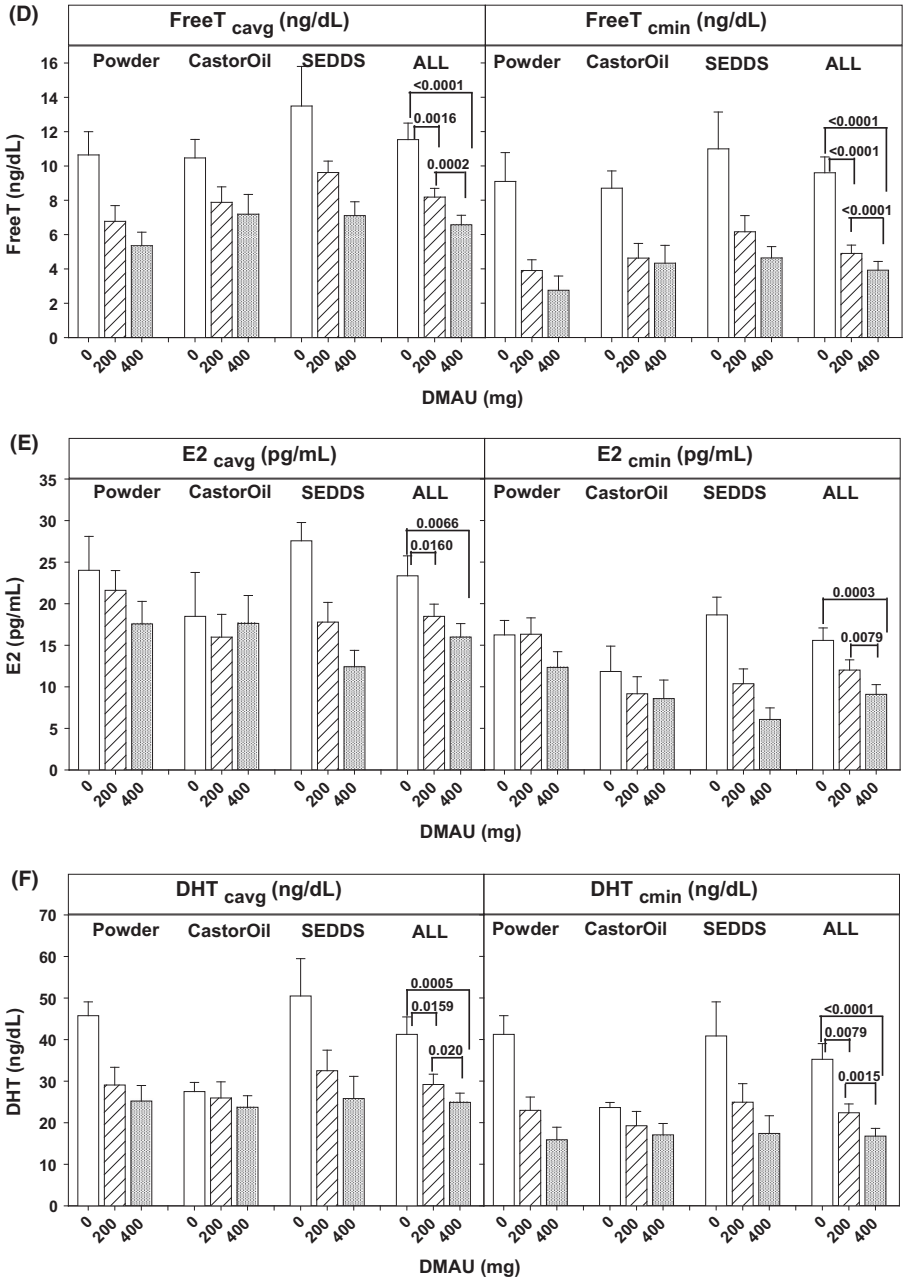
Significant dose effects were detected in C_{avg} and C_{min} for LH, T, free T, DHT, and E2 ($p < 0.01$ in all cases). Significant post hoc Bonferrroni-adjusted differences are illustrated in Fig. 2A–E, and consistently show that 400 mg was more suppressive than placebo, both in overall suppression (C_{avg}) and the minimum concentration achieved (C_{min}) (Table 2). A total quantity of 400 mg achieved a greater maximal suppression than 200 mg for steroids T, DHT, and E2: see Fig. 2C–E which shows the C_{min} for each dose and formulation. There were no significant effects of the formulation on LH or any of the sex steroids examined. In contrast, there were no significant dose effects on FSH C_{avg}, but there was a significant dose by drug interaction for FSH C_{min} ($p < 0.05$, Fig. 2B). Post hoc testing indicated that the 400 mg of SEDDS DMAU significantly suppressed FSH compared with SEDDS placebo ($p = 0.004$), but no other significant differences were detected.

Figure 2 Serum average (C_{avg}) and minimum (C_{min}) concentrations over 24 h of LH (A), FSH (B), T (C), free T (D), estradiol (E), and DHT (F) after administration of a single dose 0 (placebo), 200, 400, mg of DMAU with a high fat meal. For simplicity, the statistical differences between doses were marked by brackets only in the combined responses of gonadotropins or sex steroids of all three formulations (ALL).



 © 2016 American Society of Andrology and European Academy of Andrology

Figure 2 Continued



Pharmacokinetics and pharmacodynamics of DMAU administered when fasting

When administered fasting, oral DMAU in the SEDDS formulation resulted in higher blood concentrations of DMAU and DMAU than the other two formulations (Fig. 1). When the 400 mg dose was compared, these differences were significant, with SEDDS DMAU resulting in higher C_{avg} for both DMA and DMAU than the castor oil and powder in capsule formulations ($p < 0.005$ for each comparison for formulation by dose interaction, Bonferroni-adjusted post hoc $p < 0.03$ for each comparison).

As only negligible concentrations of DMA were achieved when DMAU was administered in powder or castor oil, their pharmacodynamics effects were not further evaluated. We investigated whether DMAU given fasting in SEDDS decreased serum LH, FSH, or T in a dose-dependent manner. When given fasting, DMAU in SEDDS resulted in a dose-dependent decrease in T C_{avg} ($p = 0.0005$) and C_{min} ($p = 0.015$), and FSH C_{min} ($p < 0.0001$) without significantly impacting LH (data not shown).

DISCUSSION

In this study, we compared the pharmacokinetics of DMAU and DMA after oral administration of single, escalating doses of three formulations of DMAU. These studies are consistent with our earlier observation that administration of DMAU with food, even when formulated in oil, markedly increases absorption (Surampudi *et al.*, 2014), even when formulated in oil, and result

© 2016 American Society of Andrology and European Academy of Andrology

in dose incremental increases in DMAU and DMA levels. Although administering DMAU in oil did not negate the significant enhancing effect of co-administration of food on DMAU absorption, oil-based formulations improved the conversion of DMAU to DMA compared to powder in capsule. Moreover, these newer formulations of DMAU dynamically suppress LH, T, and the downstream metabolites of T, DHT, and E2, when given with food. In contrast to the powder and castor oil formulations, under fasting conditions, administration of DMAU in SEDDS provided sufficient DMA to suppress T and FSH, even after a single dose, despite resulting in markedly lower DMAU and DMA concentrations than when given with food.

This study builds upon our previous work, further demonstrating that a single oral dose of DMAU of 200 or 400 mg with food significantly suppressed serum LH, T, free T, and its metabolites E_2 and DHT compared with placebo. The suppressive effect of all three formulations on sex steroids was dose dependent. The suppression of serum FSH was only evident when DMAU was given at the highest dose, 400 mg, with the SEDDS preparation, which also resulted in the highest DMA AUC (Fig 2B and Table 2). This is consistent with our prior observation that a single dose of DMAU as powder in capsule is a very effective suppressor of production of T, and suggests that oral DMAU may, when given repeatedly over time, be a potent suppressor of spermatogenesis (Surampudi *et al.*, 2014). The very rapid suppression of LH and endogenous T in response to oral DMAU may be because of the dual action of DMA on both the androgen and progesterone receptors (Attardi *et al.*, 2006). Combinations of androgens and progestins are more effective suppressors of spermatogenesis than T alone in male contraceptive clinical trials (Liu *et al.*, 2008).

We were somewhat surprised that DMAU had greater oral bioavailability when given as a powder than the other two formulations when given in the fed state (Fig. 1 and Table 2), achieving greater serum DMAU C_{avg} than the castor oil formulation. Testosterone and its short chain esters undergo rapid first-pass hepatic metabolism, limiting oral bioavailability (Tauber *et al.*, 1986), whereas testosterone undecanoate (TU) in castor oil has enhanced lipophilicity with absorption occurring via the intestinal lymphatics when TU is given with food (Shackleford *et al.*, 2003). Presumably, DMAU is also lymphatically absorbed; thus, we expected that oil emulsions would enhance absorption via this route, but this was not evident in this study. However, there was no effect of formulation on the AUC for DMA, the active metabolite of DMAU, because of the enhanced conversion of DMAU to DMA in vivo when given in either emulsified/oil formulation compared to powder (Table 2). How administration of DMAU in oil enhances de-esterification is not clear; however, given that only 5–10% of the DMAU is metabolized to DMA systemically, the improved DMA/DMAU ratio achieved when DMAU is given in oil is likely to be a significant advantage in multiple dose studies, allowing for markedly lower amounts of DMAU to be administered to achieve equivalent pharmacodynamic effects when DMAU is given in oil/SEDDS vs. powder. This hypothesis remains to be tested in repeat dose studies as we did not observe an effect of formulation in this single dose study on the degree or extent of LH or endogenous steroid production.

Although the effects were modest, we did observe a significant effect of formulation on DMA, LH, and endogenous steroid production when DMAU was administered fasting. In particular, the SEDDS formulation was superior to both powder in capsule and the castor oil formulation in achieving significant DMA and DMAU concentrations. SEDDS has also been shown to enhance the absorption of TU (Yin *et al.*, 2012). While the AUC for DMAU and DMA when DMAU is administered in SEDDS is still vastly lower than when given with food, by roughly an order of magnitude (Fig. 1), these low levels of serum DMA achieved when DMAU is given in SEDDS may be important in longer term, real use studies. As a potential contraceptive, the levels of DMA achieved with DMAU-SEDDS occasionally dosed without concomitant food, may be sufficient for maintaining gonadotropin suppression, and perhaps inhibition of spermatogenesis, in long-term daily users.

There were very few adverse events that were ascribed to DMAU. Importantly, as the studies presented here include only single doses, and involve multiple blood draws, androgenic effects such as stimulation of erythropoiesis, suppression of sex hormone-binding globulin, and potential reductions in high-density lipoprotein cholesterol concentrations could not be adequately assessed in this study. Longer term, repeat dose studies are required to further evaluate the safety of DMAU in men. To assess the effects of DMAU on spermatogenesis as well non-gonadal, androgen-sensitive organs including the prostate, bone, and muscle, longer term DMAU administration studies will be necessary. While DMAU has both androgenic and progestational activity in vitro and in pre-clinical rodent models, the long-term impact of LH and testosterone suppression on these hormonally sensitive tissues remains to be assessed and will be vital in developing DMAU as a male hormonal contraceptive.

In summary, single escalating doses of DMAU in three formulations: powder in capsule, castor oil, and SEDDS, up to 400 mg, were well tolerated in healthy male volunteers. When a single oral dose of DMAU was administered with a high fat meal, serum DMAU and DMA showed dose incremental increases sufficient to reversibly suppress LH and endogenous sex hormone production with all three formulations. Administration of DMAU in castor oil or SEDDS resulted in enhanced conversion of DMAU to DMA in vivo, which might be an advantage further development of these formulations over the powder in capsule. Furthermore, DMAU given in SEDDS was superior to the other two formulations when given in the fasting state, resulting in higher serum DMA concentrations sufficient for suppression of T and FSH. Further development of oral DMAU is ongoing with the goal of assessing its safety and efficacy in suppressing endogenous gonadotropins, sex steroids, and, in the long run, sperm production. These studies demonstrate that the formulation of DMAU in oil may have some advantages over powder in capsule; whether these observations hold true with multiple, repeat dosing remains to be evaluated. DMAU holds promise as a potential single agent, reversible, male hormonal contraceptive.

ACKNOWLEDGEMENTS

The Los Angeles Center was supported by the Eunice Kennedy Shriver National Institute of Child Health and Development (NICHD) Contraceptive Clinical Trial Network Contract HHSN27520130024I Task Orders HHSN 27500001, 27500002, 27500006; the Endocrinology, Metabolism and Nutrition training Grant (T32 DK007571), and the UCLA Clinical and Translational Science Institute (UL1TR000124) at Harbor-UCLA/LA BioMed.

 © 2016 American Society of Andrology and European Academy of Andrology

The Seattle Center was supported by NICHD Contraceptive Clinical Trial Network Contract HHSN275201100075U Task Orders HHSN 27500001, and 27500002, the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000423, the Center for Research in Reproduction and Contraception U54 HD 04245 (NICHD), and the National Institute of Diabetes and Digestive and Kidney Diseases training grant 5T32 DK007247-35. We thank Peter Christenson, PhD, formerly at LA Biomed for his advice in study design and data analysis plans. We thank our research coordinators Xiao-Dan Han, Elizabeth Ruiz, Kathryn Torrez-Duncan for their assistance with the study and the technical assistance of Feng Bai and Andrew Leung for DMA and DMAU assay development and validation. Finally, we thank our research volunteers and the staff of the Endocrine and Metabolic Research Laboratory at Harbor-UCLA/LA Biomed and the University of Washington Center for Research in Reproduction and Contraception.

DISCLOSURES

CW received research funding from Clarus Therapeutics, Lipocine, and Besins Healthcare. She is a temporary consultant to TesoRx and Lipocine. JKA has received research funding from Clarus Therapeutics. RSS received research funding from Clarus Therapeutics, Lipocine, Antares, and Aeterna Zentaris, Inc. He has served as a temporary consultant for Clarus Therapeutics, Novartis, TesoRex, Antares, and Aeterna Zentaris, Inc.

AUTHORS' CONTRIBUTIONS

STP, RSS, JKA, DB, AC, WJB, and CW designed the research study, analyzed the data, and wrote the paper. RA, JHC, AL, LH conducted the research study and analyzed the samples and reviewed and revised the paper.

REFERENCES

Attardi BJ, Hild SA & Reel JR. (2006) Dimethandrolone undecanoate: a new potent orally active androgen with progestational activity. *Endocrinology* 147, 3016–3026.

Attardi BJ, Engbring JA, Gropp D & Hild SA. (2011a) Development of dimethandrolone 17beta-undecanoate (DMAU) as an oral male hormonal contraceptive: induction of infertility and recovery of fertility in adult male rabbits. *J Androl* 32, 530–540.

Attardi BJ, Marck BT, Matsumoto AM, Koduri S & Hild SA. (2011b) Long-term effects of dimethandrolone 17beta-undecanoate and 11beta-methyl-19-nortestosterone 17beta-dodecylcarbonate on body composition, bone mineral density, serum gonadotropins, and androgenic/anabolic activity in castrated male rats. *J Androl* 32, 183–192.

Cook CE & Kepler JA. (2005) 7alpha,11beta-dimethyl-19-nortestosterone: a potent and selective androgen response modulator with prostate-sparing properties. *Bioorg Med Chem Lett* 15, 1213–1216.

Glasier A. (2010) Acceptability of contraception for men: a review. *Contraception* 82, 453–456.

Gu Y, Liang X, Wu W, Liu M, Song S, Cheng L, Bo L, Xiong C, Wang X, Liu X, Peng L & Yao K. (2009) Multicenter contraceptive efficacy trial of injectable testosterone undecanoate in Chinese men. *J Clin Endocrinol Metab* 94, 1910–1915.

Hild SA, Attardi BJ, Koduri S, Till BA & Reel JR. (2010) Effects of synthetic androgens on liver function using the rabbit as a model. *J Androl* 31, 472–481.

Liu PY, Swerdloff RS, Christenson PD, Handelsman DJ, Wang C & Hormonal Male Contraception Summit, G. (2006) Rate, extent, and modifiers of spermatogenic recovery after hormonal male contraception: an integrated analysis. *Lancet* 367, 1412–1420.

Liu PY, Swerdloff RS, Anawalt BD, Anderson RA, Bremner WJ, Elliesen J, Gu YQ, Kersemaekers WM, McLachlan RI, Meriggiola MC, Nieschlag E, Sitruk-Ware R, Vogelsong K, Wang XH, Wu FC, Zitzmann M, Handelsman DJ & Wang C. (2008) Determinants of the rate and extent of spermatogenic suppression during hormonal male contraception: an integrated analysis. *J Clin Endocrinol Metab* 93, 1774–1783.

Meriggiola MC & Bremner WJ. (1997) Progestin-androgen combination regimens for male contraception. *J Androl* 18, 240–244.

Nieschlag E. (2010) Clinical trials in male hormonal contraception. *Contraception* 82, 457–470.

Page ST, Amory JK & Bremner WJ. (2008) Advances in male contraception. *Endocr Rev* 29, 465–493.

Piotrowska K, Wang C, Swerdloff RS & Liu PY (2016) Male hormonal contraception: hope and promise. Lancet Diabetes Endocrinol 16, 34–36. *Epub* February 22, 2016.

Rothman MS, Carlson NE, Xu M, Wang C, Swerdloff R, Lee P, Goh VH, Ridgway EC & Wierman ME. (2011) Reexamination of testosterone, dihydrotestosterone, estradiol and estrone levels across the menstrual cycle and in postmenopausal women measured by liquid chromatography-tandem mass spectrometry. *Steroids* 76, 177–182.

Shackleford DM, Faassen WA, Houwing N, Lass H, Edwards GA, Porter CJ & Charman WN. (2003) Contribution of lymphatically transported testosterone undecanoate to the systemic exposure of testosterone after oral administration of two andriol formulations in conscious lymph duct-cannulated dogs. *J Pharmacol Exp Ther* 306, 925–933.

Shiraishi S, Lee PW, Leung A, Goh VH, Swerdloff RS & Wang C. (2008) Simultaneous measurement of serum testosterone and dihydrotestosterone by liquid chromatography-tandem mass spectrometry. *Clin Chem* 54, 1855–1863.

Surampudi P, Page ST, Swerdloff RS, Nya-Ngatchou JJ, Liu PY, Amory JK, Leung A, Hull L, Blithe DL, Woo J, Bremner WJ & Wang C. (2014) Single, escalating dose pharmacokinetics, safety and food effects of a new oral androgen dimethandrolone undecanoate in man: a prototype oral male hormonal contraceptive. *Andrology* 2, 579–587.

Swerdloff RS, Wang C, Cunningham G, Dobs A, Iranmanesh A, Matsumoto AM, Snyder PJ, Weber T, Longstreth J & Berman N. (2000) Long-term pharmacokinetics of transdermal testosterone gel in hypogonadal men. *J Clin Endocrinol Metab* 85, 4500–4510.

Tauber U, Schroder K, Dusterberg B & Matthes H. (1986) Absolute bioavailability of testosterone after oral administration of testosterone-undecanoate and testosterone. *Eur J Drug Metab Pharmacokinet* 11, 145.

Turner L, Conway AJ, Jimenez M, Liu PY, Forbes E, McLachlan RI & Handelsman DJ. (2003) Contraceptive efficacy of a depot progestin and androgen combination in men. *J Clin Endocrinol Metab* 88, 4659–4667.

Vermeulen A, Verdonck L & Kaufman JM. (1999) A critical evaluation of simple methods for the estimation of free testosterone in serum. *J Clin Endocrinol Metab* 84, 3666–3672.

Wang C & Swerdloff RS. (2010) Hormonal approaches to male contraception. *Curr Opin Urol* 20, 520–524.

Wang C, Festin MP & Swerdloff RS. (2016) Male hormonal contraception: where are we now? *Curr Obstet Gynecol Rep* 5, 38–47.

World Health Organization Task Force on Methods for the Regulation of Male Fertility. (1990) Contraceptive efficacy of testosterone-induced azoospermia in normal men. World health organization task force on methods for the regulation of male fertility. *Lancet* 336, 955–959.

World Health Organization Task Force on the Regulation of Male Fertility. (1996) Contraceptive efficacy of testosterone-induced azoospermia and oligozoospermia in normal men. *Fertil Steril* 65, 821–829.

Yin AY, Htun M, Swerdloff RS, Diaz-Arjonilla M, Dudley RE, Faulkner S, Bross R, Leung A, Baravarian S, Hull L, Longstreth JA, Kulback S, Flippo G & Wang C. (2012) Reexamination of pharmacokinetics of oral testosterone undecanoate in hypogonadal men with a new self-emulsifying formulation. *J Androl* 33, 190–201.

© 2016 American Society of Andrology and European Academy of Andrology

Molecular Human Reproduction, Vol. 0, No. 0 pp. 1–17, 2016
Mol. Hum. Reprod. Advance Access published March 17, 2016
Advanced Access publication on February 26, 2016 doi:10.1093/molehr/gaw016

molecular human reproduction

ORIGINAL RESEARCH

Structural analyses to identify selective inhibitors of glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme

Polina V. Danshina[1], Weidong Qu[1,5], Brenda R. Temple[2], Rafael J. Rojas[3,6], Michael J. Miley[4], Mischa Machius[4,7], Laurie Betts[1], and Deborah A. O'Brien[1,*]

[1]Department of Cell Biology and Physiology, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA [2]R.L. Juliano Structural Bioinformatics Core Facility, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA [3]Department of Pharmacology, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA [4]Macromolecular X-Ray Crystallography Core Facility, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA
[5]Present address: Key Laboratory of Public Health Safety, Ministry of Education, Department of Environmental Health, School of Public Health, Fudan University, Shanghai 200032, China
[6]Present address: Dart NeuroScience, LLC, San Diego, CA 92131, USA
[7]Present address: 230 Jamestown Road, Pittsboro, NC 27312, USA

*Correspondence address. E-mail: dao@med.unc.edu

Submitted on January 15, 2016; resubmitted on February 11, 2016; accepted on February 22, 2016

STUDY HYPOTHESIS: Detailed structural comparisons of sperm-specific glyceraldehyde 3-phosphate dehydrogenase, spermatogenic (GAPDHS) and the somatic glyceraldehyde 3-phosphate dehydrogenase (GAPDH) isozyme should facilitate the identification of selective GAPDHS inhibitors for contraceptive development.

STUDY FINDING: This study identified a small-molecule GAPDHS inhibitor with micromolar potency and >10-fold selectivity that exerts the expected inhibitory effects on sperm glycolysis and motility.

WHAT IS KNOWN ALREADY: Glycolytic ATP production is required for sperm motility and male fertility in many mammalian species. Selective inhibition of GAPDHS, one of the glycolytic isozymes with restricted expression during spermatogenesis, is a potential strategy for the development of a non-hormonal contraceptive that directly blocks sperm function.

STUDY DESIGN, SAMPLES/MATERIALS, METHODS: Homology modeling and x-ray crystallography were used to identify structural features that are conserved in GAPDHS orthologs in mouse and human sperm, but distinct from the GAPDH orthologs present in somatic tissues. We identified three binding pockets surrounding the substrate and cofactor in these isozymes and conducted a virtual screen to identify small-molecule compounds predicted to bind more tightly to GAPDHS than to GAPDH. Following the production of recombinant human and mouse GAPDHS, candidate compounds were tested in dose–response enzyme assays to identify inhibitors that blocked the activity of GAPDHS more effectively than GAPDH. The effects of a selective inhibitor on the motility of mouse and human sperm were monitored by computer-assisted sperm analysis, and sperm lactate production was measured to assess inhibition of glycolysis in the target cell.

MAIN RESULTS AND THE ROLE OF CHANCE: Our studies produced the first apoenzyme crystal structures for human and mouse GAPDHS and a 1.73 Å crystal structure for NAD^+-bound human GAPDHS, facilitating the identification of unique structural features of this sperm isozyme. In dose–response assays T0501_7749 inhibited human GAPDHS with an IC_{50} of 1.2 μM compared with an IC_{50} of 38.5 μM for the somatic isozyme. This compound caused significant reductions in mouse sperm lactate production ($P = 0.017$ for 100 μM T0501_7749 versus control) and in the percentage of motile mouse and human sperm (P values from <0.05 to <0.0001, depending on incubation conditions).

LIMITATIONS, REASONS FOR CAUTION: The chemical properties of T0501_7749, including limited solubility and nonspecific protein binding, are not optimal for drug development.

WIDER IMPLICATIONS OF THE FINDINGS: This study provides proof-of-principle evidence that GAPDHS can be selectively inhibited, causing significant reductions in sperm glycolysis and motility. These results highlight the utility of structure-based drug design and support further exploration of GAPDHS, and perhaps other sperm-specific isozymes in the glycolytic pathway, as contraceptive targets.

© The Author 2016. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

LARGE SCALE DATA: None. Coordinates and data files for three GAPDHS crystal structures were deposited in the RCSB Protein Data Bank (http://www.rcsb.org).

STUDY FUNDING AND COMPETING INTEREST(S): This work was supported by grants from the National Institutes of Health (NIH), USA, including U01 HD060481 and cooperative agreement U54 HD35041 as part of the Specialized Cooperative Centers Program in Reproduction and Infertility Research from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, and TW/HD00627 from the NIH Fogarty International Center. Additional support was provided by subproject CIG-05-109 from CICCR, a program of CONRAD, Eastern Virginia Medical School, USA. There are no conflicts of interest.

Key words: sperm motility / sperm metabolism / glycolysis / male contraception / glyceraldehyde 3-phosphate dehydrogenase-S / sperm-specific isozyme

Introduction

Extensive tissue-specific gene expression is a hallmark of spermatogenesis, providing a large number of sperm-specific proteins that are potential targets for male contraception (Schultz *et al.*, 2003; Matzuk and Lamb, 2008). The central metabolic pathway of glycolysis presents an extraordinary example of tissue specificity within the male germ line. Mammalian sperm have distinct isozymes at most steps of this pathway, resulting from germ cell-specific expression of at least three genes and from novel alternative splicing events (Eddy *et al.*, 2003; Vemuganti *et al.*, 2007; Ijiri *et al.*, 2013). Enzymes are typically considered good candidates for pharmaceutical intervention and account for >25% of the molecular targets for all known drugs (Nass and Strauss, 2004).

Glyceraldehyde 3-phosphate dehydrogenase, spermatogenic (GAPDHS), in particular, is a promising contraceptive target. This sperm-specific glycolytic isozyme is expressed only in the post-meiotic period of spermatogenesis, replacing the somatic isozyme (GAPDH) in mammalian sperm (Welch *et al.*, 1992, 2000; Bunch *et al.*, 1998). In multiple mammalian species the sperm isozyme shares ~70% amino acid identity with the somatic isozyme and has a novel proline-rich extension at the N-terminus. This N-terminal extension plays a role in anchoring GAPDHS to the fibrous sheath in the principal piece of the sperm flagellum (Bunch *et al.*, 1998; Krisfalusi *et al.*, 2006), but is not required for enzymatic activity (Elkina *et al.*, 2010; Sexton *et al.*, 2011). Gene targeting studies in mice established that GAPDHS (Miki *et al.*, 2004) and other sperm-specific isozymes in the glycolytic pathway (Odet *et al.*, 2008; Danshina *et al.*, 2010; Nakamura *et al.*, 2013) are essential for sperm motility and male fertility. These studies demonstrate the importance of glycolysis for sperm energy production. Spermatogenesis and sperm production appear normal in mice lacking GAPDHS, providing evidence that inhibition of this isozyme should not impair testicular function (Miki *et al.*, 2004).

Like the somatic isozyme, homotetrameric GAPDHS catalyzes the oxidation and phosphorylation of glyceraldehyde 3-phosphate (GAP) to form 1,3-bisphosphoglycerate, a reaction that requires NAD^+ and inorganic phosphate. This reaction occurs at an important transition point in glycolysis between the enzymatic steps that consume and generate ATP. Consequently, ATP levels in mouse sperm lacking GAPDHS are ~10% of wild-type levels immediately after collection in glucose-containing M16 medium, which fuels the ATP-consuming steps of glycolysis (Miki *et al.*, 2004). The requirement for glycolytic ATP production has also been demonstrated for human sperm (Williams and Ford, 2001; Nascimento *et al.*, 2008) and sperm from several other mammalian species (Storey, 2008; Mukai and Travis, 2012).

As a contraceptive target, GAPDHS must be subject to selective inhibition that does not disrupt glycolysis in other tissues. Initial evidence for selective inhibition was provided by early studies of α-chlorohydrin (3-chloro-1,2-propanediol) conducted before GAPDHS was identified as a distinct sperm-specific isozyme (reviewed in Jones, 1978; Jones and Cooper, 1999). S-3-chlorolactaldehyde, the major metabolite of α-chlorohydrin and related compounds, inhibited GAPDHS, sperm glycolysis and motility in a dose-dependent manner. These effects on sperm occurred at concentrations that did not inhibit GAPDH activity in other tissues (Brown-Woodman and White, 1975; Brown-Woodman *et al.*, 1978; Ford and Harrison, 1983). Efforts to develop these agents as contraceptives were abandoned due to toxic effects at high doses, although impurities and racemic mixtures of reactive compounds may have contributed to the observed toxicity (Jones and Cooper, 1999).

To assess the potential for selective inhibition of GAPDHS, we used structural analyses to highlight differences between the sperm and somatic isozymes that are conserved between species. We initially constructed homology models of human and mouse GAPDHS and GAPDH templated on the 2.0 Å crystal structure of *Palinurus versicolor* GAPDH (PDB 1CRW; Shen *et al.*, 2000). Active site structural features of our models were later confirmed in high resolution crystal structures of human GAPDHS (PDB 3H9E, Chaikuad *et al.*, 2011; PDB 5C7L and PDB 5C7O, this study), mouse GAPDHS (PDB 5C7I, this study) and human GAPDH (PDB 1U8F, Jenkins and Tanner, 2006). We conducted virtual screening to identify small-molecule compounds that are predicted to bind more tightly to GAPDHS than GAPDH. These compounds were then tested in enzymatic assays to compare inhibition of the sperm and somatic enzymes, and in functional assays to determine their effects on sperm metabolism and motility.

Materials and Methods

Reagents

Standard reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) with the exception of glucose, magnesium sulfate heptahydrate, sodium bicarbonate, sodium chloride, sodium isopropyl-β-D-thiogalactopyranoside (IPTG), sodium pyruvate, tris(2-carboxyethyl)phosphine (TCEP) (ThermoFisher Scientific, Waltham, MA, USA); potassium chloride, potassium phosphate (Mallinckrodt Baker, Phillipsburg, NJ, USA); penicillin/streptomycin 100X stock solution containing 10 000 U/ml of penicillin G and 10 mg/ml of streptomycin (Gemini Bio-Products, West Sacramento, CA, USA); cOmplete protease inhibitor cocktail (Roche Life Science, Indianapolis, IN, USA); and PEG 3350 (Emerald Biosystems, Bainbridge Island, WA, USA).

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

Homology modeling and virtual screening

The 2.0 Å crystal structure of *Palinurus versicolor* GAPDH (PDB 1CRW; Shen *et al.*, 2000) was used as a template for homology modeling of human and mouse GAPDH and GAPDHS isozymes, and binding sites for GAP and NAD^+ were determined by reference to structures PDB 1DC4 and 1DC6 (Yun *et al.*, 2000). The proline-rich N-terminal extensions that are specific to the sperm isozymes were not included in these models. Therefore, mouse GAPDHS was modeled without the first 105 amino acids of its sequence, and human GAPDHS without the first 75 amino acids. Homology models were constructed using the modeler module of the Insight II molecular modeling program (Accelrys, San Diego, CA, USA). Final structures were tested for sequence-structure compatibility using the Verify function in the Profiles-3D module. Results were displayed graphically using the PyMOL Molecular Graphics System (Schrödinger, LLC, New York, NY, USA).

To assess the potential for selective inhibition of the sperm isozymes, we conducted virtual screening with the SYBYL 6.9 molecular modeling program (Tripos, St. Louis, MO, USA). The SiteID module was used to identify three solvent-accessible binding pockets surrounding the active site in our homology models. After pre-filtering by size, compounds from the Maybridge small molecule database were docked into each of the binding pockets using the SYBYL FlexX algorithm for the flexible docking of small ligands (Rarey *et al.*, 1995). Protein-ligand interaction scores were computed for each isozyme, and the predicted discrimination between sperm and somatic isozymes was calculated as the difference between interaction scores. Docking was repeated thirty times for the compounds with the largest discrimination scores for the human isozymes to confirm that these compounds consistently docked in the predicted binding pocket.

Expression and purification of recombinant proteins

As in previous studies (Frayne *et al.*, 2009; Elkina *et al.*, 2010; Sexton *et al.*, 2011), we expressed only trace amounts of full-length mouse or human GAPDHS even though we purchased DNA sequences with codon usage optimized for expression in *Escherichia coli* (GeneArt, Regensburg, Germany). Therefore, we subcloned the optimized sequences and expressed truncated forms of the sperm isozymes lacking their proline-rich N-terminal extensions (tGAPDHS). Several fusion constructs were tested to optimize expression of the sperm isozymes.

The DNA fragment encoding human tGAPDHS (amino acids 76–408) was cloned into the pGEX-4T-1 vector (GE Healthcare Life Sciences, Piscataway, NJ, USA) for expression as a glutathione S-transferase (GST) fusion protein. Recombinant protein was expressed in gapA-deficient *E. coli* DS112, strain K-12 (Yale Coli Genetic Stock Center, New Haven, CT, USA) to avoid the formation of mixed tetramers that contain bacterial GAPDH (Frayne *et al.*, 2009; Sexton *et al.*, 2011). Cells were grown at 37°C in M9 minimal media supplemented with 1% glycerol, 0.1% casamino acids, 0.4% sodium succinate, 0.001% thiamine, 0.004% methionine, 0.004% tryptophan and 0.002% uracil until absorbance at 600 nm was 0.5–0.6. Expression was induced by 0.1 mM IPTG during overnight culture at 18°C. Cells were harvested by centrifugation and resuspended in phosphate-buffered saline (PBS: 136 mM NaCl, 2.7 mM KCl, 10 mM $NaH_2PO_4{\cdot}7H_2O$, 1.7 mM KH_2PO_4; pH 7.4) supplemented with 2 mM dithiothreitol (DTT), 5 mg/ml lysozyme, 5 μg/ml DNase I and cOmplete protease inhibitor cocktail. After incubation for 30 min on ice, the suspension was sonicated with short pulses for 2 min and centrifuged at 30 000*g* for 1 h at 4°C. The resulting supernatant was loaded onto a glutathione Sepharose 4B (GE Healthcare Life Sciences) column prepared according to the manufacturer's instructions and washed with PBS containing 2 mM DTT. To remove the GST tag, the column was incubated overnight at room temperature with ~40 units of bovine thrombin/ml bed volume. Cleaved tGAPDHS was eluted, frozen in liquid nitrogen and stored at −70°C.

The DNA fragment encoding mouse tGAPDHS (amino acids 106–438) was subcloned into the pMal vector (New England Biolabs, Ipswich, MA, USA), which incorporates a thrombin-cleavable maltose-binding protein (MBP) tag. Recombinant mouse tGAPDHS was expressed using the same procedure described for human tGAPDHS, except that buffer A (20 mM Tris–HCl, 200 mM NaCl, 10 mM EDTA, pH 7.4) replaced PBS in the cell lysis solution. The clarified supernatant was loaded onto an amylose column (New England Biolabs) equilibrated with buffer A and 5% glycerol, followed by overnight incubation at room temperature with ~40 units of bovine thrombin/ml bed volume. The eluate from this column, containing recombinant protein and a fraction of the cleaved MBP tag, was dialyzed against 2000 volumes of 20 mM Tris–HCl, 25 mM NaCl, 2 mM β-mercaptoethanol, pH 8.0. Following dialysis, the protein solution was loaded onto a diethylaminoethyl (DEAE)-Sepharose column equilibrated with 20 mM Tris–HCl, 10 mM NaCl, 2 mM DTT, pH 8.0, which retained the MBP tag. Mouse tGAPDHS was eluted, frozen in liquid nitrogen and stored at −70°C.

Mouse somatic GAPDH was expressed as a GST-fusion protein and purified according to the same procedures used for human tGAPDHS. The purity of all recombinant enzyme preparations was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (Sexton *et al.*, 2011). Human somatic GAPDH, purified from erythrocytes, was purchased from Sigma-Aldrich (Cat. No. G6019).

Crystallization of tGAPDHS

Crystal structures were obtained using resources in the Macromolecular X-Ray Crystallography Core Facility at the University of North Carolina School of Medicine (Chapel Hill, NC, USA) as well as the SER-CAT beamline at the Advanced Photon Source of the Argonne National Laboratory (Lemont, IL, USA). The PACT Suite screening kit (Qiagen, Valencia, CA, USA) was used to assess multiple conditions for crystallization with the hanging-drop vapor-diffusion method. Apoenzyme crystals of mouse tGAPDHS were obtained when purified protein (3 mg/ml in PBS) was mixed in a 1:1 volume ratio with a reservoir solution containing 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane, pH 6.5, 20% PEG 3350. These crystals, which appeared in the hanging drop after 6 days at 20°C, were cryoprotected with 25% glycerol in crystal growth solution and then flash-frozen in liquid nitrogen.

Human tGAPDHS was concentrated to 10 mg/ml in 10 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 0.5 mM TCEP, 0.01% of sodium azide and stored at −80°C. Purified protein was mixed in a 1:1 volume ratio with each of the selected crystallization conditions from the screening kit. Crystals appeared within two days at 20°C. Apoenzyme crystals were obtained with F8 PACT Suite reservoir conditions (0.2 M Na_2SO_4, 0.1 M Bis Tris propane, pH 6.5, 20% PEG 3350) and holoenzyme crystals of tGAPDHS complexed with NAD^+ were obtained with E5 PACT Suite reservoir conditions (0.2 $NaNO_3$, 20% PEG3350). Crystals were cryoprotected and flash-frozen as described for mouse tGAPDHS.

Diffraction data were collected from flash-frozen crystals at 100 K either on a Rigaku rotating anode generator using the Saturn944 CCD detector (mouse apoenzyme), or the Advanced Photon Source SER-CAT beamline (human apo- and holoenzymes). Crystal structures were solved by molecular replacement with PHASER (McCoy *et al.*, 2007) in the CCP4 program suite (Winn *et al.*, 2011), using human placenta GAPDH (PDB 1U8F, Jenkins and Tanner, 2006) and human sperm GAPDHS (PDB 3H9E, Chaikuad *et al.*, 2011) structures as search models for the mouse and human sperm enzymes, respectively. The 2.01 Å mouse tGAPDHS apoenzyme structure was refined initially using CCP4 RefMac5 (Vagin *et al.*, 2004), with final refinements using PHENIX (Adams *et al.*, 2010) and keeping the R_{free} constant. Human tGAPDHS structures were refined to 1.86 Å (apoenzyme) and 1.73 Å (holoenzyme) using PHENIX (Adams *et al.*, 2010). Data collection and refinement statistics are shown in Table I.

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

Table I Data collection and refinement statistics for human and mouse truncated glyceraldehyde 3-phosphate dehydrogenase, spermatogenic (tGAPDHS) structures.

Parameter	Human tGAPDHS-NAD$^+$	Human tGAPDHS-apo	Mouse tGAPDHS-apo
PDB code	5C7O	5C7L	5C7I
Data collection			
Beamline	Advanced Photon Source, 22-ID	Advanced Photon Source, 22-ID	Rigaku 007-HF/Saturn944 CCD
Wavelength (Å)	1.0746	1.0746	1.5418
Space group	C2	P 3_121	P 3_121
Monomers/asymmetric unit	2	2	2
Unit cell dimensions	$a = 144.4$, $b = 71.8$, $c = 80.8$; $\alpha = \gamma = 90.0^\circ$, $\beta = 123.01^\circ$	$a = 86.9$, $b = 86.9$, $c = 159.1$; $\alpha = \beta = 90.0^\circ$, $\gamma = 120.0^\circ$	$a = 86.7$, $b = 86.7$, $c = 158.4$; $\alpha = \beta = 90.0^\circ$, $\gamma = 120.0^\circ$
Resolution range (Å)	31.71–1.73 (1.79–1.73)	38.2–1.86 (1.93–1.86)	30.6–2.01 (2.08–2.01)
# of unique reflections	72 313 (1819)	59 104 (1462)	44 704 (4260)
Completeness	93.9 (55.7)	99.8 (99.9)	95.9 (92.9)
$I/\sigma I$	19.2 (2.0)	17.5 (2.1)	23.7 (4.6)
R_{meas}	0.091	0.169	0.071
R_{pim}	0.044	0.052	Not calculated
$CC_{1/2}$	0.922	0.895	Not calculated
CC*	0.981	0.952	Not calculated
Redundancy	4.2 (4.0)	8.7 (6.3)	3.5 (3.1)
Refinement			
# of non-hydrogen atoms (protein/ligand/water O)	5218/98/629	5210/384	5107/447
Number of reflections and cutoff	68 115/$F > 0\sigma F$	54 421/$F > 0\sigma F$	44 704/$F > 0\sigma F$
R_{work} (%)	13.7 (18.9)	18.7 (22.1)	19.2 (23.2)
R_{free} (%)	16.6 (21.2)	22.7 (26.0)	21.3 (27.8)
Average B-factor (Å^2) (protein/ligand/water O)	23.8/26.5/35.2	29.5/NA/33.9	29.2/NA/35.7
Wilson B-factor	19.7	23.1	24.6
RMSD bond length (Å)	0.005	0.013	0.004
RMSD bond angles (°)	1.03	1.4	0.907
Ramachandran favored/allowed (%)	95.9/3.8	96.3/3.4	95.2/4.5
Ramachandran outliers (%)	0.3	0.3	0.3

Values in parentheses are for highest resolution shells.
Standard crystallography statistics are shown for three structures determined for tGAPDHS. X-ray diffraction provides the intensity (I) of n measurements of reflections, which yield structure factor (F) values with defined indices (h, k, l). R factor statistics measure agreement between the experimental X-ray diffraction data and the crystallographic model. R_{meas} and R_{pim} are indicators of data quality and R_{work} and R_{free} assess the quality of the model. RMSD refers to root-mean-square deviation. Formulas for calculating the statistics shown are:
$I/\sigma I$ = average intensity/average standard deviation of intensity.
$R_{meas} = (\sum_{hkl}(n/n-1)^{1/2}\sum_{i=1,n}|I_i(hkl) - \langle I\rangle(hkl)|)/\sum_{hkl}\sum_{i=1,n}I_i(hkl))$, where $\langle I\rangle$ = average I of the n measurements.
$R_{pim} = (\sum_{hkl}(1/n-1)^{1/2}\sum_{i=1,n}|I_i(hkl) - \langle I\rangle(hkl)|)/\sum_{hkl}\sum_{i=1,n}I_i(hkl))$, where $\langle I\rangle$ = average I of the n measurements.
$R_{work} = (\sum_{hkl}|F_{obs}(hkl) - F_{calc}(hkl)|/\sum_{hkl}F_{obs}(hkl))$, ALL reflections, obs = observed, calc = calculated.
$R_{free} = (\sum_{hkl}|F_{obs}(hkl) - F_{calc}(hkl)|/\sum_{hkl}F_{obs}(hkl))$, 5% randomly chosen reflections NOT included in refinement.
$CC_{1/2}$ = Pearson's correlation coefficient between randomly chosen 'half' of each data set.
$CC^* = (2\,CC_{1/2}/1 + CC_{1/2})^{\frac{1}{2}}$.

Enzyme inhibition and kinetics

The dehydrogenase activity of GAPDH or GAPDHS was monitored in a kinetic assay that measures NADH accumulation at 340 nm (Schmalhausen *et al.*, 1997). The assay was adapted for 96-well plates and the final reaction mixture in each well (200 µl) contained 100 mM glycine, 100 mM potassium phosphate, 5 mM EDTA, 0.5 mM NAD$^+$, 0.5 mM D-glyceraldehyde-3-phosphate (Sigma-Aldrich, 39705) and 0.5 µg of protein (GAPDH or tGAPDHS). One enzyme unit is defined as the amount of enzyme necessary for the formation of 1 µmole 1,3-diphosphoglycerate/min. Twenty-five compounds identified as potential selective inhibitors of GAPDHS in our virtual screen were purchased (Ryan Scientific, Mount Pleasant, SC, USA) and tested for inhibition of both sperm and somatic isozymes. Stock solutions (10 mM) of each compound were prepared in dimethyl sulfoxide (DMSO), aliquoted and frozen at −20°C. Following pre-incubation of the test compounds with enzyme and NAD$^+$ for 30 min at 37°C, enzymatic reactions were initiated by the addition of GAP. Triplicate samples were included for each concentration tested and DMSO controls were included in each assay. IC_{50} values for each compound were calculated in Prism (GraphPad Software, La Jolla, CA, USA) using the four parameter nonlinear regression model. Results are expressed as the mean ± SEM for at least three different experiments with different batches of recombinant isozymes.

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

Kinetic assays of inhibition were conducted for each substrate (GAP or NAD^+) at several inhibitor concentrations (1–100 μM) and the results were analyzed with Enzyme Kinetics Pro (ChemSW, Fairfield, CA, USA). Duplicate samples were measured in all assays, which were repeated three times with different enzyme preparations. Data were fit using the Lineweaver–Burk linearization method to identify the mode of inhibition and estimates K_i values.

Inhibitor effects on sperm motility and metabolism

Mouse sperm were collected from the cauda epididymides of adult CD1 males (>8 weeks old; Charles River Laboratories, Raleigh, NC, USA) in human tubal fluid (HTF) medium as described previously (Goodson *et al.*, 2011). All procedures involving mice were performed according to the Guide for the Care and Use of Laboratory Animals with prior approval by the Institutional Animal Care and Use Committee within the AAALAC accredited program at the University of North Carolina at Chapel Hill (UNC-CH) (Animal Welfare Assurance Number: A-3410-01).

De-identified human semen samples were obtained from excess stocks collected by the UNC-CH Fertility Clinic, Department of Obstetrics and Gynecology. All human samples were from healthy donors with normal semen parameters and screened for HIV (human immunodeficiency virus), RPR (rapid plasma reagin: syphilis) and hepatitis. Criteria for selection of donor sperm included normal morphology using strict criteria (World Health Organization, 2010) and a post-thaw count of $>80 \times 10^6$ motile sperm. Semen samples were diluted 1:1 with TEST-yolk buffer (Prins and Weidel, 1986) following collection, divided into aliquots with $\geq 20 \times 10^6$ sperm, and stored in liquid nitrogen. After thawing, sperm were isolated from seminal plasma and cryoprotectant by low speed centrifugation (365*g*) for 10 min through a 90% colloidal silica suspension (ISolate, Irvine Scientific, Santa Ana, CA, USA) at a 1:1 volume ratio. The upper layer was discarded and the remaining sperm pellet was resuspended into 2 ml HEPES buffered Sperm Washing Medium (Irvine Scientific) and centrifuged for 5 min at 365*g*. The washing step was repeated twice and the final pellet was resuspended in 1–2 ml of HTF medium without human serum albumin.

Sperm were incubated at 37°C under 5% CO_2 in air for 2 h in HTF medium containing 50 or 100 μM T0501_7749. Effects on motility were monitored by computer-assisted sperm analysis (CASA) with the Hamilton Thorne CEROS imaging system, version 12.3H IVOS software (Goodson *et al.*, 2011). Sperm tracks (1.5 s) were captured at a frame acquisition rate of 60 Hz. Aliquots were removed at 30 min intervals to compare sperm motility in the presence of inhibitors with motility in the DMSO vehicle control. Motility analyses were repeated in modified HTF medium, omitting bovine serum albumin (BSA) or human serum albumin or replacing BSA with 0.4 mM methyl-β-cyclodextrin and 0.01% polyvinyl alcohol (Hasegawa *et al.*, 2012). For each condition, experiments were repeated at least three times with sperm from different mice or human donors. Sperm viability was monitored by the uptake of propidium iodide by cells with damaged plasma membranes, as described previously (Goodson *et al.*, 2012).

Lactate production and accumulation in the medium was measured in mouse sperm to monitor inhibitor effects on sperm glycolysis. Sperm were incubated for 2 h ± T0501_7749 under the same conditions used for motility analyses in HTF medium with 0.4 mM methyl-β-cyclodextrin and 0.01% polyvinyl alcohol, except that lactate and pyruvate were omitted from the medium. Osmolality of the medium was adjusted to ~315 mOsm/kg with 5 M NaCl using a Model 3300 micro-osmometer (Advanced Instruments, Norwood, MA, USA). Duplicate aliquots were removed at time 0 and 2 h, centrifuged to remove sperm and assayed to measure lactate accumulation in the medium. The spectrophotometric assay monitors the conversion of lactate to pyruvate by lactate dehydrogenase in the presence of NAD^+ and hydrazine (Pesce *et al.*, 1975; Danshina *et al.*, 2001). In this assay the concentration of lactate in the sample is proportional to the increase in absorbance at 340 nm as NAD^+ is reduced to NADH.

Statistical analyses were performed using Prism (GraphPad Software, La Jolla, CA, USA). Data are shown as mean ± SEM and statistical significance ($P < 0.05$) was calculated using either two-tailed unpaired *t*-tests or two-way analysis of variance followed by Tukey's multiple-comparison test.

Results

Structural differences between sperm and somatic isozymes

Both mouse and human sperm require glycolysis (Miki *et al.*, 2004; Nascimento *et al.*, 2008) and have GAPDHS isozymes with 83% amino acid identity, suggesting that the mouse may be useful as a model system for testing GAPDHS inhibitors as potential contraceptives. Alignment of the amino acid sequences for human and mouse GAPDHS and GAPDH shows that 68 amino acids are identical in the sperm isozymes (highlighted in red in Fig. 1), but distinct from the corresponding amino acids in the somatic isozymes. The proline-rich N-terminal regions of human (amino acids 1–75) and mouse (amino acids 1–105) GAPDHS are not included in the alignment since homologous regions do not exist in their somatic counterparts. Unless noted otherwise, we will refer to amino acid numbering based on the sequences of human GAPDHS (top row in Fig. 1) and human GAPDH (bottom row in Fig. 1). Each subunit of the tetrameric enzyme has an NAD^+-binding domain in the N-terminal segment of the protein preceding the active site C224 and a catalytic domain in the C-terminal half. Eight amino acids in the C-terminal domain that have crucial roles in catalysis (denoted by dots above the sequences in Fig. 1; Seidler, 2013) are conserved in both sperm and somatic isozymes. Forty-six of the sperm-specific residues are in the NAD^+-binding domain and the remaining 22 are in the catalytic domain.

We constructed homology models for mouse and human GAPDHS and GAPDH based on the 2.0 Å crystal structure of *Palinurus versicolor* GAPDH (PDB 1CRW; Shen *et al.*, 2000) to assess the three-dimensional location of the sperm-specific amino acids. Models of the sperm isozymes represent tGAPDHS without the proline-rich N-terminal extensions. Most of the 68 sperm-specific amino acids noted in Fig. 1 are located on the surface of the protein subunit (pink in Supplementary Video S1). However, two regions with sperm-specific amino acids (red in Supplementary Video S1) are closer to the active site. One region near the center of the model is adjacent to the catalytic site where both GAP (orange) and the nicotinamide ring of NAD^+ (gray) bind. The second region near the top of the model is at the opposite end of the NAD^+ binding site surrounding the adenine moiety of the cofactor.

To evaluate the potential for differential binding to the sperm and somatic isozymes, we used SiteID in the SYBYL molecular modeling program to identify three binding pockets surrounding the active site in our tGAPDHS homology models (Fig. 2A). Subsequent comparisons with high resolution crystal structures for human tGAPDHS (PDB 3H9E, Chaikuad *et al.*, 2011 and crystal structures solved in this study) and human GAPDH (PDB 1U8F, Jenkins and Tanner, 2006) confirmed the conformations of all three pockets. Pockets 1 and 2 (blue and green, respectively, in Fig. 2A) contain the two regions with sperm-specific amino acids noted in Supplementary Video S1. These amino acid differences are highlighted in Fig. 2B, where structures of the sperm (pink) and somatic (gray) isozymes are aligned and superimposed.

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

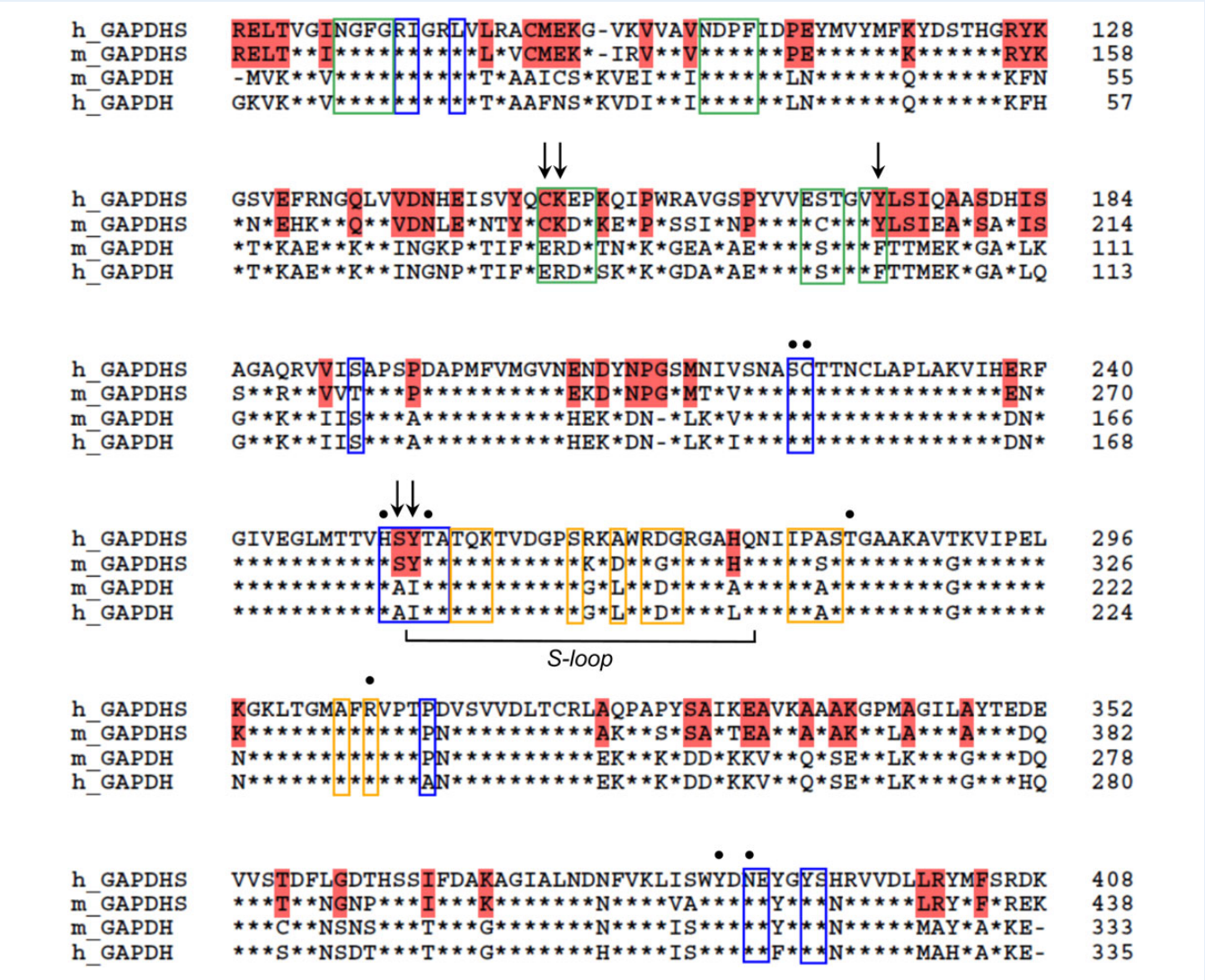


Figure 1 Alignment of human (h) and mouse (m) glyceraldehyde 3-phosphate dehydrogenase, spermatogenic (GAPDHS) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) amino acid sequences. The human GAPDHS sequence (NP_055179) is shown on the top row, without the proline-rich N-terminal region (amino acids 1–75) that is characteristic of the sperm isozyme. Similarly, mouse GAPDHS (NP_032111) in the second row does not include the N-terminal 105 residues. Highlighted in red are amino acids that are conserved in these sperm isozymes, but distinct from the corresponding amino acids in the somatic GAPDH isozymes shown in the third and fourth rows (mouse, AAH83149 and human, NP_002037). The sperm isozymes have 83% amino acid identity, compared with the ~70% identity between the sperm and somatic isozymes in each species. Dots above the sequences denote amino acids required for catalysis. Colored boxes indicate amino acids that form pockets 1 (blue), 2 (green) and 3 (yellow) surrounding the active site, as identified by SiteID in the SYBYL molecular modeling program. Arrows denote sperm-specific amino acids in pockets 1 and 2. Amino acids forming the S-loop are labeled.

Pocket 1 (blue in Fig. 2A) surrounds the substrate and the nicotinamide moiety of NAD^+. It is formed by the side-chains of R85, I86, L89, S193, S223, C224, H251, S252, Y253, T254, A255, P310, N388, E389, Y392 and S393 of human GAPDHS (blue boxes in Fig. 1). Pocket 1 includes two polar residues, S252 and Y253 (arrows in Fig. 1 and in the lower box of Fig. 2B), that are specific to the sperm isozymes and are located within 6 Å of the active site where they are flanked by two of the amino acids required for catalysis (Fig. 1). Unlike the smaller, hydrophobic residues (A180 and I181) present at corresponding positions in the somatic enzymes, S252 and Y253 have the potential for hydrogen bond formation. Pocket 1 of human GAPDHS has a third amino acid (P310) that is distinct from the corresponding residue in human GAPDH (A238).

Pocket 2 (green in Fig. 2A) surrounds the adenine moiety of NAD^+ and is formed by the side-chains of N81, G82, F83, G84, N105, D106, P107, F108, C150, K151, E152, P153, E168, S169, T170, V172 and Y173 of human GAPDHS (green boxes in Fig. 1). There are three sperm-specific amino acids in pocket 2, C150, K151 and Y173 (arrows in Fig. 1 and in the upper box of Fig. 2B) replacing E79, R80 and F102 in the somatic isozymes. Both K151 and Y173 are within 2.5 Å of the adenine portion of NAD^+. C150 and K151 in GAPDHS are smaller than the corresponding residues in GAPDH, and the C150 residue alters charge and provides a

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

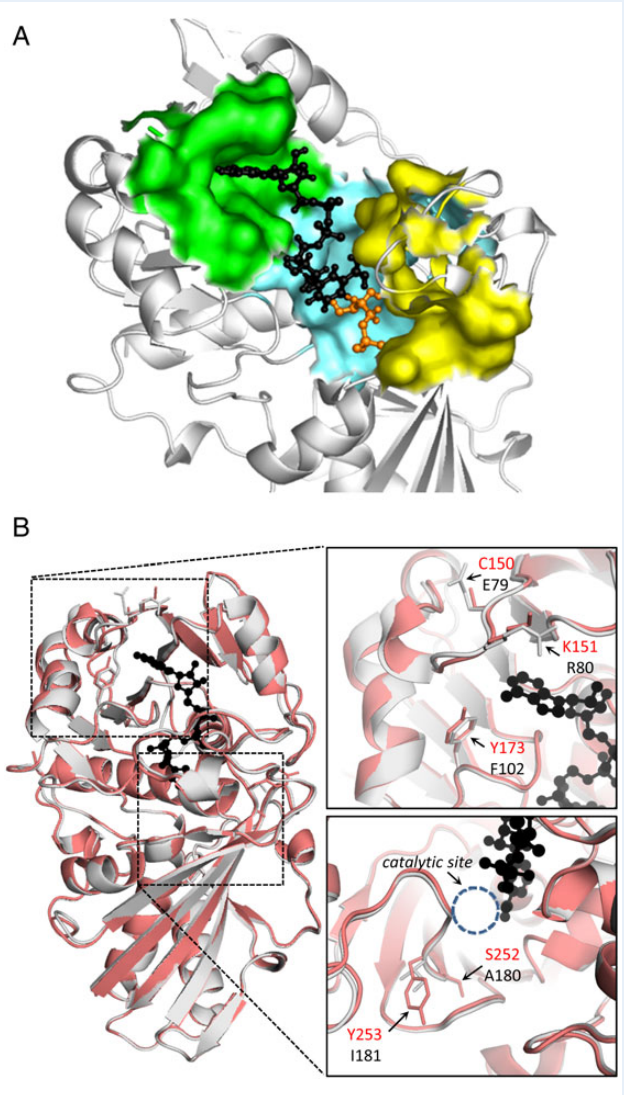


Figure 2 Binding pockets and sperm-specific amino acids near the active site. (**A**) To facilitate virtual screening, three binding pockets were identified in our homology model of truncated GAPDHS (tGAPDHS) using SiteID in the SYBYL molecular modeling program. Surfaces are shown for pockets 1 (blue), 2 (green) and 3 (yellow) which surround the glyceraldehyde 3-phosphate (GAP) substrate (orange) and NAD^+ cofactor (black). (**B**) Superimposed structures of human tGAPDHS (pink, PDB 3H9E) and GAPDH (gray, PDB 1U8F) highlighting the position of sperm-specific residues in pockets 1 (lower box) and 2 (upper box).

nucleophilic thiol group. The polar Y173 residue replaces a smaller, non-polar residue in GAPDH (F102), shown previously to interact directly with NAD^+ through hydrophobic contacts (Moras *et al.*, 1975). It is interesting to note that the five sperm-specific residues identified in pockets 1 and 2 are all highly conserved in mammalian GAPDHS orthologs identified in Ensembl (Supplementary Table SI). C150 and Y173 in pocket 2 are also conserved in reptile and fish orthologs.

Pocket 3 (yellow in Fig. 2A) surrounds the inorganic phosphate binding site and is formed by the side-chains of T256, Q257, K258, S264, A267, R269, D270, G271, I279, P280, A281, S282, A304 and R306 of human GAPDHS (yellow boxes in Fig. 1). This pocket does not contain sperm-specific amino acids that are conserved between human and mouse GAPDHS, although L195 in the somatic isozyme is replaced by A267 in human GAPDHS and D297 in mouse GAPDHS. Two additional residues (G300 and S311) are distinct only in mouse GAPDHS, replacing D198 and A209 in the somatic isozyme.

To explore GAPDHS structures in greater detail, we expressed recombinant human and mouse tGAPDHS without their proline-rich N-terminal regions and compared multiple crystal structures for these sperm isozymes. We solved three distinct crystal structures (Table 1), including the first apoenzyme crystal structures for human tGAPDHS (PDB 5C7L, resolution 1.86 Å) and mouse tGAPDHS (PDB 5C7I, 2.01 Å). We also obtained a 1.73 Å holoenzyme structure for human tGAPDHS complexed with NAD^+ (PDB 5C7O). The coordinates and data files for these crystal structures were deposited in the RCSB Protein Data Bank (http://www.rcsb.org). Using subunit nomenclature previously established for GAPDH (Buehner *et al.*, 1974), the asymmetric unit of the human sperm NAD^+-bound holoenzyme is a dimer of O and P subunits, while the asymmetric unit in the human and mouse apoenzyme structures is an O-Q dimer. In both cases, the consensus tetramer of the GAPDH family is formed by crystallographic symmetry that recapitulates the biological tetramer. As reported for multiple GAPDH and GAPDHS structures (Cowan-Jacob *et al.*, 2003; Ismail and Park, 2005; Frayne *et al.*, 2009), the active site cysteines (C224 in human GAPDHS, C256 in mouse GAPDHS) in our structures were oxidized to the sulfinic acid form, with clear electron density for two oxygens attached to the cysteine sulfur.

The C2 space group of our NAD^+-bound structure for human tGAPDHS (PDB 5C7O) is the same as the published NAD^+-bound structures for this isozyme (PDB 3H9E and 3PWF, Chaikuad *et al.*, 2011), but distinct from the $P3_121$ space group of both apoenzyme structures. A novel feature of our NAD-bound tGAPDHS structure is the presence of disulfide bonds between tetramers, linking the C150 residue of each monomer with the C150 of the symmetry-related monomer in the adjacent tetramer. C150 is one of the highly conserved sperm-specific amino acids in GAPDHS (pocket 2 in our homology model), replacing E79 in the somatic isozyme. Figure 3A shows the tetramers stacked and translated as they are in the holoenzyme crystal, with the C150 residues shown as black spheres. Disulfide bonds between these residues facilitate stacking of the tetramers parallel to the *R*-axis of the C2 space group, which would continue along the direction of this axis. The C150 residue is within 5 Å of the adenine ring in the NAD^+ bound within the crystal. As shown in the electron density plot of this tetramer interface (Fig. 3B), the disulfide bond between C150 residues is the major conformation observed in this crystal structure (occupancy = 0.88). The minor non-disulfide-bonded conformation of C150 has an occupancy of only 0.12.

Inter-subunit disulfide bonds are not found in either the human or mouse tGAPDHS crystal structures without NAD^+ (PDB 5C7L and PDB 5C7I), where the loop containing C150 (residues 148–158) has a slightly different conformation. Figure 3C shows two tetramers of the human apoenzyme that are rotated and translated, illustrating that they cannot stack as they do in the C2 holoenzyme (Fig. 3A). When the human apo- and holoenzyme structures are superimposed (CCP4 Superpose, Krissinel and Henrick, 2004), the root-mean-square deviation (RMSD) for all atoms of the C150 loop (including hydrogens and side-chains) is 0.7 Å, larger than the 0.38 RMSD over the entire structure. In the apoenzyme, the crystal packing changes since the relationship

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

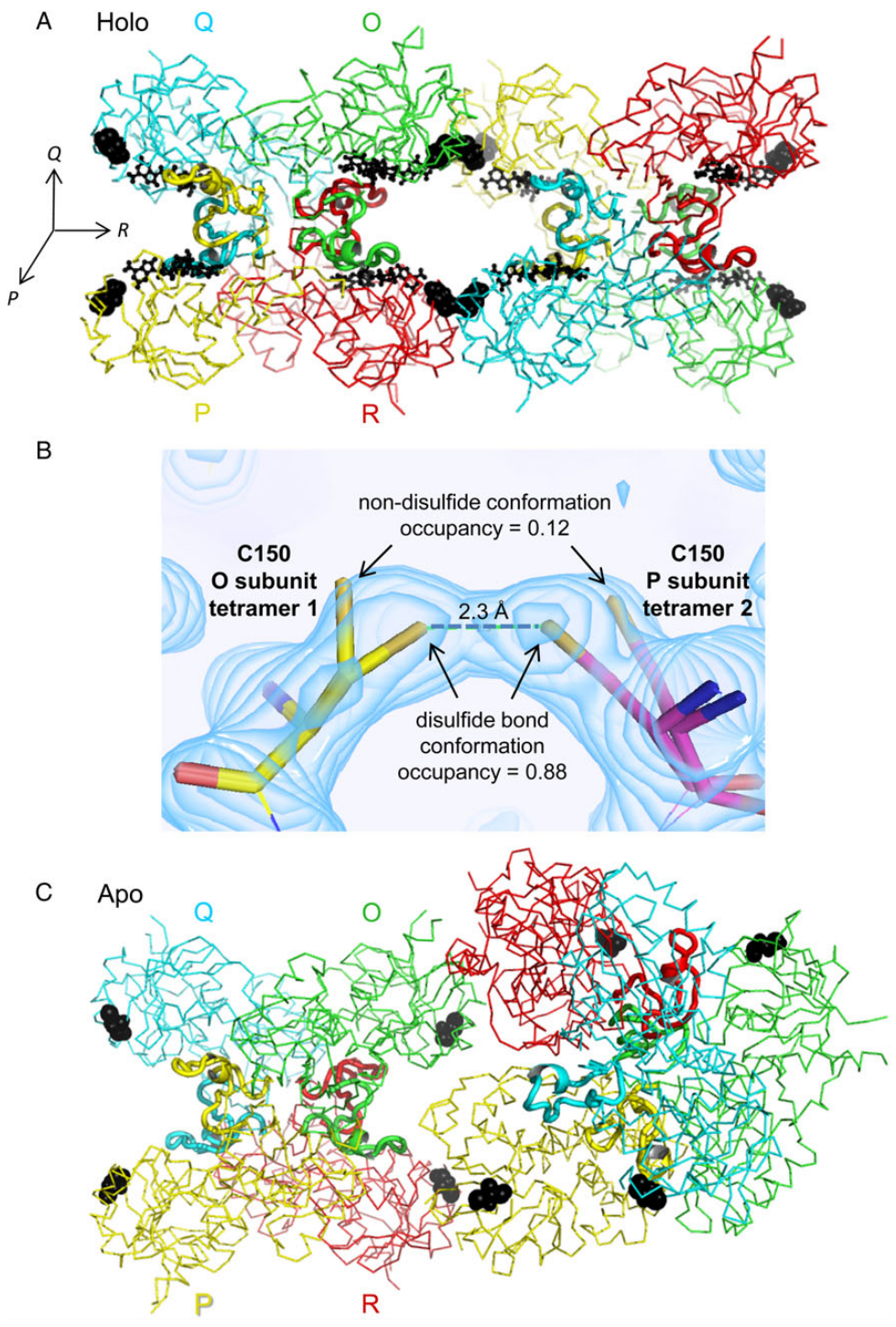


Figure 3 Features of the holo- (PDB 5C7O) and apoenzyme (PDB 5C7L) crystal structures of human tGAPDHS. (**A**) Two holoenzyme tetramers stacked and translated as they are in the C2 crystal form. Subunits and axes are labeled as described by Buehner *et al.* (1974). The bound NAD^+ cofactors are shown in each subunit as black ball-and-stick structures. In this holoenzyme structure the sperm-specific C150 residues (highlighted as black spheres) form disulfide bonds between tetramers, facilitating stacking parallel to the *R*-axis of the C2 space group. The S-loops of each subunit are shown as thicker ribbons. (**B**) Electron density (2Fo-Fc contoured at 1σ, where Fo is the observed structure factor and Fc is the calculated structure factor) around the inter-tetramer disulfide bond seen in the NAD^+-bound holoenzyme. The figure shows that there is a major disulfide-bonded conformation and a minor non-disulfide-bonded conformation of C150. The distance between sulfur atoms (2.3 Å) is slightly longer than the consensus disulfide bond, but there is continuous electron density between them. (**C**) Apoenzyme tetramers are more open, packing in the $P3_121$ crystal form without disulfides between the C150 residues (black spheres).

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

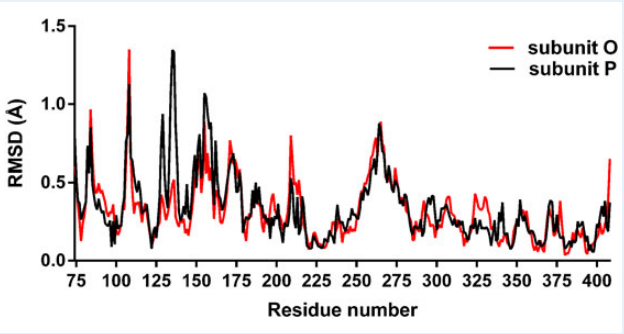


Figure 4 Root-mean-square deviation (RMSD, Å) between equivalent Cα atoms along the polypeptide chains of the O and P subunits when the human apo- and holo-tGAPDHS crystal structures (PDB 5C7L and PDB 5C7O) are superimposed. Displacement is greatest within the N-terminal NAD^+-binding domain and in the S-loop (residues 253–276) within the catalytic domain.

between monomers in the tetramer is slightly more open. For this reason, the apposition of C150 residues and consequent disulfide bonds that are facilitated in the holo space group are not allowed in the trigonal space group that the apoenzyme forms.

Previous comparisons of GAPDH crystal structures indicate that differences between the apo- and holoenzyme occur primarily in the N-terminal NAD^+-binding domain (Skarzynski and Wonacott, 1988; Shen *et al.*, 2000; Yun *et al.*, 2000), reflecting the conformational changes that occur during cofactor binding. To assess conformational changes in the sperm isozyme, we superimposed our human apo- and holo-tGAPDHS structures with CCP4 Superpose (Krissinel and Henrick, 2004) and plotted the RMSD between equivalent Cα backbone atoms along the polypeptide chain for both subunits of the crystallographic unit (Fig. 4). Comparable to the somatic isozyme, the greatest displacement for tGAPDHS was seen in the NAD^+-binding domain (residues 76–223). Only minor variations (<0.5 Å) were apparent in most of the C-terminal catalytic domain. However, there was a larger conformational difference between the apo- and holo-tGAPDHS structures in the S-loop region (residues 253–276, thicker ribbon regions in Fig. 3, Banas *et al.*, 1987), with maximum displacement (0.88 Å) centered at R265 and K266. Several S-loop amino acids are included in binding pockets 1 and 3 identified with SiteID, including the sperm-specific Y253 residue in pocket 1. Structural analyses of GAPDH indicate that the S-loops form the core of the tetrameric enzyme, making contacts with neighboring subunits across the *R* and *P* axes (Biesecker *et al.*, 1977). Supplementary Video S2 illustrates the conformational change in the GAPDHS S-loop that occurs when NAD^+ binds.

We also compared the mouse and human tGAPDHS apoenzyme structures (Fig. 5) by superimposing the Cα backbone atoms using CCP4 Superpose (Krissinel and Henrick, 2004). The overall arrangements of helices and strands in these structures were in very good agreement, with an average RMSD of 0.356 Å for all 1320 main chain atoms in the O subunit and 0.373 for the P subunits, not including side-chains since there are sequence differences between the orthologs. The most significant relative displacement between the apoenzyme structures was in the surface loop formed by residues W158-P164 of human tGAPDHS (black arrow in Fig. 5). Four of the seven amino acids in this loop are different in

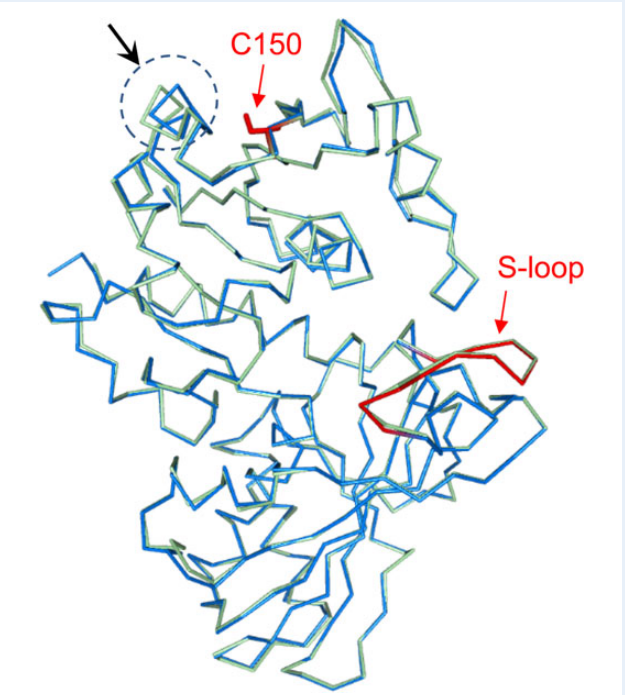


Figure 5 Mouse (green, PDB 5C7I) and human (blue, PDB 5C7L) tGAPDHS apoenzyme structures superimposed along their Cα backbones. The largest displacement between the structures was seen in a surface loop (black arrow) that is not close to the S-loops or the C150 residues that form disulfides in the holoenzyme (highlighted in red).

the human (WRAVGSP) and mouse (WSSIGNP) isozymes. The RMSD of the 28 main chain atoms for this loop is 1.08 Å. The position of this loop is unlikely to alter the novel structural features as described for human tGAPDHS, including potential disulfide bonds and S-loop conformational changes (which occur in the regions highlighted in red in Fig. 5).

Virtual screening to identify compounds predicted to inhibit GAPDHS selectively

With the goal of identifying selective inhibitors of GAPDHS, we used the SYBYL FlexX algorithm to compare docking of compounds from the Maybridge small molecule database into the binding pockets of the sperm and somatic isozymes. The Maybridge collection used for this screen included over 300 000 organic compounds designed to exhibit diverse ring structures with drug-like characteristics. We first prescreened the Maybridge compounds on the basis of size to eliminate compounds that did not fit into the three pockets identified by SiteID. The remaining compounds (up to 8000) were docked with FlexX into each of the pockets, comparing predicted binding in both mouse and human tGAPDHS with the somatic isozyme in each species. Protein-ligand interaction scores were determined for each isozyme, with more negative scores reflecting tighter binding. Discrimination scores were calculated as the difference between the interaction scores for the sperm and somatic isozymes. For compounds with the largest discrimination scores for the human isozymes, docking was repeated

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

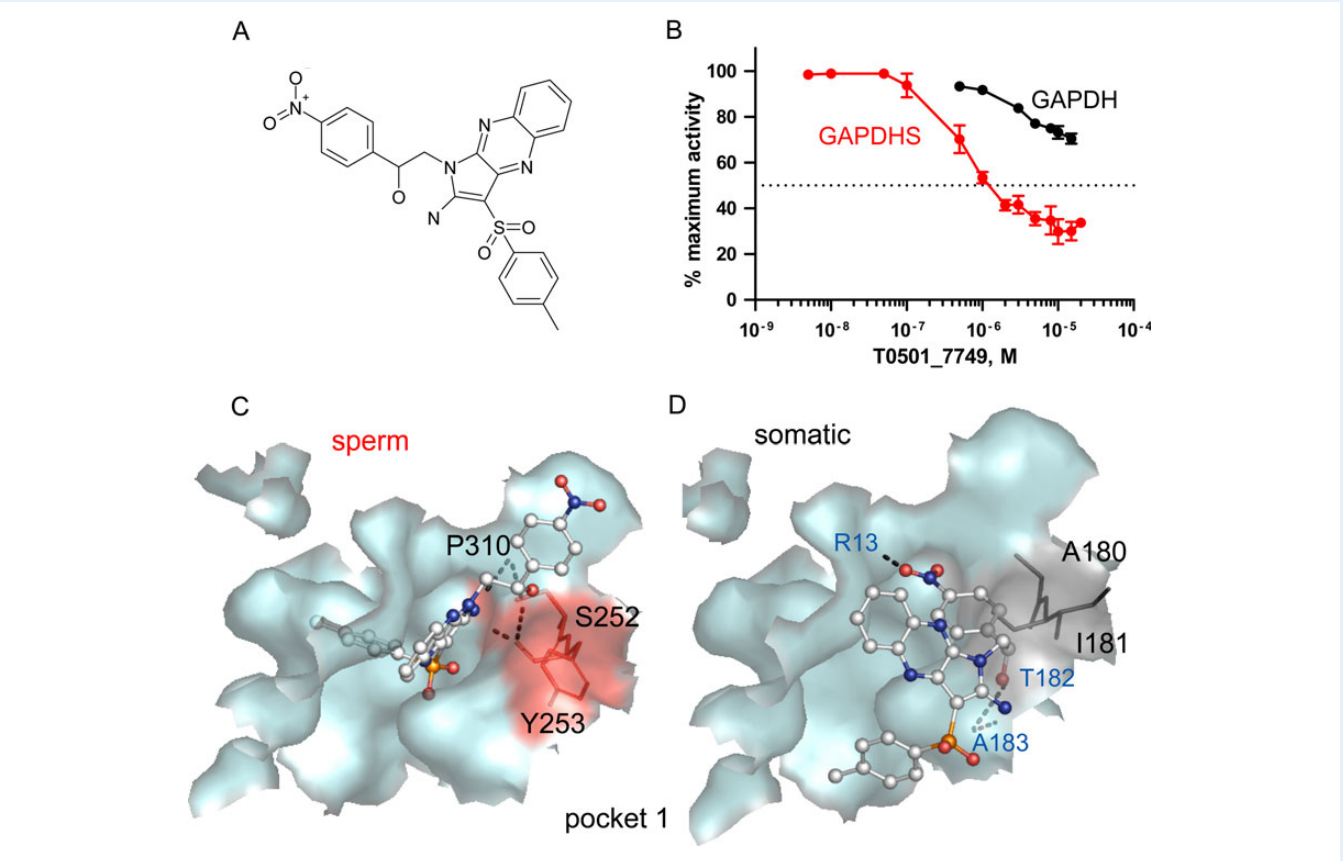


Figure 6 Properties of T0501_7749. (**A**) Chemical structure of this small-molecule inhibitor. (**B**) Dose–response curves monitoring inhibition of human tGAPDHS (red, $n = 3–6$ for all concentrations except first and last, where $n = 2$) and GAPDH (black, $n = 4–16$) in the presence of 2 mM GAP and 2 mM NAD^+. (**C**) FlexX predicted binding pose of T0501_7749 within pocket 1 (blue) of human tGAPDHS, forming hydrogen bonds (dotted lines) with Y253 and P310, which are distinct from the corresponding residues (I181 and A238) in GAPDH. The sperm-specific residues (S252 and Y253) in pocket 1 are highlighted in red. (**D**) In the somatic isozyme, FlexX predicted that T0501_7749 forms hydrogen bonds with three pocket 1 residues that are shared with the sperm isozyme (R13, T182, A183). GAPDH amino acids (A180, I181) corresponding to the sperm-specific residues are highlighted in gray.

thirty times to identify compounds that docked consistently within the predicted binding pocket. Following cluster analysis and visual inspection for drug-like properties and distinct scaffolds, we selected 25 representative compounds for further testing in enzyme assays. These compounds had discrimination scores between −2.2 and −17.6 and were predicted to form hydrogen bonds within one of the pockets of the sperm isozyme (eleven pocket 1, eleven pocket 2 and three pocket 3 compounds).

Testing compounds to compare inhibition of GAPDHS and GAPDH

We tested predicted inhibitors in a spectrophotometric assay of enzyme activity that measures NADH accumulation at 340 nm (Schmalhausen *et al.*, 1997), initially comparing inhibition of mouse tGAPDHS and GAPDH. Two compounds inhibited ~50% mouse tGAPDHS activity at concentrations that did not alter activity of the somatic enzyme, but complete inhibition was not achieved because these compounds exhibited limited solubility at higher concentrations. To estimate IC_{50} values, detailed dose–response assays were conducted comparing inhibition of human tGAPDHS and GAPDH.

T0501_7749 (IUPAC Name: 2-[2-amino-3-(4-methylphenyl)sulfonyl pyrrolo[3,2-b]quinoxalin-1-yl]-1-(4-nitrophenyl)ethanol; structure shown in Fig. 6A) inhibited >60% human tGAPDHS activity at a concentration of 5 μM (Fig. 6B). IC_{50} values, calculated using the four parameter nonlinear model (Prism, Graphpad Software), were 1.2 μM for human tGAPDHS and 38.5 μM for human GAPDH, indicating that T0501_7749 inhibited the sperm isozyme with >10-fold selectivity. FlexX docking predicted that T0501_7749 binds in pocket 1 of human tGAPDHS (Fig. 6C), forming hydrogen bonds with Y253 and P310, two residues that are distinct from the somatic isozyme. T0501_7749 did not form bonds with the corresponding amino acids (I181 and A238) when docked in human GAPDH, interacting instead with pocket 1 residues (R13, T182, A183) that are shared in the sperm and somatic isozymes (Fig. 6D).

T0506_9350 (IUPAC Name: 1-cyclohexyl-3-[4-[(4-methoxyphenyl) sulfamoyl]-2-nitroanilino]urea; structure shown in Fig. 7A) inhibited ~60% human tGAPDHS activity at a concentration of 200 μM (Fig. 7B). Consistent inhibition was observed with this compound only when the temperature of the 30 min pre-incubation period with enzyme was reduced to 4°C. Although T0506_9350 is not a very potent inhibitor, it did exhibit partial selectivity for the sperm isozyme, with estimated IC_{50} values of 95.8 μM for GAPDHS compared with

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

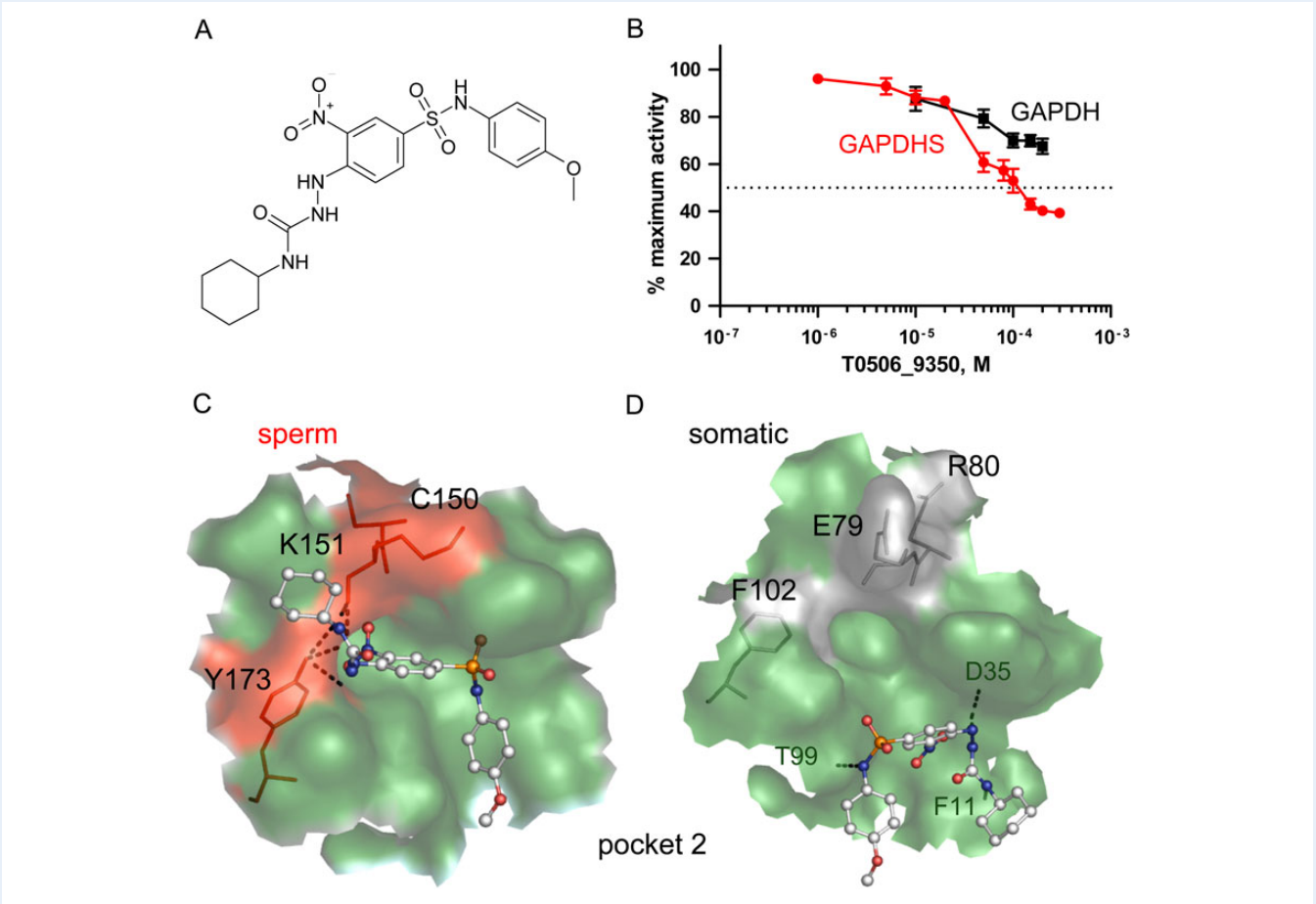


Figure 7 Properties of T0506_9350. (**A**) Chemical structure of this small-molecule inhibitor. (**B**) Dose–response curves monitoring inhibition of human tGAPDHS (red, $n = 4–6$) and GAPDH (black, $n = 5–6$) in the presence of 2 mM GAP and 2 mM NAD^+. (**C**) FlexX predicted binding pose of T0506_9350 within pocket 2 (green) of human tGAPDHS, forming hydrogen bonds (dotted lines) with Y173 and K151. These residues flank the adenine ring of NAD^+ (see Fig. 2B) and are two of the three sperm-specific residues in pocket 2 (highlighted in red). (**D**) FlexX predicted that T0506_9350 binding to GAPDH is removed from the corresponding amino acids (highlighted in gray), forming hydrogen bonds with pocket 2 residues that are shared with the sperm isozyme (F11, D35, T99).

860 μM for GAPDH. FlexX docking predicted that T0506_9350 binds in pocket 2 of tGAPDHS forming five hydrogen bonds with two sperm-specific residues, Y173 and K151 (Fig. 7C). The corresponding residues in GAPDH (F102 and E79) were not predicted to bind to T0506_9350. Instead, docking predicted hydrogen-bond interactions with GAPDH residues (F11, D35 and T99) that are shared with the sperm isozyme (Fig. 7D).

Kinetic analyses were conducted to determine the modes of inhibition and K_i values for T0501_7749 and T0506_9350, and the resulting Lineweaver–Burk plots for human tGAPDHS are shown in Fig. 8. Each double reciprocal line ($1/V_o$ versus $1/[S]$) shows the effects of a single inhibitor concentration with variable concentrations of either GAP (Fig. 8A and C) or NAD^+ (Fig. 8B and D). When GAP concentrations vary, lines representing higher concentrations of T0501_7749 have steeper slopes (K_m/V_{max}) and smaller x intercepts ($-1/K_m$) with little variation in the y intercept ($1/V_{max}$) (Fig. 8A). This pattern of increasing K_m and unchanging V_{max} indicates that T0501_7749 inhibition is competitive with the GAP substrate. In contrast, the double reciprocal lines for T0501_7749 are approximately parallel when NAD^+ concentrations vary (Fig. 8B), reflecting decreases in both V_{max} and K_m. This pattern reflects uncompetitive inhibition, indicating that T0501_7749 binds following formation of the NAD^+-tGAPDHS complex. Mean K_i values for T0501_7749 were 10.9 μM for GAP and 7.1 μM for NAD^+. Identical modes of inhibition were observed in kinetic analyses of mouse tGAPDHS (Supplementary Fig. S1A and B), with mean K_i values of 19.9 μM for GAP and 31.6 μM for NAD^+. Kinetic analyses indicate that T0506_9350 inhibition of human and mouse tGAPDHS is competitive with both GAP (Fig. 8C and Supplementary Fig. S1C) and NAD^+ (Fig. 8D and Supplementary Fig. S1D). Mean K_i values were 63.3 μM for GAP and 60.2 μM for NAD^+ for human tGAPDHS, and 71.0 μM for GAP and 56.9 μM for NAD^+ for mouse tGAPDHS.

GAPDHS inhibitor effects on sperm motility and metabolism

Based on the phenotype of GAPDHS knockout mice (Miki *et al.*, 2004), we expect inhibitors of GAPDHS to impair sperm motility by blocking the glycolytic pathway. Our studies of inhibitor effects on sperm

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

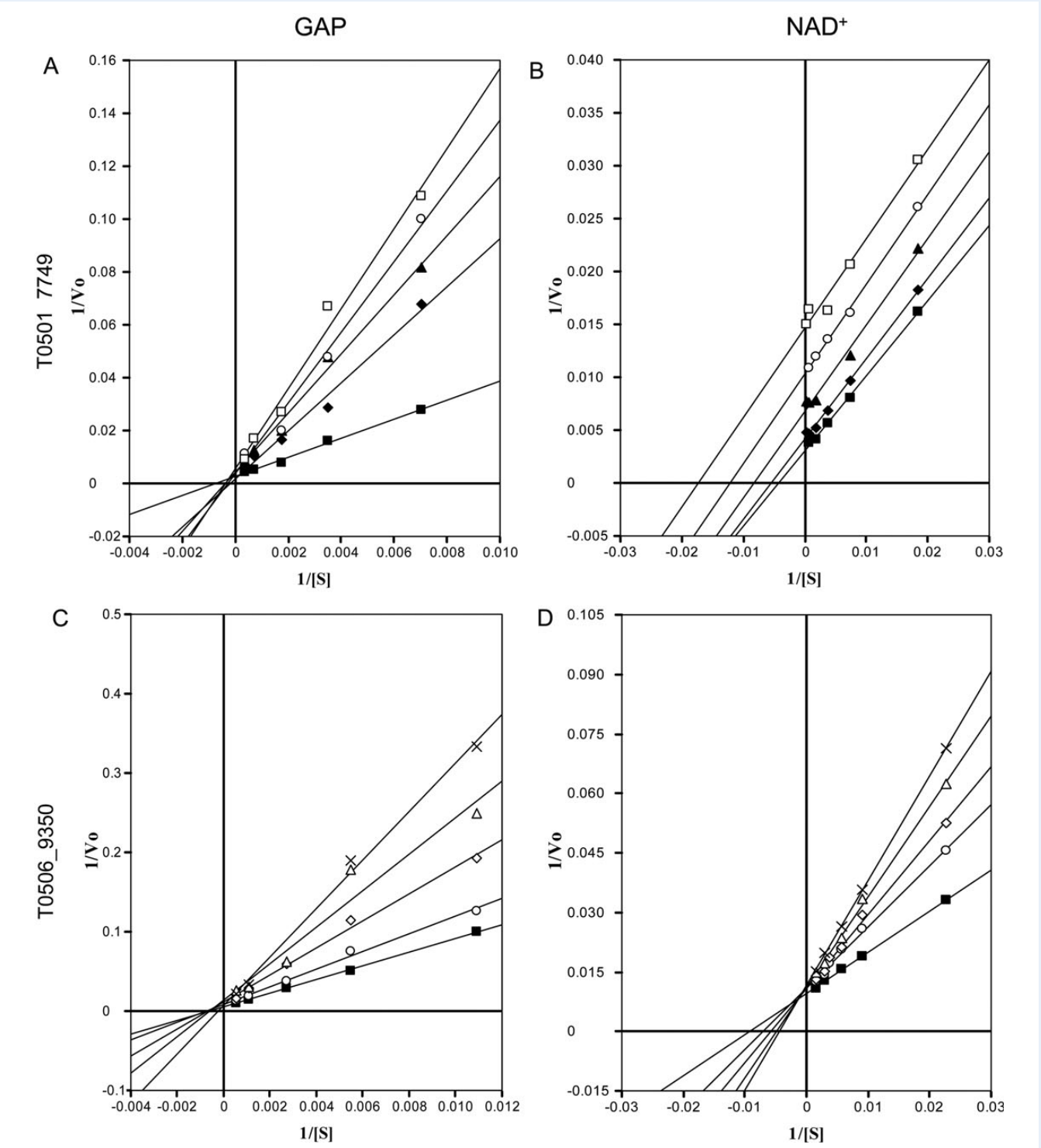


Figure 8 Lineweaver–Burk plots for the kinetic analysis of T0501_7749 (**A** and **B**) and T0506_9350 (**C** and **D**) inhibition of human tGAPDHS. Each line (1/initial velocity V_o versus 1/substrate concentration [S]) shows the effects of a single inhibitor concentration with varying concentrations (0.05–2 mM) of either GAP (A and C) or NAD^+ (B and D). Inhibitors concentrations were: no inhibitor (■), 1 μM (◆), 3 μM (▲),10 μM (○), 20 μM (□), 30 μM (◇), 80 μM (Δ), and 100 μM (x).

focused on T0501_7749, since T0506_9350 had a very high IC_{50} and limited stability above 4°C. Using CASA, we first monitored effects on mouse sperm motility in HTF medium containing 5 mg/ml BSA throughout a 2 h incubation at 37°C, conditions which support capacitation (Goodson *et al.*, 2011). Under these conditions, T0501_7749 did not reduce the percentage of motile sperm. In other preliminary

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

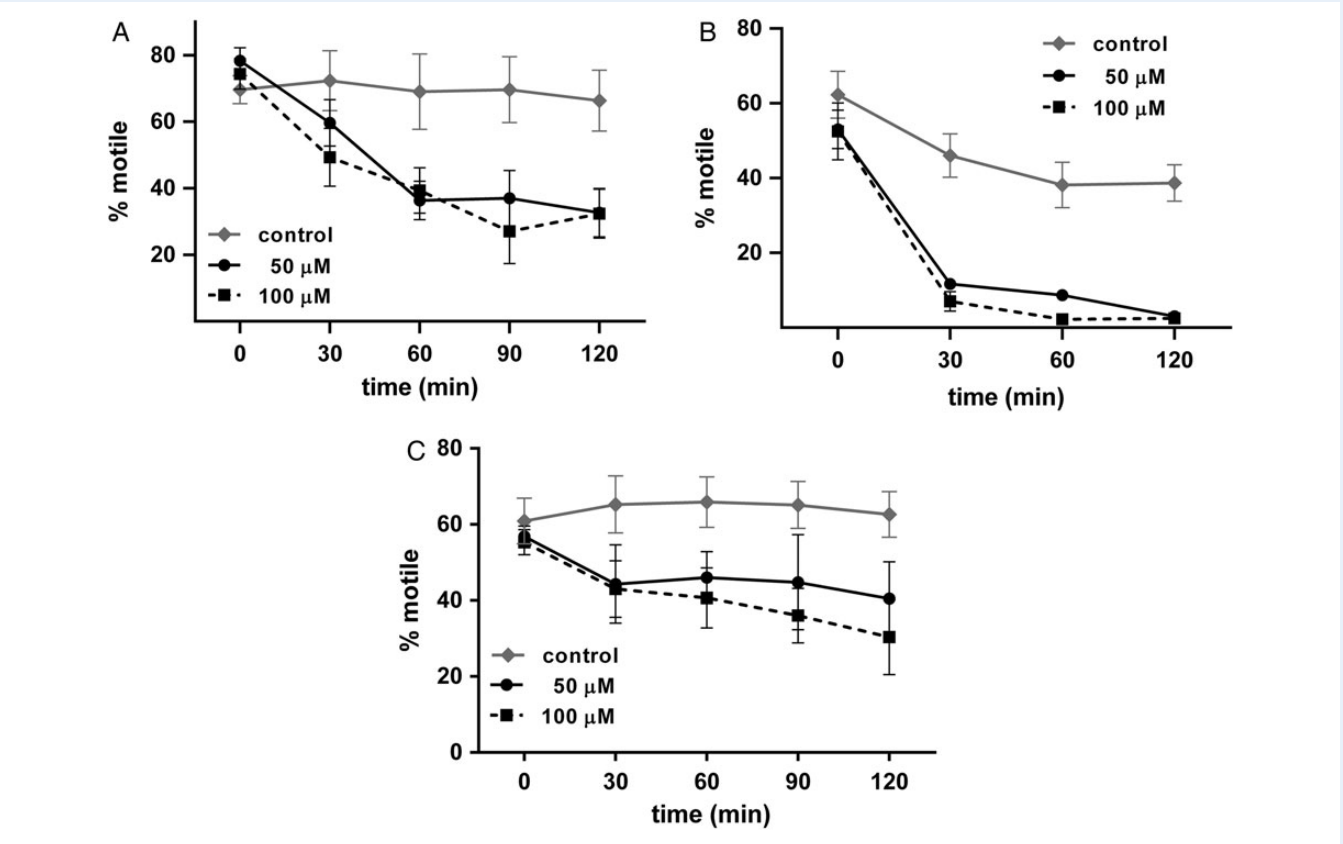


Figure 9 Inhibition of sperm motility by T0501_7749 monitored by computer-assisted sperm analysis. The percentages of motile sperm were determined at 30 min intervals, plotted as mean values ± SEM, with treatments compared by two-way analysis of variance followed by Tukey's multiple-comparison test. Mouse sperm were incubated in (**A**) human tubal fluid (HTF) medium with 0.4 mM methyl-β-cyclodextrin and 0.01% polyvinyl alcohol replacing bovine serum albumin (BSA) or (**B**) HTF with neither these constituents nor BSA. In both media T0501_7749 caused significant reductions in the percentage of motile sperm compared with DMSO vehicle controls. For (A), $n = 3$ and $P \leq 0.03$ for 50 or 100 μM at 60, 90 and 120 min. For (B), $n = 3$ to 7; $P \leq 0.003$ for 50 μM at 30, 60 and 120 min; $P < 0.0001$ for 100 μM at these time points. (**C**) T0501_7749 also significantly reduced the motility of human sperm incubated in HTF without human serum albumin ($n = 8$; P for 100 μM compared with control < 0.05 at 60 and 90 min, < 0.01 at 120 min).

studies, however, we found that T0501_7749 inhibition of GAPDHS enzymatic activity was substantially reduced by the addition of BSA or sperm from GAPDHS knockout mice (not shown). Since these results suggested that nonspecific protein binding to T0501_7749 reduced its effectiveness, we assayed sperm motility in modified HTF medium. T0501_7749 caused significant reductions in the percentage of motile sperm during *in vitro* capacitation when BSA was replaced by 0.4 mM methyl-β-cyclodextrin and 0.01% polyvinyl alcohol (Fig. 9A). More pronounced reductions to <15% motile sperm occurred within 30 min when T0501_7749 incubations were conducted in HTF medium without BSA or replacement compounds that facilitate capacitation (Fig. 9B). After human samples were washed to remove seminal plasma and cryoprotectant, T0501_7749 also significantly reduced the percentage of motile sperm when incubated in HTF without human serum albumin (Fig. 9C). Mean sperm viabilities at 120 min were comparable for control samples (48.0%) and samples treated with 100 μM T0501_7749 (56.7%).

Sperm lactate production and accumulation in the medium during *in vitro* incubations has long been used as an indicator of glycolytic activity (Mann and Lutwak-Mann, 1981). T0501_7749 reduced lactate production by ≥50% ($P = 0.017$ for 100 μM T0501_7749 versus control) when mouse sperm were incubated for 2 h in HTF containing 0.4 mM methyl-β-cyclodextrin, 0.01% polyvinyl alcohol and glucose as the sole substrate (Fig. 10), providing evidence that this compound inhibits glycolysis in the target cell.

Discussion

Our structural analyses of GAPDHS focused on differences between the sperm and somatic isozymes that are conserved between species, with the goal of identifying regions uniquely important for GAPDHS function that are suitable for selective targeting to achieve contraception. We solved the first apoenzyme structures for human and mouse GAPDHS and demonstrated very close alignment between these orthologs. Comparison of our human apo- and holoenzyme structures also identified features not previously observed in GAPDHS or GAPDH crystal structures. Our virtual screen compared the predicted binding of compounds to the sperm and somatic isozymes and identified two small-molecule inhibitors with partial selectivity for the sperm isozyme. Furthermore, T0501_7749, which inhibited GAPDHS with micromolar potency and >10-fold selectivity, exhibited the expected inhibitory effects on sperm motility and lactate production.

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

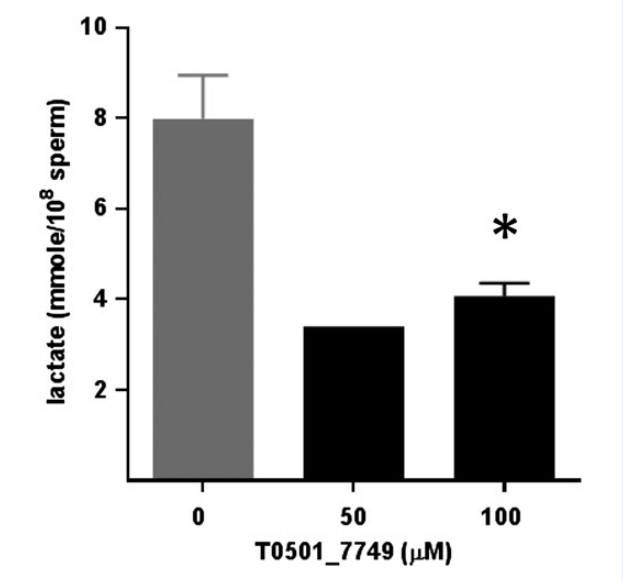


Figure 10 Inhibition of mouse sperm lactate production by T0501_7749. Mouse sperm were incubated for 2 h under the same conditions as in Fig. 9A, except that lactate and pyruvate were omitted from the HTF medium. Data are shown as mean values ± SEM for sperm treated with 50 μM ($n = 1$) or 100 μM T0501_7749 ($n = 3$) compared with control sperm incubated with equal amounts of the DMSO vehicle ($n = 3$). The reduction was statistically significant when the 100 μM and control treatments were compared using a two-tailed unpaired *t*-test (*$P = 0.017$).

In addition to its distinctive N-terminus, GAPDHS has 68 amino acids that are conserved in the human and mouse sperm isozyme, but are different from the corresponding amino acid in the somatic isozymes of both species. Although most of these sperm-specific residues are on the surface, five are clustered in pockets 1 and 2 surrounding the binding sites for GAP and NAD^+. Chaikuad *et al.* (2011) noted that differences in electrostatic properties in these two regions may contribute to kinetic differences between the sperm and somatic enzymes, notably a lower K_m and increased catalytic efficiency for tGAPDHS with respect to NAD^+. Two sperm-specific residues, S252 and Y253, are in pocket 1 surrounding the substrate and the nicotinamide portion of NAD^+. Y253 is predicted to bind to the T0501_7749 selective inhibitor. This residue also provides a potential phosphorylation site that is not present in the somatic isozyme. The sperm-specific K151 and Y173 residues are in binding pocket 2 where they are localized within 2.5 Å of the adenine moiety of NAD^+. In multiple GAPDH orthologs this region is important for cofactor binding and induction of conformational changes that precede catalysis (Skarzynski and Wonacott, 1988; Shen *et al.*, 2000; Yun *et al.*, 2000). Another sperm-specific residue, C150, is located in the same pocket between K151 and Y173, but is oriented in the opposite direction away from NAD^+ (Fig. 2).

Highly conserved in mammalian, reptile and fish orthologs of GAPDHS, C150 is an interesting feature of the sperm isozyme since it mediates disulfide bond formation between adjacent tetramers in our holoenzyme crystal structure (PDB 5C7O). Inter-tetramer disulfide bonds were not observed in the previously published human sperm holoenzyme structure (PDB 3H9E, Chaikuad *et al.*, 2011), with the same space group and very similar unit cell constants. Perhaps differences in protein production and crystallization conditions may have contributed to a slightly more 'oxidized' tGAPDHS in this study, resulting in packing that allows disulfides and the oxidation of the active site cysteine to sulfinic acid. Recombinant tGAPDHS has enhanced stability compared with the somatic isozyme, exhibiting greater resistance to inactivation by heat or guanidine hydrochloride (Elkina *et al.*, 2010; Kuravsky *et al.*, 2014). Furthermore, the native protein with its N-terminal extension is tightly bound to the fibrous sheath in mature sperm (Westhoff and Kamp, 1997; Bunch *et al.*, 1998), persisting when this cytoskeletal structure is isolated by sequential extraction with 1% Triton X-100, 0.6 M potassium thiocyanate and 6 M urea (Krisfalusi *et al.*, 2006). Disulfide bond formation between GAPDHS tetramers may contribute to the stability of this protein and the protein complexes that form the fibrous sheath. Further studies are needed to assess disulfide bond formation between the highly conserved C150 residues of GAPDHS during flagellar formation and during maturation in the epididymis, where disulfide linkages increase in both head and tail proteins of mammalian sperm (Calvin and Bedford, 1971; Balhorn, 2007; Baker *et al.*, 2015).

Although we were unable to obtain crystals of tGAPDHS complexed with T0501_7749, our crystal structures should prove useful in the design and analysis of selective inhibitors. The monoclinic C2 space group of tGAPDHS holoenzyme structures (PDB 5C7O and PDB 3H9E, Chaikuad *et al.*, 2011) is distinct from our apoenzyme structures (PDB 5C7L and 5C7I), which belong to the trigonal $P3_121$ space group and cannot form disulfide bonds between tetramers. Comparison of our apo- and holoenzyme structures confirms that GAPDHS, like GAPDH from multiple species (Skarzynski and Wonacott, 1988; Shen *et al.*, 2000; Yun *et al.*, 2000), undergoes significant conformational change in the N-terminal NAD^+-binding domain when the cofactor binds. Unlike GAPDH in these earlier structural comparisons, GAPDHS also exhibits distinct conformational change in the S-loop region (residues 253–276) of the catalytic domain upon cofactor binding. The maximum displacement in this region occurs at basic residues R265 and K266. The basic residue at position 265 in GAPDHS (R in human or K in mouse) replaces a nonpolar glycine in the somatic isozyme. There are also sperm-specific amino acids (Y253 and H275) at both ends of the S-loop. Previous studies suggest that the S-loops, which extend across the *R*-axis (see Fig. 3) and interact with NAD^+ in the adjacent subunit, may contribute to cooperativity between subunits for cofactor binding (Moras *et al.*, 1975; Biesecker *et al.*, 1977). A recent study found that human GAPDHS exhibits positive cooperativity, opposite to the negative cooperativity observed in mammalian GAPDH (Kuravsky *et al.*, 2015). Although this property was eliminated by mutagenesis to disrupt the D311-H124 salt bridge (Kuravsky *et al.*, 2015), the role of S-loop amino acids in GAPDHS cooperativity has not been explored to our knowledge.

S-loop residues also form one side of the selectivity cleft, which is wider in trypanosomatid GAPDH structures and has been targeted for the design of selective inhibitors to combat illnesses caused by these parasitic protozoa (Verlinde *et al.*, 1994; Suresh *et al.*, 2001). This is an intersubunit cleft across the *R*-axis that is adjacent to the adenosine ribose of NAD^+ in the neighboring subunit. 'Open' (PDB 1ZNQ; Ismail and Park, 2005) and 'closed' (PDB 1U8F; Jenkins and Tanner, 2006) conformations of the S-loop have been reported in human GAPDH structures, although the distance between residues closest to NAD^+ (S-loop P191 and F37 in the

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

neighboring subunit) is similar in these two structures. These residues are conserved in the sperm isozyme (P263 and F108). Only the 'closed' S-loop conformation has been observed in both apoenzyme (human PDB 5C7L and mouse 5C7I) and holoenzyme (human PDB 5C7O and PDB 3H9E, Chaikuad *et al.*, 2011) structures of GAPDHS. The S-loop P263 in our holoenzyme structure (PDB 5C7O) is 3.8 Å from the 2′-hydroxyl of the NAD^+ in the neighboring subunit. Consequently, adenosine analogs designed to fit in the larger trypanosomatid selectivity cleft are unlikely to inhibit the sperm isozyme.

Two compounds predicted to bind more tightly to GAPDHS than to GAPDH in our virtual screen inhibited GAPDHS in enzymatic assays and exhibited partial selectivity for the sperm isozyme. FlexX docking predicted that both of these inhibitors interact with amino acids that are conserved in human and mouse GAPDHS, but are distinct from the corresponding amino acids in the somatic isozymes in both species. Although both compounds exhibited limited solubility, we found that T0501_7749 inhibited human GAPDHS with an estimated IC_{50} of 1.2 μM and >10-fold selectivity. Kinetic analyses with both human and mouse GAPDHS indicate that T0501_7749 competes with GAP substrate binding following formation of the cofactor-enzyme complex.

T0501_7749 appears to bind nonspecifically to BSA and other proteins, hampering full evaluation of its effects on sperm function. Nevertheless, we demonstrated that this compound inhibited both glycolysis and motility in mouse sperm when albumin in the medium was replaced with methyl-β-cyclodextrin and polyvinyl alcohol. The percentage of motile sperm was further reduced to <15% when these additives and BSA were omitted completely. T0501_7749 also caused significant inhibition of human sperm motility, although the effect was not as pronounced. It should be noted that the human sperm used in this study were exposed to proteins from seminal fluid and egg yolk in the cryopreservation medium, potentially modulating the effectiveness of this compound.

This study provides proof-of-principle evidence that GAPDHS can be inhibited with at least partial selectivity and that small-molecule inhibitors of this enzyme significantly reduce sperm glycolysis and motility. We are currently using medicinal chemistry to optimize structure-activity relationships within the T0501_7749 scaffold, with the goal of improving potency, selectivity and drug-like chemical properties. In addition, we will continue to use structure-based cheminformatics approaches to search for additional GAPDHS inhibitors, efforts that are informed by our crystal structures.

Supplementary data

Supplementary data are available at http://molehr.oxfordjournals.org/.

Acknowledgements

The authors thank Vira Ayzenbart, Sanjana Bhat and Patricia Magyar for excellent technical assistance, and Drs. Masuo Goto and Kiyoshi Miki for advice on the production of recombinant proteins. We also gratefully acknowledge the University of North Carolina Andrology Laboratory and Dr. Michael O'Rand for providing human sperm samples used in this study. We thank the Yale Coli Genetic Stock Center for providing the gapA-deficient *E. coli* strain.

Authors' roles

P.V.D. had primary responsibility for all aspects of data acquisition and analysis, and prepared the initial draft of the manuscript. WD.Q. developed the initial constructs and procedures for the expression of recombinant proteins. B.R.T. directed our homology modeling and virtual screening and R.J.R. assisted with the virtual screen. M.J.M., M.M. and L.B. provided guidance for the preparation of tGAPDHS crystals, collecting diffraction data and analyzed the crystal structures. D.A.O. is responsible for the experimental design and supervision of this project, including data analysis and manuscript preparation.

Funding

This work was supported by National Institutes of Health grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development: U01 HD060481 and cooperative agreement U54 HD035041 as part of the Specialized Cooperative Centers Program in Reproduction and Infertility Research. Also supported by National Institutes of Health Fogarty International Center grant D43 TW/HD00627 and subproject CIG-05-109 from CICCR, a program of CONRAD, Eastern Virginia Medical School, USA.

Conflict of interest

None declared.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010;**66**:213–221.

Baker MA, Weinberg A, Hetherington L, Villaverde AI, Velkov T. Analysis of protein thiol changes occurring during rat sperm epididymal maturation. *Biol Reprod* 2015;**92**:11, 1–10.

Balhorn R. The protamine family of sperm nuclear proteins. *Genome Biol* 2007;**8**:227.

Banas T, Malarska A, Krotkiewska B, Marcinkowska A. Glyceraldehyde-3-phosphate dehydrogenase. Investigation on the regions responsible for self-assembly of subunits. *Comp Biochem Physiol B* 1987;**87**:391–401.

Biesecker G, Harris JI, Thierry JC, Walker JE, Wonacott AJ. Sequence and structure of D-glyceraldehyde 3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *Nature* 1977;**266**:328–333.

Brown-Woodman PDC, White IG. Effect of α-chlorohydrin on cauda epididymis and spermatozoa of the rat and general physiological status. *Contraception* 1975;**11**:69–78.

Brown-Woodman PD, Mohri H, Mohri T, Suter D, White IG. Mode of action of α-chlorohydrin as a male anti-fertility agent. Inhibition of the metabolism of ram spermatozoa by α-chlorohydrin and location of block in glycolysis. *Biochem J* 1978;**170**:23–37.

Buehner M, Ford GC, Olsen KW, Moras D, Rossman MG. Three-dimensional structure of D-glyceraldehyde-3-phosphate dehydrogenase. *J Mol Biol* 1974;**90**:25–49.

Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA. Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis. *Biol Reprod* 1998;**58**:834–841.

Calvin HI, Bedford JM. Formation of disulphide bonds in the nucleus and accessory structures of mammalian spermatozoa during

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

maturation in the epididymis. *J Reprod Fertil Suppl* 1971;**13**(Suppl 13): 65–75.

Chaikuad A, Shafqat N, Al-Mokhtar R, Cameron G, Clarke AR, Brady RL, Oppermann U, Frayne J, Yue WW. Structure and kinetic characterization of human sperm-specific glyceraldehyde-3-phosphate dehydrogenase, GAPDS. *Biochem J* 2011;**435**:401–409.

Cowan-Jacob SW, Kaufmann M, Anselmo AN, Stark W, Grutter MG. Structure of rabbit-muscle glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2003;**59**:2218–2227.

Danshina PV, Schmalhausen EV, Avetisyan AV, Muronetz VI. Mildly oxidized glyceraldehyde-3-phosphate dehydrogenase as a possible regulator of glycolysis. *IUBMB Life* 2001;**51**:309–314.

Danshina PV, Geyer CB, Dai Q, Goulding EH, Willis WD, Kitto GB, McCarrey JR, Eddy EM, O'Brien DA. Phosphoglycerate kinase 2 (PGK2) is essential for sperm function and male fertility in mice. *Biol Reprod* 2010;**82**:136–145.

Eddy EM, Toshimori K, O'Brien DA. Fibrous sheath of mammalian spermatozoa. *Microsc Res Tech* 2003;**61**:103–115.

Elkina YL, Kuravsky ML, El'darov MA, Stogov SV, Muronetz VI, Schmalhausen EV. Recombinant human sperm-specific glyceraldehyde-3-phosphate dehydrogenase: structural basis for enhanced stability. *Biochim Biophys Acta* 2010;**1804**:2207–2212.

Ford WC, Harrison A. The activity of glyceraldehyde 3-phosphate dehydrogenase in spermatozoa from different regions of the epididymis in laboratory rodents treated with α-chlorohydrin or 6-chloro-deoxyglucose. *J Reprod Fertil* 1983;**69**:147–156.

Frayne J, Taylor A, Cameron G, Hadfield AT. Structure of insoluble rat sperm glyceraldehyde-3-phosphate dehydrogenase (GAPDH) via heterotetramer formation with *Escherichia coli* GAPDH reveals target for contraceptive design. *J Biol Chem* 2009;**284**:22703–22712.

Goodson SG, Zhang Z, Tsuruta JK, Wang W, O'Brien DA. Classification of mouse sperm motility patterns using an automated multiclass support vector machines model. *Biol Reprod* 2011;**84**:1207–1215.

Goodson SG, Qiu Y, Sutton KA, Xie G, Jia W, O'Brien DA. Metabolic substrates exhibit differential effects on functional parameters of mouse sperm capacitation. *Biol Reprod* 2012;**87**:75, 1–15.

Hasegawa A, Mochida K, Matoba S, Yonezawa K, Ohta A, Watanabe G, Taya K, Ogura A. Efficient production of offspring from Japanese wild-derived strains of mice (*Mus musculus molossinus*) by improved assisted reproductive technologies. *Biol Reprod* 2012;**86**:167, 1–7.

Ijiri TW, Vadnais ML, Lin AM, Huang AP, Cao W, Merdiushev T, Gerton GL. Male mice express spermatogenic cell-specific triosephosphate isomerase isozymes. *Mol Reprod Dev* 2013;**80**:862–870.

Ismail SA, Park HW. Structural analysis of human liver glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2005; **61**:1508–1513.

Jenkins JL, Tanner JJ. High-resolution structure of human D-glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2006; **62**:290–301.

Jones AR. The antifertility actions of α-chlorohydrin in the male. *Life Sci* 1978; **23**:1625–1645.

Jones AR, Cooper TG. A re-appraisal of the post-testicular action and toxicity of chlorinated antifertility compounds. *Int J Androl* 1999;**22**:130–138.

Krisfalusi M, Miki K, Magyar PL, O'Brien DA. Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa. *Biol Reprod* 2006;**75**:270–278.

Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 2004;**60**:2256–2268.

Kuravsky M, Barinova K, Marakhovskaya A, Eldarov M, Semenyuk P, Muronetz V, Schmalhausen E. Sperm-specific glyceraldehyde-3-phosphate dehydrogenase is stabilized by additional proline residues and an interdomain salt bridge. *Biochim Biophys Acta* 2014;**1844**:1820–1826.

Kuravsky ML, Barinova KV, Asryants RA, Schmalhausen EV, Muronetz VI. Structural basis for the NAD binding cooperativity and catalytic characteristics of sperm-specific glyceraldehyde-3-phosphate dehydrogenase. *Biochimie* 2015;**115**:28–34.

Mann T, Lutwak-Mann C. *Male Reproductive Function and Semen*. Berlin, Heidelberg, New York: Springer-Verlag, 1981.

Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nat Med* 2008;**14**:1197–1213.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr* 2007; **40**:658–674.

Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA. Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility. *Proc Natl Acad Sci USA* 2004;**101**:16501–16506.

Moras D, Olsen KW, Sabesan MN, Buehner M, Ford GC, Rossmann MG. Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase. *J Biol Chem* 1975; **250**:9137–9162.

Mukai C, Travis AJ. What sperm can teach us about energy production. *Reprod Domest Anim* 2012;**47**(Suppl 4):164–169.

Nakamura N, Dai Q, Williams J, Goulding EH, Willis WD, Brown PR, Eddy EM. Disruption of a spermatogenic cell-specific mouse enolase 4 (Eno4) gene causes sperm structural defects and male infertility. *Biol Reprod* 2013;**88**:90.

Nascimento JM, Shi LZ, Tam J, Chandsawangbhuwana C, Durrant B, Botvinick EL, Berns MW. Comparison of glycolysis and oxidative phosphorylation as energy sources for mammalian sperm motility, using the combination of fluorescence imaging, laser tweezers, and real-time automated tracking and trapping. *J Cell Physiol* 2008;**217**:745–751.

Nass SJ, Strauss JF (eds). *New Frontiers in Contraceptive Research: A Blueprint for Action*. Washington, DC: National Academies Press, 2004.

Odet F, Duan C, Willis WD, Goulding EH, Kung A, Eddy EM, Goldberg E. Expression of the gene for mouse lactate dehydrogenase C (Ldhc) is required for male fertility. *Biol Reprod* 2008;**79**:26–34.

Pesce MA, Bodourian SH, Nicholson JF. Rapid kinetic measurement of lactate in plasma with a centrifugal analyzer. *Clin Chem* 1975;**21**:1932–1934.

Prins GS, Weidel L. A comparative study of buffer systems as cryoprotectants for human spermatozoa. *Fertil Steril* 1986;**46**:147–149.

Rarey M, Kramer B, Lengauer T. Time-efficient docking of flexible ligands into active sites of proteins. *Proc Int Conf Intell Syst Mol Biol* 1995;**3**: 300–308.

Schmalhausen EV, Muronetz VI, Nagradova NK. Rabbit muscle GAPDH: non-phosphorylating dehydrogenase activity induced by hydrogen peroxide. *FEBS Lett* 1997;**414**:247–252.

Schultz N, Hamra FK, Garbers DL. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci USA* 2003;**100**:12201–12206.

Seidler NW. Basic biology of GAPDH. *Adv Exp Med Biol* 2013;**985**:1–36.

Sexton JZ, Danshina PV, Lamson DR, Hughes M, House AJ, Yeh LA, O'Brien DA, Williams KP. Development and implementation of a high throughput screen for the human sperm-specific isoform of glyceraldehyde 3-phosphate dehydrogenase (GAPDHS). *Curr Chem Genomics* 2011; **5**:30–41.

Shen Y, Song S, Lin Z. Structure of D-glyceraldehyde-3-phosphate dehydrogenase from *Palinurus versicolor* in a tetragonal crystal form. *Sci China C Life Sci* 2000;**43**:96–104.

Skarzynski T, Wonacott AJ. Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *J Mol Biol* 1988;**203**:1097–1118.

Storey BT. Mammalian sperm metabolism: oxygen and sugar, friend and foe. *Int J Dev Biol* 2008;**52**:427–437.

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

Suresh S, Bressi JC, Kennedy KJ, Verlinde CL, Gelb MH, Hol WG. Conformational changes in *Leishmania mexicana* glyceraldehyde-3-phosphate dehydrogenase induced by designed inhibitors. *J Mol Biol* 2001;**309**:423–435.

Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 2004;**60**:2184–2195.

Vemuganti SA, Bell TA, Scarlett CO, Parker CE, de Villena FP, O'Brien DA. Three male germline-specific aldolase A isozymes are generated by alternative splicing and retrotransposition. *Dev Biol* 2007;**309**:18–31.

Verlinde CL, Callens M, Van Calenbergh S, Van Aerschot A, Herdewijn P, Hannaert V, Michels PA, Opperdoes FR, Hol WG. Selective inhibition of trypanosomal glyceraldehyde-3-phosphate dehydrogenase by protein structure-based design: toward new drugs for the treatment of sleeping sickness. *J Med Chem* 1994;**37**:3605–3613.

Welch JE, Schatte EC, O'Brien DA, Eddy EM. Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells. *Biol Reprod* 1992;**46**:869–878.

Welch JE, Brown PL, O'Brien DA, Magyar PL, Bunch DO, Mori C, Eddy EM. Human glyceraldehyde 3-phosphate dehydrogenase-2 gene is expressed specifically in spermatogenic cells. *J Androl* 2000;**21**:328–338.

Westhoff D, Kamp G. Glyceraldehyde 3-phosphate dehydrogenase is bound to the fibrous sheath of mammalian spermatozoa. *J Cell Sci* 1997; **110**:1821–1829.

Williams AC, Ford WC. The role of glucose in supporting motility and capacitation in human spermatozoa. *J Androl* 2001;**22**:680–695.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 2011;**67**:235–242.

World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Geneva: WHO Press, 2010.

Yun M, Park CG, Kim JY, Park HW. Structural analysis of glyceraldehyde 3-phosphate dehydrogenase from *Escherichia coli*: direct evidence of substrate binding and cofactor-induced conformational changes. *Biochemistry* 2000;**39**:10702–10710.

Downloaded from http://molehr.oxfordjournals.org/ at Washington State University Libraries on April 14, 2016

Human Reproduction, Vol.29, No.4 pp. 842–851, 2014
Advanced Access publication on February 2, 2014 doi:10.1093/humrep/deu005

human reproduction

ORIGINAL ARTICLE ***Reproductive genetics***

In vitro screening of embryos by whole-genome sequencing: now, in the future or never?

Raf Winand[1,2,†*], Kristien Hens[3,†], Wybo Dondorp[4], Guido de Wert[4], Yves Moreau[1,2], Joris Robert Vermeesch[5], Inge Liebaers[6], and Jan Aerts[1,2]

[1]Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium [2]iMinds Future Health Department, KU Leuven, Heverlee, Belgium [3]Health, Ethics & Society, Research Institute GROW, Maastricht University, Maastricht, Netherlands [4]Department of Health, Ethics and Society, Research Institutes CAPHRI and GROW, Maastricht University, Maastricht, Netherlands [5]Laboratory of Cytogenetics and Genome Research, Centre for Human Genetics (CME), KU Leuven, Leuven, Belgium [6]Centre for Medical Genetics, University Hospital Brussels, Brussels, Belgium

*Correspondence address. E-mail: raf.winand@esat.kuleuven.be

Submitted on October 9, 2013; resubmitted on November 26, 2013; accepted on December 9, 2013

STUDY QUESTION: What are the analytical and clinical validity and the clinical utility of *in vitro* screening of embryos by whole-genome sequencing?

SUMMARY ANSWER: At present there are still many limitations in terms of analytical and clinical validity and utility and many ethical questions remain.

WHAT IS KNOWN ALREADY: Whole-genome sequencing of IVF/ICSI embryos is technically possible. Many loss-of-function mutations exist in the general population without serious effects on the phenotype of the individual. Moreover, annotations of genes and the reference genome are still not 100% correct.

STUDY DESIGN, SIZE, DURATION: We used publicly available samples from the 1000 Genomes project and Complete Genomics, together with 42 samples from in-house research samples of parents from trios to investigate the presence of loss-of-function mutations in healthy individuals.

PARTICIPANTS/MATERIALS, SETTING, METHODS: In the samples, we looked for mutations in genes that are associated with a selection of severe Mendelian disorders with a known molecular basis. We looked for mutations predicted to be damaging by PolyPhen and SIFT and for mutations annotated as disease causing in Human Genome Mutation Database (HGMD).

MAIN RESULTS AND THE ROLE OF CHANCE: More than 40% of individuals who can be considered healthy have mutations that are predicted to be damaging in genes associated with severe Mendelian disorders or are annotated as disease causing.

LIMITATIONS, REASONS FOR CAUTION: The analysis relies on current knowledge and databases are continuously updated to reflect our increasing knowledge about the genome. In the process of our analysis several updates were already made.

WIDER IMPLICATIONS OF THE FINDINGS: At this moment it is not advisable to use whole-genome sequencing as a tool to set up health profiles to select embryos for transfer. We also raise some ethical questions that have to be addressed before this technology can be used for embryo selection.

STUDY FUNDING: This research was supported by: Research Council KU Leuven (Projects: GOA/10/09 MaNet, KUL PFV/10/016 SymBioSys); Flemish Government: IWT – Agency for Innovation by Science and Technology (Project: O&O ExaScience Life), Hercules Foundation (Project: Hercules III PacBio RS), iMinds Future Health Department (Projects: SBO 2013, Art&D Instance), Flemish tier-1 Supercomputer (Project: VSC Tier 1 Exome sequencing); K.H. was supported by the Centre for Society and Life Sciences (CSG, non-profit organization) (Project number: 70.1.074).

COMPETING INTEREST(S): None of the authors has any conflict of interest to declare.

[†] The authors consider that the first two authors should be regarded as joint First Authors.

© The Author 2014. Published by Oxford University Press on behalf of the European Society of Human Reproduction and Embryology. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

TRIAL REGISTRATION NUMBER: N/A.

Key words: embryo / ethics / PGS / PGD / whole-genome sequencing

Introduction

Genetic testing of preimplantation embryos is a generally accepted approach in the context of preimplantation genetic diagnosis (PGD): patients with a known risk to transmit a specific genetic condition, or with a known chromosomal rearrangement, can opt for PGD to select embryos without the relevant disease-causing mutation. A second application of genetic testing of embryos is preimplantation genetic screening (PGS) for aneuploidy. Although not yet sufficiently proven by randomized controlled trials, this is offered by some centers to subfertile patients undergoing IVF as a treatment of infertility or as part of PGD, with the aim of improving their chances of a successful pregnancy. Whole-genome sequencing (WGS) and analysis, which determines and analyzes the entire DNA sequence of an individual in one procedure, has been performed in single cells and single blastomeres (Navin *et al.*, 2011; Xu *et al.*, 2012; Voet *et al.*, 2013). WGS (followed by a targeted analysis) could be a generic approach for PGD, avoiding time-consuming and labor-intensive customized PGD workups. WGS might also be an elegant alternative method for PGS, since full chromosomal aneuploidies and expected segmental imbalances associated with chromosome rearrangements can be easily identified (Harper and Sengupta, 2012). As WGS costs are now approaching the costs for array-based single-cell PGD or PGS, WGS may become a useful auxiliary technique for embryo testing as already performed in such context (Baslan *et al.*, 2012; Hens *et al.*, 2013; Simpson *et al.*, 2013).

In addition to these applications, WGS could theoretically be used to extend considerably the scope of embryo testing. This would entail a widening of the aims of the procedure. In addition to helping people either to have a child without a specific disorder (as in PGD), or to attempt to increase the chances of a successful IVF pregnancy (as in PGS for aneuploidy), the aim of WGS in embryo testing would be to ensure that children born after IVF or IVF/PGD are free from major disorders. Unlike classical PGD, embryo testing with these aims would be a form of medical screening, as it is a form of indiscriminate genetic testing without clinical data. One of the accepted criteria for responsible screening is that there should be a suitable test (Wilson and Jungner, 1968). This means that both the *analytical and clinical validity* of the test must have been demonstrated. The analytical validity of a genetic test is its ability to determine accurately the genotype of interest. Clinical validity is the accuracy with which the test can then predict a phenotype. If the test performs poorly in these regards, this will adversely affect the *clinical utility* of the screening. This last concept refers to the balance of aim-related advantages and unavoidable disadvantages (drawbacks and costs), and is as such directly linked with the ethical acceptability of screening programs (Sanderson *et al.*, 2005; Dondorp *et al.*, 2010). In this paper, we assess the analytical and clinical validity of WGS-based testing as a necessary (though not a sufficient) condition for the clinical utility and ethical acceptability of extended or comprehensive embryo screening using this technology.

Materials and Methods

We ascertained whether state-of-the-art technology and know-how could adequately distinguish benign polymorphisms from disease-causing mutations and help in deciding which embryo to transfer. We investigated how many apparently healthy adults carry mutations predicted to be damaging or annotated as disease causing—under the rationale that if such mutations were sufficient to cause the severe early-onset phenotype (and therefore predictive in a screening setting), they should be absent from the exome of apparently healthy adults.

Disease selection

We obtained the Online Mendelian Inheritance in Man (OMIM; McKusick-Nathns Institute of Genetic Medicine, 2012) list of diseases with a known molecular basis, consisting of >3000 diseases, from the OMIM website. This list was cross-referenced with the Human Phenotype Ontology database (Robinson and Mundlos, 2010) to provide an overview of diseases and their associated phenotypes. The resulting set contained 2172 diseases from which we selected diseases characterized by dysmorphology and early-onset symptoms. By selecting early-onset disorders we ensure that individuals affected by one of these disorders would already show symptoms at the time of sequencing. In addition, only diseases that had the inheritance annotated in OMIM as 'autosomal dominant', 'autosomal dominant type; high penetrance' and 'autosomal recessive' were selected. As a result 132 autosomal dominant and 215 autosomal recessive diseases were retained. The complete list of these diseases can be found in Supplementary data, Tables SI and SII.

Samples

For our analysis, we used both private and publicly available samples from adult individuals who were considered healthy at the time of sequencing (i.e. who did not exhibit clear signs of a congenital disorder at the time of sampling). We had access to 42 exome sequences from in-house research samples. These are trio samples where the parents are considered healthy, as they are symptom free. The sequences of affected children were also available but were not used in the initial analysis. In addition to our high-quality exome sequences, we downloaded two freely available sets from the 1000 Genomes project (1000 Genomes Project Consortium, 2012) (1kG) and Complete Genomics (Drmanac *et al.*, 2010) (CG). The 1kG data came from the phase 1 integrated release version 3 from 30 April 2012, containing genotype calls for >1000 individuals. The data from CG were from 69 individuals, the sample names of which can be found in Supplementary data, Table SIII. All these sequences are from people who do not express symptoms of the selected disorders and therefore considered 'healthy'. Because some of the samples of these publicly available sets are from trios, we removed the samples from related individuals from the data sets leaving 1004 samples from the 1kG data and 50 from the CG data.

Transcripts

To localize the mutations in the genes and the different transcripts of those genes, we used the Ruby Ensembl API (Strozzi and Aerts, 2011) to connect to the Ensembl core database version 70. In addition, we also

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

looked at transcripts present in the Consensus CDS Project (CCDS) database Hs37.3 (Pruitt et al., 2009).

Mutations predicted to be damaging

When analyzing genomes in the search of causative mutations, prediction algorithms are often used to predict the effect of a mutation on the protein. We used an in-house database called Annotate-it (Sifrim et al., 2012) that holds detailed information on the selected genes, to retrieve prediction scores from SIFT (Ng, 2003) and PolyPhen (Adzhubei et al., 2010) for the identified mutations. More information about these algorithms can be found in the Supplementary data. Because PolyPhen and SIFT only give scores to missense mutations, we considered nonsense and splice site mutations also to be damaging in the analysis. Indels were not included in the analysis.

Mutations described to be damaging in literature

In addition to the prediction algorithms, we also looked for mutations that are described as disease causing in the literature. For information on these mutations, we used the Human Gene Mutation Database (HGMD) containing >70 000 disease-causing mutations (Stenson et al., 2003). For our analysis, we only selected those mutations with associations not considered tenuous by the curators of the HGMD and in the disease-causing category 'DM'.

Results

The results from this analysis show that many healthy individuals have mutations predicted to be damaging or annotated as disease causing in HGMD in genes associated with severe developmental disorders. For the 1kG samples all mutations are included, i.e. mutations that were found in both the low coverage and exome sequences. For the in-house data sets we only retained mutations with a read depth of at least 30× and phred score of 30. Relaxing these constraints leads to a higher number of mutations, which are described in Supplementary data, Tables SIV and SV.

Autosomal dominant disorders: mutations predicted to be damaging

When looking for mutations predicted to be damaging, we found that depending on the data set, 98–100% of healthy individuals had damaging mutations in genes associated with the selected autosomal dominant disorders. In the 1kG data set, we found a median of 8 mutations (min. 3, max. 14), in the Complete Genomics data set a median of 9 mutations (min. 4, max. 14) and in the in-house data set a median of 2 mutations (min. 0, max. 5) per individual. In only one sample from the in-house data set no damaging mutations were found. The distribution of the number of mutations can be seen in Fig. 1. Because variants that are frequently found in the population are unlikely to cause these severe Mendelian disorders, we only retained the variants with a minor allele frequency (MAF) in the 1000 genomes of <1%. Applying this filter leads to a large decrease in identified variants but still 40–94% of individuals were found to carry damaging mutations. In the 1kG data set we found a median of 0 (min. 0, max. 5) mutations, in the Complete Genomics data set a median of 2 mutations (min. 0, max. 6) and in the in-house data set a median of 0 mutation (min. 0, max. 4) per individuals. The distribution of the number of mutations can be seen in Fig. 2.

As explained by MacArthur et al. (2012) faulty gene annotation is a likely cause for these genes containing a large number of deleterious mutations. An example of this can be found in EDARADD (Ectodysplasin-A receptor-associated death domain), which is the gene with the highest number of unfiltered damaging mutations in all data sets. Mutations in EDARADD cause the autosomal dominant form of ectodermal dysplasia, characterized by sparse hair, missing or abnormal teeth and the inability to sweat (Cluzeau et al., 2011). For this gene we found mutations in 83–98% of the samples. These mutations however are not annotated as disease causing in HGMD. In the 1kG data the most occurring mutation in EDARADD is at chromosome 1, position 236557771 G>A (dbSNP id: rs966365). So even though it is predicted to be damaging by both SIFT and PolyPhen, it is frequently found in that population (91%) suggesting an annotation error in the reference genome. Another observation we made was that most of the genes with the most common mutations have multiple transcripts and most have at least one transcript that is not affected by these mutations.

After filtering out the variants with a MAF in the 1000 genomes of >1%, we find differences in the percentage of samples showing mutations between the data sets. For instance, the most occurring variant in the in-house data set was found in NOTCH2, which is associated with Hajdu–Cheney syndrome, a disease characterized by coarse face, short neck, hirsutism, joint laxity and bone dysplasias (Ramos et al., 1998). In this case, mutations were found in 14% of the in-house samples, in 1% of the 1kG samples and 6% of the CG samples. The fact that a predicted damaging variant is frequently found in the local population but not in the 1000 genomes data set and dbSNP can indicate that the variant is a benign polymorphism in this population. An example of this is a variant that we identified at chromosome 3, position 98300354 (A>C) in CPOX, which is associated with hereditary coproporphyria. This mutation is found in almost 10% of the in-house samples but does not occur in the other data sets and is not found in dbSNP.

In total we identified 323, 120 and 21 distinct mutations with a MAF of <1% in the 1kG data in 69, 59 and 18 genes in, respectively, the 1kG, CG and in-house data sets. An overview of the genes with the largest number of predicted damaging mutations and some phenotypic information about the disorder is shown in Supplementary data, Tables SVI and SVII.

Autosomal dominant disorders: mutations present in HGMD

Taking only the mutations into account that are annotated as disease causing in HGMD for the severe disorders from our list, we identified mutations in 20% of the 1kG samples, 22% of the Complete Genomics samples and 12% of the in-house samples. The number of affected samples with curated disease-causing variants is thus much lower compared with those identified by the prediction programs. When looking at the in-house samples, we found that the mutations were considered to be damaging by either PolyPhen or SIFT but not by both. A list of the diseases of which causative mutations were identified can be found in Supplementary data, Table SVIII.

In the in-house data, we identified a total of six heterozygous mutations in five individuals. One of these mutations (CM023740) had a MAF >1% in the 1000 genomes indicating a possibly suspicious annotation or a variant with reduced penetrance. At the time of the initial analysis, this mutation was annotated as causing extrahepatic biliary atresia, a

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

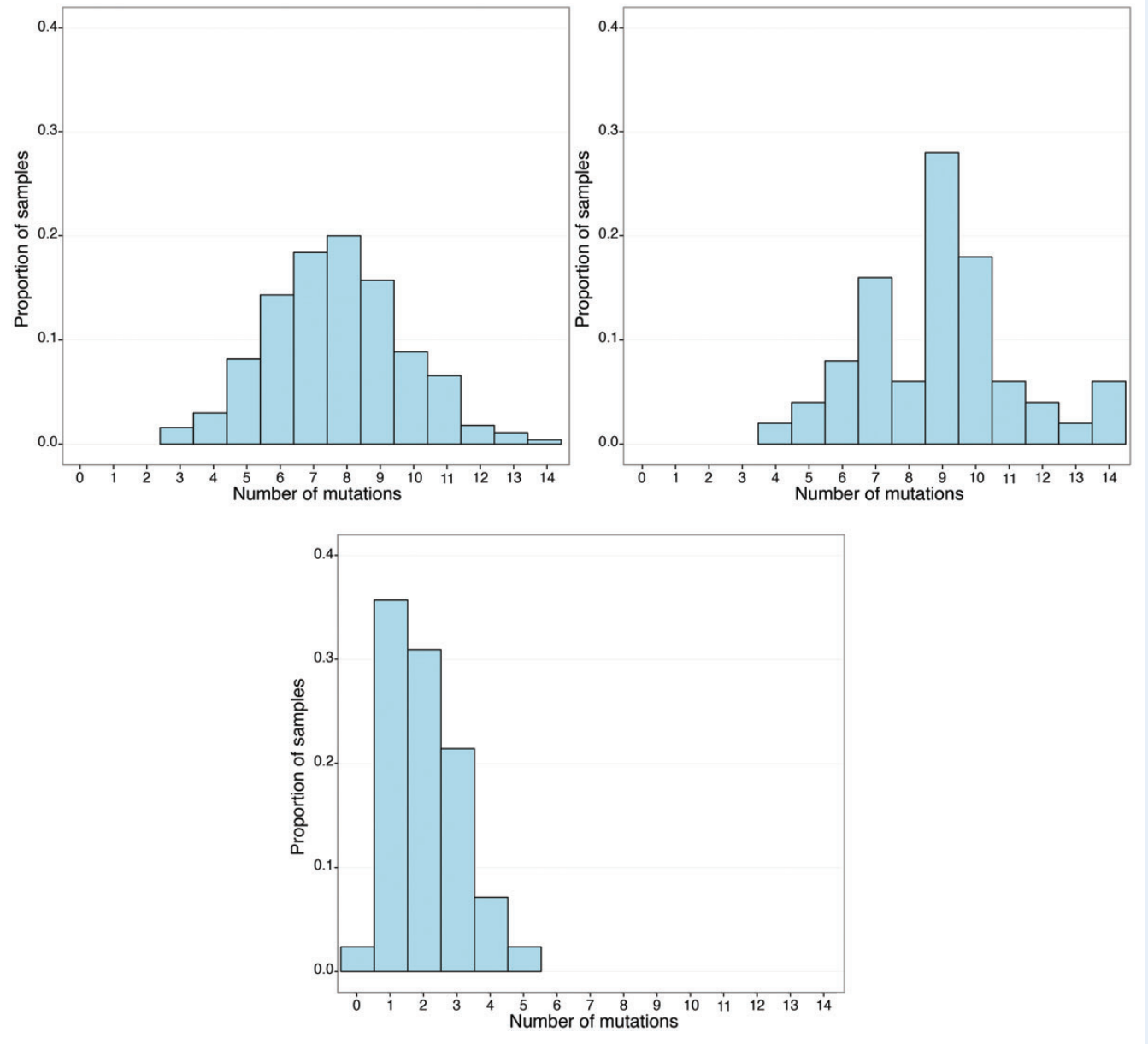


Figure 1 Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging. Samples are from the 1000 genomes, Complete Genomics and in-house (clockwise starting top left).

feature of Alagille syndrome, but it was removed from HGMD at a later time. For Greig cephalopolysyndactyly syndrome two distinct mutations were found in GLI3, i.e. P707S (CM970684) (Wild *et al.*, 1997) and I808M (CM990707) (Kalff-Suske *et al.*, 1999). In a functional analysis, both mutations were found to cause misregulation of GLI3-localization by Krauss *et al.* (2009). The mutation causing Rubinstein-Taybi syndrome—A981T (CM021081) in CREBBP—was identified by Coupry *et al.* (2002) in a set of 60 patients. Because these samples are part of trios, we also had access to the samples of the children. Out of a total of five children, four children were heterozygous for the same mutation as their parent(s) but also did not express the disease. The fact that these variants are present in apparently healthy individuals may hint towards (i) sequencing errors, (ii) false-positive entries in HGMD or (3) incomplete penetrance of certain variants that would make them of low predictive value in a PGS context.

Autosomal recessive disorders: mutations predicted to be damaging

For autosomal recessive disorders, there are two categories of affected individuals. Either they are homozygous for a mutation or they are compound heterozygous. We found almost the same number of samples that were homozygous for damaging mutations in genes associated with autosomal recessive disorders as we found mutations for the autosomal dominant disorders. Approximately 98–100% of samples were homozygous for at least one mutation predicted to be damaging. The number of

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

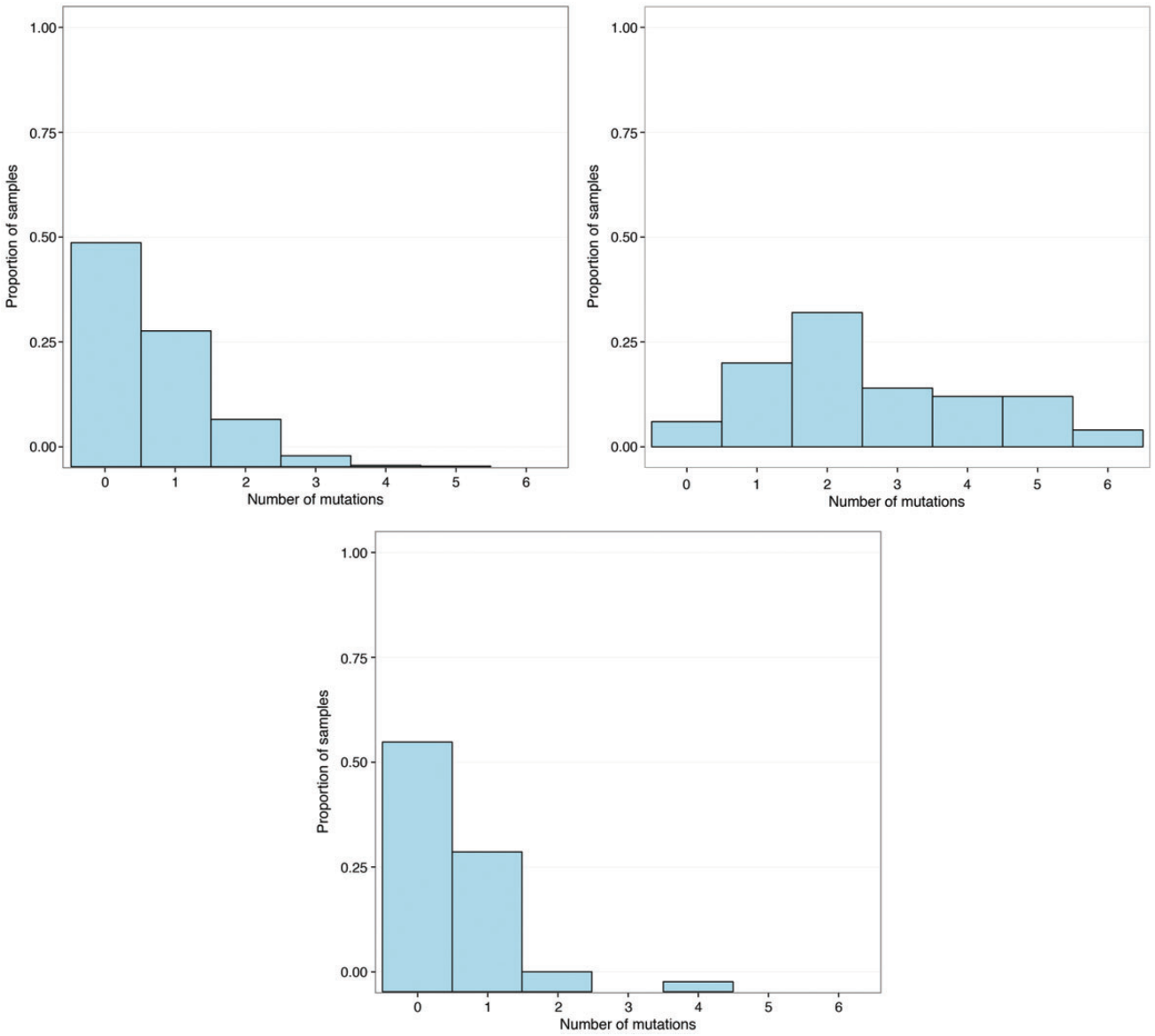


Figure 2 Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging and are found with a minor allele frequency of $<1\%$ in the 1000 genomes data. Samples are from the 1000 genomes, Complete Genomics and in-house (clockwise starting top left).

damaging mutations per individual was lower however. In the 1kG data set we found a median of 4 mutations (min. 0, max. 9), in the Complete Genomics data set a median of 4 mutations (min. 1, max. 7) and also in the in-house data set a median of 4 mutations (min. 0, max. 7) per individual. The distribution of the mutation counts can be seen in Fig. 3. Limiting the variants to those with a MAF of $<1\%$ in the 1000 genomes data produced a large decrease especially in the 1kG data set. Less than 2% of the samples are homozygous for a damaging mutation in the 1kG data set and 56–69% in the CG and in-house data set. The median in the 1kG data sets is 0 mutation (min. 0, max. 1), in the Complete Genomics data set 1 mutation (min. 0, max. 2) and in the in-house data set 1 mutation (min. 0, max. 1). The distribution of the mutation counts can be seen in Fig. 4. In total, we identified 17, 4 and 1 distinct homozygous mutations with MAF in 1kG of $<1\%$ in 16, 4 and 1 genes in, respectively, the 1kG, CG and in-house data set. As for the autosomal dominant disorders, Supplementary data, Tables SIX and SX show an overview of the genes with the largest number of mutations and the corresponding diseases.

In addition to homozygosity, we also investigated the case of compound heterozygosity. Because no haplotype information was available for the in-house data set, we identified the number of samples that have two different mutations predicted to be damaging in each gene. In this case, the difference between the data sets is also large. While for some diseases a number of individuals are identified in one data set, there are no individuals with damaging mutations in another and vice versa. In this case, the gene NEB, which is associated with nemaline myopathy, showed the highest number of individuals that had at least two

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

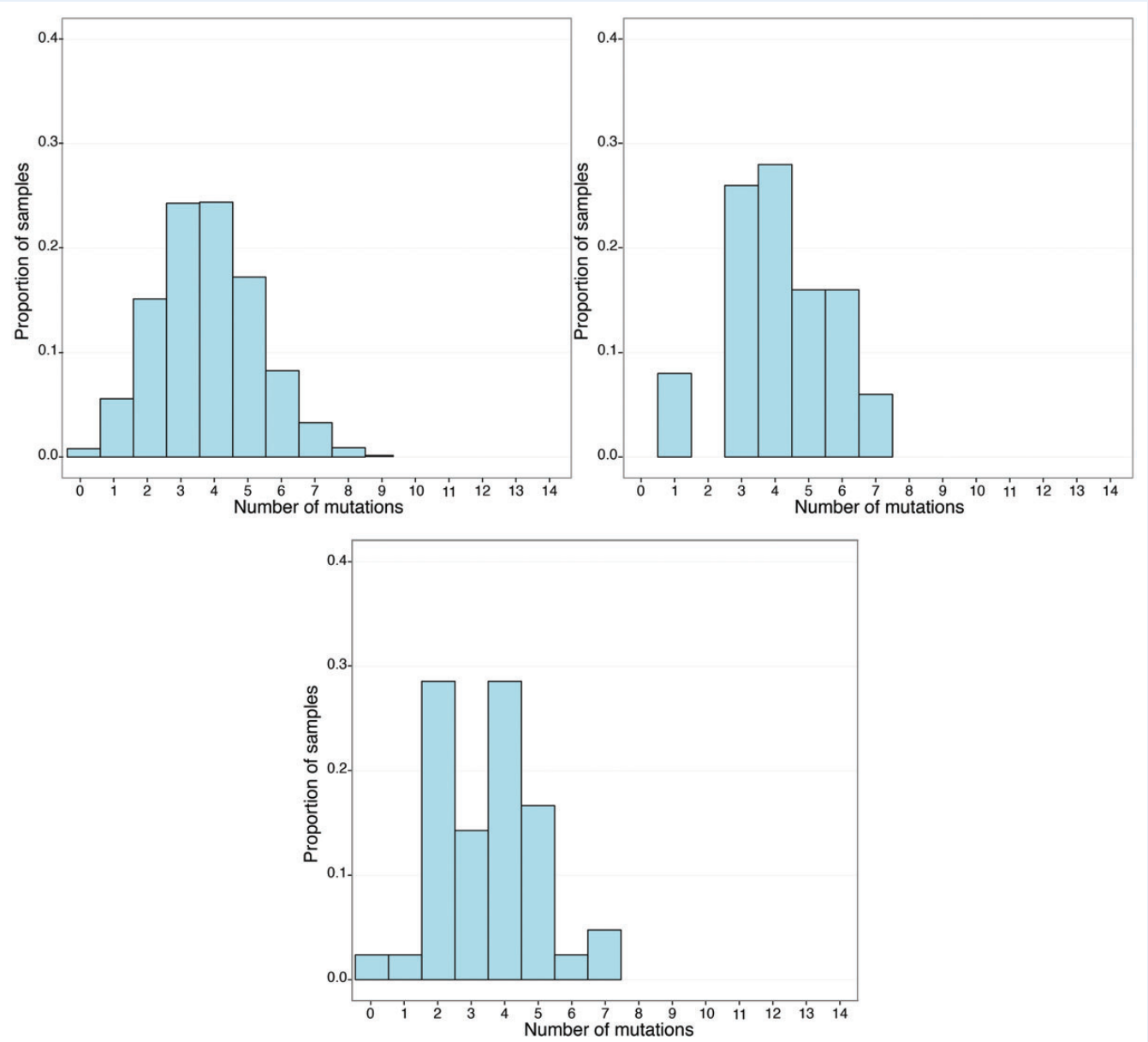


Figure 3 Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging. Samples are from the 1000 genomes, Complete Genomics and in-house (clockwise starting top left). Only homozygous mutations are counted.

distinct mutations with 2% of samples in the 1kG data set, 2% in the CG data set but no samples in the in-house data set affected. An overview of the number of affected samples and genes can be found in Supplementary data, Table SXI.

Autosomal recessive disorders: mutations present in HGMD

For HGMD mutations, we found that the percentage of samples that are homozygous and therefore are expected to express the disease is much lower than with autosomal dominant diseases. In the in-house data set, no homozygous HGMD mutations were found. An overview of all affected genes can be found in Supplementary data, Table SXII. The most frequently identified mutation was found in 16 individuals in the 1kG data (1.8%) on chromosome 17, position 7915912 C>T (dbSNP id: rs34598902). This mutation was thought to be causative for Leber Congenital Amaurosis (Zernant *et al.*, 2005), a disease characterized by congenital blindness that affects 10–20% of all blind children (INSERM, 1997). However, it was later found in the normal population (Ito, 2004) with a MAF in the 1000 genomes of 8.26%. While the mutation was annotated as disease causing in HGMD at the time of our analysis, it has been reclassified at a later time to the category 'DM?' indicating a tenuous association. Another mutation in the same gene at position 7912879 C>T (dbSNP id: rs28743021) was found in two individuals in the 1kG data who were homozygous for this mutation. This mutation is annotated as disease causing in HGMD based on a study by Koenekoop *et al.* (2002).

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

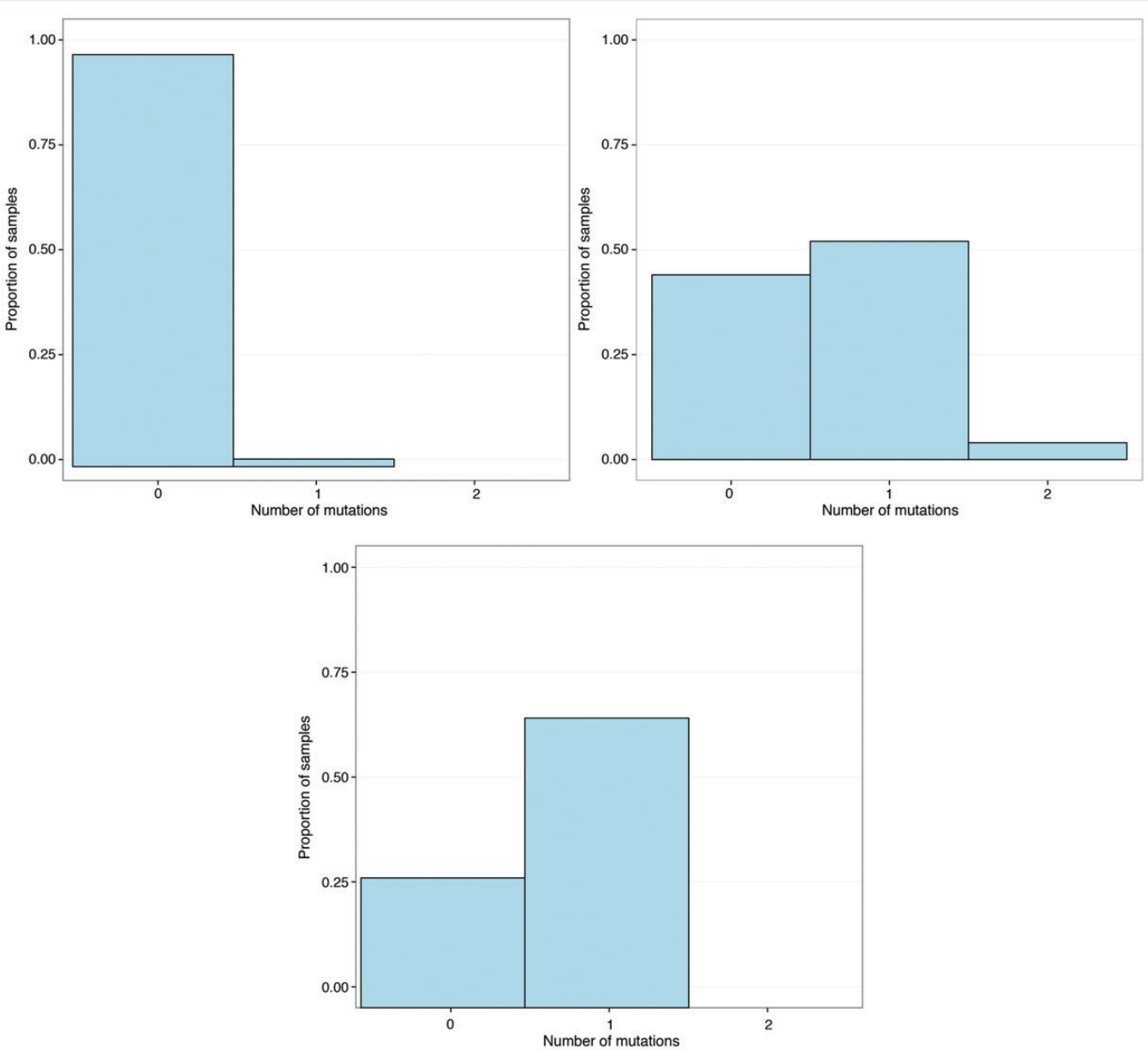


Figure 4 Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging and are found with a minor allele frequency of < 1% in the 1000 genomes data. Samples are from the 1000 genomes, Complete Genomics and in-house (clockwise starting top left). Only homozygous mutations are counted.

For compound heterozygous mutations, again counted as two distinct mutations, no positive samples were identified in the in-house data, only one in the CG data set with a MAF > 1% and a few in the 1kG data set. An overview of the number of affected samples is shown in Supplementary data, Table SXIII.

Discussion

Analytical validity

The analytical validity of WGS-based embryo testing is constituted by the quality of single-cell sequencing and the accuracy of interpretation. At present, single-cell genome-wide sequencing is not as good as sequencing based on multiple cells. On the assumption that this limitation will be overcome, we focus here on the accuracy of the interpretation of sequencing data.

We found that with increasing quality thresholds the number of individuals that carry mutations predicted to be damaging decreases in the samples from our in-house data set. Except for a common mutation found in the CG and in-house data set, we found fewer individuals with mutations predicted to be damaging in genes associated with autosomal recessive disorders. This may be because homozygous mutations occur less frequently than heterozygous mutations or it may be linked to the clinical validity of the test: autosomal dominant disorders are more

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

likely to show reduced penetrance or variable expressivity and therefore symptoms of the disorder might not be present in these individuals although the mutations are.

Clinical validity

More than 40% of genomes of healthy individuals in our study had a genetic mutation thought to be causative for severe autosomal dominant congenital disorders. Moreover, some healthy individuals were homozygous or compound heterozygous for mutations associated with autosomal recessive disorders. These results are in line with the results of Xue *et al.* (2012) and MacArthur and Tyler-Smith (2010) who studied loss-of-function and disease-causing variants in healthy individuals. Our analysis relies on current prediction programs and databases. It is obvious from this analysis that current programs predicting protein damage based on exonic sequence information alone show a number of false positives and will need to improve to enable proper future phenotype prediction (Sifrim *et al.*, 2013). In our results we see that filtering the identified mutations on the MAF in the 1000 genomes significantly reduces the number of identified mutations per individual. In the CG and in-house data set however, a larger percentage of individuals still show mutations. This suggests that these mutations might be local polymorphisms that are not identified in the 1000 genomes project. An in-depth knowledge of the variants found in the local population can therefore lead to improved filters. Finally, the HGMD database is frequently updated and thus the results of each analysis might be different with each version of the database.

A specific challenge for embryo screening is that predictions have to be made in the absence of phenotypic information. This is different from the use of WGS-based testing in the post-natal context where this technology is being introduced as a means of finding a diagnosis for existing patients with a phenotype that could not be clarified using more traditional diagnostic approaches. However, in embryo selection, genetic information is all that is available, except for classical PGD where the phenotype of the parents can be taken into account. We considered to what extent adding WGS-based preconception testing of the prospective parents might help fill in this lacuna by providing additional context information. This approach may perhaps be helpful where findings related to dominant disorders are concerned, as information about genotype and phenotype of the parents will contribute to making better predictions about the health of children resulting from embryos with the relevant mutations found in WGS-based screening. However, preconception testing of the prospective parents seems less useful for the interpretation of findings related to recessive disorders, given our observation that healthy persons may be compound heterozygous or even homozygous for mutations in this category.

Advancing knowledge in genomics may enhance the clinical validity of WGS-based embryo testing, as more will be known about the influence of modifier and protective genes and possible epigenetic influences that may explain our findings (von Kanel *et al.*, 2013). Also, other projects, such as the 'Deciphering Developmental Disorders' (Firth and Wright, 2011) study, may contribute to identifying the genetic cause of disorders, making knowledge about the genome more complete. Although it can be expected that with increasing up-to-date knowledge about the relation between the genome and the phenotype it will become possible in the future to make better predictions about the health of children resulting from the transfer of an embryo with a specific genotype, we conclude that at present the clinical validity of WGS-based embryo screening is limited.

Clinical utility

As the analytical and clinical validity are still insufficient, a necessary condition for the introduction of extended or comprehensive embryo screening ('suitable test') is not met. Clearly, this would adversely affect the *clinical utility* of the screening, as it would lead to discarding embryos that may well develop into healthy children. Some may still defend the rapid introduction of WGS-based screening, arguing that mutations will be found that are thought to be causative at least in some cases, and that non-transfer of embryos with such mutations may still be warranted. However, as the number of available good-quality embryos in an IVF cycle is limited, there may not be much scope for choosing a mutation-free embryo, and the option of a new cycle just to avoid a 'suspected' embryo may well be disproportional. Moreover, with the current state of the art, prospective parents would be faced with choices based on unreliable predictions about the health of the children they could have as a result of transferring this or that embryo. Overestimation of the predictive value of adverse findings of WGS-based embryo screening may lead to the couple remaining childless or to undermining of their confidence in the health of any children they may still decide to have. In addition to this, they may also be confronted with equally unreliable information suggesting that they themselves are carriers of a potentially severe disease.

Conclusion and final remarks

At present, the drawbacks of WGS-based embryo screening appear to outweigh the possible benefits for prospective parents, making the introduction of such screening in clinical practice unwarranted and at best premature. It may be that further scientific developments will lead to improving the predictive accuracy of WGS-based embryo screening. Although that would take away the drawback of decision-making based on unreliable information, it does not automatically follow that WGS-based screening would then be unproblematic.

As screening aimed at simultaneously excluding a more than a limited number of genetic risk factors would very soon run into the problem of leaving no embryo for transfer (Hens *et al.*, 2012), a possible approach would be to always select the embryo(s) with the best health profile (while maintaining a threshold of at least not transferring high-risk embryos). However, the clinical utility of that approach would depend on whether meaningful choices between embryos with various health profiles can indeed be made. To say the least it is not clear whether that is the case. And even if it were, the amount of relevant data and the fact that (where genetic susceptibilities are concerned) even reliable information would be about risks rather than certainties, the feasibility of well-considered decision-making would not be obvious. Moreover, whose decisions should this be? As is it can be argued that both prospective parents (whose child it will be) and professionals (given their active involvement in the creation of the child) have a say in this matter, WGS-based screening may have the potential of leading to conflicts between those stakeholders. Last but not least, a difficult problem is that choosing the embryo with the best profile will inevitably mean that children are born for whom some health prospects might already be known. The ethical question here is whether the future child should

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

be allowed to decide for herself what she wants to know about her genome (Hens et al., 2013).

We conclude that even if current limitations in terms of analytical and clinical validity can be overcome, the notion of WGS-based embryo screening still raises some difficult questions. It is presently unclear whether these can be satisfactorily answered. A possible alternative approach with its own advantages and disadvantages (De Wert, 2009) would consist of preconception screening of the prospective parents followed by targeted PGD.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Acknowledgements

We thank the reviewers for their helpful suggestions.

Authors' roles

All authors participated equally in study design. The list of severe disorders was generated by I.L. and R.W. Data analysis was performed by R.W. Input on the ethics debate was provided by K.H., G.d.W. and W.D. The manuscript was drafted by K.H. and R.W. All authors revised the manuscript and approved the final version. The work was supervised by J.A.

Funding

This research is supported by: Research Council KU Leuven (Projects: GOA/10/09 MaNet, KUL PFV/10/016 SymBioSys); Flemish Government: IWT - Agency for Innovation by Science and Technology (Project: O&O ExaScience Life), Hercules Foundation (Project: Hercules III PacBio RS), iMinds Future Health Department (Projects: SBO 2013, Art&D Instance), Flemish tier-1 Supercomputer (Project: VSC Tier 1 Exome sequencing); K.H. was supported by the Centre for Society and Life Sciences (CSG, non-profit organization) (Project number: 70.1.074)

Conflict of interest

None of the authors has any conflict of interest to declare

References

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–249.

Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, Troge J, Ravi K, Esposito D, Lakshmi B *et al.* Genome-wide copy number analysis of single cells. *Nat Protoc* 2012;**7**:1024–1041.

Cluzeau C, Hadj-Rabia S, Jambou M, Mansour S, Guigue P, Masmoudi S, Bal E, Chassaing N, Vincent M-C, Viot G *et al.* Only four genes (EDA1, EDAR, EDARADD, and WNT10A) account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases. *Hum Mutat* 2011;**32**:70–72.

Coupry I, Roudaut C, Stef M, Delrue MA, Marche M, Burgelin I, Taine L, Cruaud C, Lacombe D, Arveiler B. Molecular analysis of the CBP gene in 60 patients with Rubinstein-Taybi syndrome. *J Med Genet* 2002;**39**:415–421.

De Wert G. Preimplantation genetic testing: normative reflections. In: Harper J (ed), *Preimplantation Genetic Diagnosis*. Cambridge: Cambridge University Press, 2009, 259–273.

Dondorp W, Wert G, Cornel M. The quality of genetic screening: an integral approach. In: Kristoffersson U, Schmidtke J, Cassiman JJ (eds). *Quality Issues in Clinical Genetic Services*. Netherlands: Springer, 2010, 165–172.

Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010;**327**:78–81.

Firth HV, Wright CF. The Deciphering Developmental Disorders (DDD) study. *Dev Med Child Neurol* 2011;**53**:702–703.

Harper JC, Sengupta SB. Preimplantation genetic diagnosis: state of the art 2011. *Hum Genet* 2012;**131**:175–186.

Hens K, Dondorp W, de Wert G. Embryos without secrets: an expert panel study on comprehensive embryo testing and the responsibility of the clinician. *Eur J Med Genet* 2012;**56**:67–71.

Hens K, Dondorp W, Handyside AH, Harper J, Newson AJ, Pennings G, Rehmann-Sutter C, de Wert G. Dynamics and ethics of comprehensive preimplantation genetic testing: a review of the challenges. *Hum Reprod Update* 2013;**19**:366–375.

INSERM. Orphanet: an online database of rare diseases and orphan drugs. 1997. http://www.orpha.net (24 January 2013, date last accessed)

Ito S. Novel complex GUCY2D mutation in Japanese family with cone-rod dystrophy. *Invest Ophthalmol Vis Sci* 2004;**45**:1480–1485.

Kalff-Suske M, Wild A, Topp J, Wessling M, Jacobsen EM, Bornholdt D, Engel H, Heuer H, Aalfs CM, Ausems MG *et al.* Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. *Hum Mol Genet* 1999;**8**:1769–1777.

Koenekoop RK, Fishman GA, Iannaccone A, Ezzeldin H, Ciccarelli ML, Baldi A, Sunness JS, Lotery AJ, Jablonski MM, Pittler SJ *et al.* Electroretinographic abnormalities in parents of patients with Leber congenital amaurosis who have heterozygous GUCY2D mutations. *Arch Ophthalmol* 2002;**120**:1325–1330.

Krauss S, So J, Hambrock M, Köhler A, Kunath M, Scharff C, Wessling M, Grzeschik K-H, Schneider R, Schweiger S. Point mutations in GLI3 lead to misregulation of its subcellular localization. *PLoS One* 2009;**4**:e7471.

MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. *Hum Mol Genet* 2010;**19**:R125–R130.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;**335**:823–828.

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. Online Mendelian Inheritance in Man, OMIM. http://omim.org (28 October 2012, date last accessed)

Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* 2011;**472**:90–94.

Ng PC. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;**31**:3812–3814.

Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ *et al.* The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* 2009;**19**:1316–1323.

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016

Ramos FJ, Kaplan BS, Bellah RD, Zackai EH, Kaplan P. Further evidence that the Hajdu-Cheney syndrome and the 'serpentine fibula-polycystic kidney syndrome' are a single entity. *Am J Med Genet* 1998;**78**:474–481.

Robinson PN, Mundlos S. The human phenotype ontology. *Clin Genet* 2010; **77**:525–534.

Sanderson S, Zimmern R, Kroese M, Higgins J, Patch C, Emery J. How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. *Genet Med* 2005;**7**:495–500.

Sifrim A, Van Houdt JK, Tranchevent L-C, Nowakowska B, Sakai R, Pavlopoulos GA, Devriendt K, Vermeesch JR, Moreau Y, Aerts J. Annotate-it: a Swiss-knife approach to annotation, analysis and interpretation of single nucleotide variation in human disease. *Genome Med* 2012;**4**:73.

Sifrim A, Popovic D, Tranchevent L-C, Ardeshirdavani A, Sakai R, Konings P, Vermeesch JR, Aerts J, De Moor B, Moreau Y. eXtasy: variant prioritization by genomic data fusion. *Nat Methods* 2013;**10**:1083–1084.

Simpson JL, Rechitsky S, Kuliev A. Next-generation sequencing for preimplantation genetic diagnosis. *Fertil Steril* 2013;**99**:1203–1204.

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, Abeysinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;**21**:577–581.

Strozzi F, Aerts J. A Ruby API to query the Ensembl database for genomic features. *Bioinformatics* 2011;**27**:1013–1014.

Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin ML, Zamani Esteki M, Van der Aa N, Mateiu L, McBride DJ *et al*. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 2013;**41**:6119–6138.

von Kanel T, Stanke F, Weber M, Schaller A, Racine J, Kraemer R, Chanson M, Tummler B, Gallati S. Clinical and molecular characterization of the potential CF disease modifier syntaxin 1A. *Eur J Hum Genet* 2013; **21**:1462–1466.

Wild A, Kalff-Suske M, Vortkamp A, Bornholdt D, König R, Grzeschik KH. Point mutations in human GLI3 cause Greig syndrome. *Hum Mol Genet* 1997;**6**:1979–1984.

Wilson JM, Jungner YG. Principles and practice of screening for disease. *WHO Chronicle: Public Health Papers* 1968;**22**:473.

Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, Li F, Tsang S, Wu K, Wu H *et al*. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 2012;**148**:886–895.

Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, Phillips AD, Shaw K, Stenson PD, Cooper DN *et al*. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 2012;**91**:1022–1032.

Zernant J, Kulm M, Dharmaraj S, den Hollander AI, Perrault I, Preising MN, Lorenz B, Kaplan J, Cremers FP, Maumenee I *et al*. Genotyping microarray (disease chip) for Leber congenital amaurosis: detection of modifier alleles. *Invest Ophthalmol Vis Sci* 2005;**46**:3052–3059.

Downloaded from http://humrep.oxfordjournals.org/ at Washinton State University Libraries on April 14, 2016