Available online at www.sciencedirect.com
SCIENCE DIRECT
Pharmacology, Biochemistry and Behavior 78 (2004) 269–274
PHARMACOLOGY BIOCHEMISTRY AND BEHAVIOR
www.elsevier.com/locate/pharmbiochembeh

Automated tests for measuring the effects of antidepressants in mice

James J. Crowley[a], Michelle D. Jones[b], Olivia F. O'Leary[a,c], Irwin Lucki[a,b,*]

[a]Department of Pharmacology, University of Pennsylvania, Philadelphia, PA 19104, USA
[b]Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA
[c]Department of Pharmacology, National University of Ireland, Galway, Ireland

Received 30 December 2003; received in revised form 18 March 2004; accepted 29 March 2004
Available online 10 May 2004

Abstract

The forced swim test (FST) and the tail suspension test (TST) are used widely for measuring the pharmacological effects of antidepressant drugs or changes in stress-evoked behavior in mice. However, inconsistent scoring techniques and poor reproducibility may result from their reliance on subjective ratings by observers to score behavioral changes. In this paper, automated versions of the mouse FST and TST were characterized and validated against observer ratings. For the FST, a commercially available video tracking system (SMART II; San Diego Instruments) measured the duration that mice swam in water-filled cylinders at a set velocity. For the TST, a commercially available automated device (Med Associates, St. Albans, VT) measured input from a strain gauge to detect movements of mice suspended from an elevated bar. Dose-dependent effects of the antidepressant desipramine on FST and TST immobility were measured in CD-1 mice using both automated devices and manual scoring from videotapes. Similar dose–response curves were obtained using both methods. However, a wide range of correlations for raters in the FST indicated that scoring criteria varied for individual raters despite similar instructions. Automated versions of the mouse FST and TST are now available and provide several advantages, including an opportunity to standardize methods across laboratories.

Keywords: Forced swim test; Tail suspension test; Automated; Antidepressant; Mouse

1. Introduction

The forced swim test (FST) and the tail suspension test (TST) are used widely for measuring the pharmacological effects of antidepressant drugs or changes in stress-evoked behavior in mice (Cryan et al., 2002). In the mouse FST (Porsolt et al., 1977a,b), a mouse is placed in a cylinder of water from which there is no escape. Although procedures can vary, behavior is most commonly measured in a single session lasting for 6 min. Mice usually display escape-oriented behaviors when immediately placed in the cylinder, consisting primarily of swimming across the water surface, although climbing behavior could be directed at the cylinder walls. After several minutes, the behavior of mice consist predominantly of bouts of immobility and passive floating, and pretreatment with antidepressant drugs reduce the amount of time spent in the latter two behaviors. The FST is sensitive to all major classes of antidepressant drugs (Borsini and Meli, 1988), including tricyclics, selective norepinephrine and serotonin reuptake inhibitors, monoamine oxidase inhibitors, and atypical antidepressants (Koe et al., 1983; Kulkarni and Mehta, 1985; Cesana et al., 1993; Nixon et al., 1994; Bourin et al., 1996; Redrobe et al., 1996; Da-Rocha et al., 1997; Sanchez and Meier, 1997). The FST is also performed with rats (Porsolt et al., 1977b), usually using different procedures, and modifications of it allow one to distinguish serotonergic and noradrenergic antidepressants (Detke et al., 1995; Lucki, 1997).

In the TST (Steru et al., 1985), mice are suspended by the tail from an elevated bar for several minutes. Typically, they immediately engage in several escape-oriented behaviors, such as leg kicks and body jerks, followed temporally by increasing bouts of immobility. The frequency of immobility is reduced by antidepressant treatments. The TST has

* Corresponding author. Department of Psychiatry, University of Pennsylvania, 538 Clinical Research Building, 415 Curie Boulevard, Philadelphia, PA 19104, USA. Tel.: +1-215-573-3305; fax: +1-215-573-2149.
E-mail address: lucki@pharm.med.upenn.edu (I. Lucki).

0091-3057/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.pbb.2004.03.014

been shown to be sensitive to an array of antidepressants, including tricyclics, selective serotonin reuptake inhibitors, monoamine oxidase inhibitors, atypical antidepressants, and electroconvulsive therapy (Steru et al., 1985, 1987; Perrault et al., 1992; Teste et al., 1990, 1993). Tests of locomotor activity are often used to distinguish the effects of antidepressants from general psychomotor stimulants, because most antidepressants do not increase activity at doses that reduce FST and TST immobility.

Although used most often in drug discovery research, the mouse FST and TST are used by divergent laboratories in the field of psychopharmacology and genetics research (Cryan et al., 2002). Significant procedural drawbacks to the use of these models include their lack of consistency in reference experimental conditions and reliance on subjective ratings by observers to score behavioral changes. Individual raters and different laboratories apparently use distinct criteria for judging immobility, based on the range of immobility scores reported in publications. For example, baseline immobility values for DBA/2 mice on the FST range from 99 (Lucki et al., 2001) to 234 (David et al., 2003) s out of a possible 240 s (final 4 min of a 6 min test). Although some laboratories use videotaped sessions to archive reference standards for training, subjective rating methods can hamper attempts at replicating findings across laboratories and the maintenance of standardized scoring methods over time in the same laboratory. Manual scoring is also tedious and time consuming, especially when a large number of animals are being tested. Extra personnel may also be required so that raters are unaware of the experimental conditions for individual animals. Automated versions of tests would be extremely desirable for several reasons: (1) assurance of consistent immobility ratings, (2) an opportunity to standardize methods across different laboratories, and (3) greatly increased throughput. However, any proposed automated system needs to be validated against manual scoring techniques and for the ability to detect antidepressant drug effects. In this article, automated versions of the mouse FST and TST are described and validated.

2. Materials and methods

2.1. Animals

Male CD-1 mice 10 to 14 weeks old were purchased from Charles River (Wilmington, MA). Subjects were housed in groups of four per cage (cage size: 28.5 × 17.5 × 13.0 cm) in a temperature-controlled environment (22 ± 1 °C) under a 12-h light–dark cycle, with lights turned on at 0700 h, for at least 1 week prior to testing. Food and water were freely available. All behavioral testing was performed between 1000 and 1800 h. All procedures are in compliance with the National Institutes of Health Guide for Care and Use of Laboratory Animals.

2.2. Drug

Desipramine hydrochloride (Sigma, St. Louis, MO) was prepared fresh daily by solution in deionized water. Drug was administered by injection in a volume of 10 ml/kg ip. Drug doses were calculated as mg/kg base. Control animals received injections of 0.9% saline in a volume of 10 ml/kg.

2.3. FST

Swim sessions were conducted by placing mice in individual polycarbonate cylinders (25.3 cm tall × 22.2 cm in diameter; Fisher Scientific, Pittsburgh, PA) filled with water (23–25 °C) to a depth of 15 cm for 6 min. The water was changed between each animal. A camcorder (Sony Handycam) positioned directly above the cylinders (using a tripod with horizontal extension) recorded the swim sessions on videotape. To allow comparison with automated scores, all swim sessions were videotaped and scored for immobility by three highly experienced raters who were blind to treatment groups. Immobility was defined as the absence of movement, except that necessary to keep afloat, and included passive floating, where animals drifted on the water surface across the cylinder without any initiation of swimming. Digital video output was analyzed by an IBM-compatible computer running SMART II Video Tracking System software (San Diego Instruments). Indirect lighting was used and black poster board was placed beneath the cylinders, allowing SMART to more easily track the albino CD-1 mice. SMART recorded the horizontal velocity (in cm/s) of mice in 100-ms intervals. The software then calculated the total amount of immobility in the 6-min test by measuring the duration of time (in seconds) that the mouse traveled below the specified threshold velocity of 2.0 cm/s. This threshold velocity was chosen because it produced immobility scores similar in magnitude to those determined from manually scored videotapes in preliminary studies. Mice were injected with desipramine (5, 10, or 20 mg/kg) or saline 30 min prior to the FST. The number of animals in each group ranged from 11–20. Mice were run individually in this study.

2.4. TST

An automated TST device (Med Associates) was used to measure the duration of behavioral immobility. In addition, to allow comparison with automated scores, all TSTs were videotaped and scored for immobility by three highly experienced raters who were blind to treatment groups. Immobility was defined as the absence of initiated movements and included passive swaying. The automated device consists of a box-like enclosure (box size: 32 × 33 × 33 cm) that was open on the front side, allowing videotaping. A vertical aluminum bar (bar size: 11.5 × 2.2 × 0.15 cm), suspended from the top, was connected to a strain gauge that detected any movements by the mouse. Mice were

suspended by the tail with tape for 6 min and were positioned such that the base of their tail was aligned with the bottom of the bar. The total duration of immobility was calculated as the time the force of the mouse's movements was below a preset threshold. An optimum threshold was determined by comparing scores rated manually from videotapes with scores from the automated device in preliminary studies. The following settings were used in all experiments: Threshold 1 = 10, gain = 16, time constant = 0.25, resolution = 200 ms. Mice were injected with desipramine (0.5, 1.0, or 5.0 mg/kg) or saline 30 min prior to the TST. The number of animals in each group ranged from 11–15. Two experimental boxes were used simultaneously in this study.

2.5. Statistics

Immobility scores in the FST and TST were analyzed by two-way analysis of variance (ANOVA), consisting of Raters (four, including the automatic device) and Dose of desipramine as main effects. Within individual raters, Fishers LSD test was used to determine doses that produced values that differed significantly than saline ($P < .05$). Interrater correlations (Pearson Product–Moment Correlation Coefficients, r values) were calculated by linear regression and tested for significance using the critical ratio z test. Statview software (SAS Institute, Cary, NC) was used for all analyses.

3. Results

Fig. 1 shows the effect of desipramine on FST immobility values determined by the automated FST device and three experienced raters (all authors; identified by numbers). Two-way ANOVA indicated that both the dose of desipramine [$F(3,268) = 32.6$, $P < .0001$] and the identity of the rater [$F(3,268) = 15.5$, $P < .0001$] had a significant effect on immobility values, but no significant interaction between dose of desipramine and rater identity was detected [$F(9,268) = 0.68$, $P = .72$]. Follow-up tests indicated that desipramine significantly decreased immobility ($P < .01$) at the 10- and 20-mg/kg doses for all but Rater 3, for whom only the 20-mg/kg dose was significant. Interrater correlations of immobility values scored by the three experienced raters and the automated FST for 71 individual mice are shown in Table 1. Values of immobility differed among the raters, with one of the raters showing consistently lower scores. Although all of the correlations between the human raters and the automated device were significant ($P < .01$), the individual correlations ranged from .81–.93 ($P < .01$). This range was similar to the range of correlations among the human raters (.83–.91; $P < .01$).

Table 1
Interrater correlations (r values derived from linear regression) for FST immobility values determined by the automated FST method and three trained raters for 71 individual mice treated with different doses of desipramine or saline

Rater	1	2	3	Auto
1	–	–	–	–
2	.91	–	–	–
3	.88	.83	–	–
Auto	.93	.85	.81	–

Fig. 2 shows the effect of desipramine on TST immobility values recorded by the automated TST device and the same three human raters. Two-way ANOVA indicated that the dose of desipramine had a significant effect on immobility values [$F(3,184) = 28.2$, $P < .0001$], but there was no significant effect of the rater [$F(3,184) = 0.54$, $P = .81$] or interaction between dose of desipramine and rater [$F(9,184) = 0.41$, $P = .87$]. Follow-up tests indicated that desipramine significantly decreased immobility at the 1- and 5-mg/kg doses for all raters ($P < .01$). Desipramine was more potent in the TST than FST, as evidenced by the 5-mg/kg dose which was not active in the FST but produced

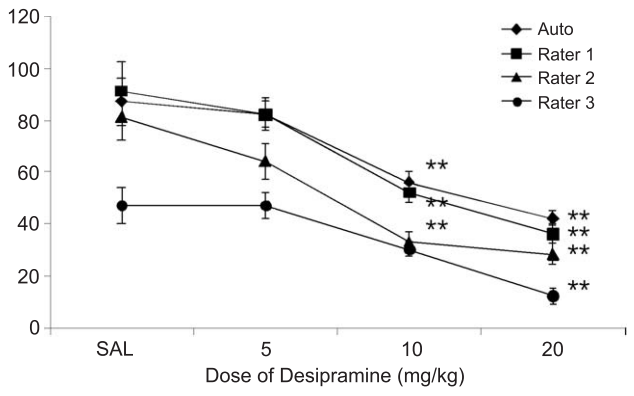

Fig. 1. Effect of desipramine on FST immobility values (mean ± S.E.M.) recorded by an automated FST method and three experienced raters (all authors; identified by numbers). The number of animals in each group was the following: saline (18); 5 mg/kg (15); 10 mg/kg (14); 20 mg/kg (14). ** $P < .01$ from corresponding saline-treated values.

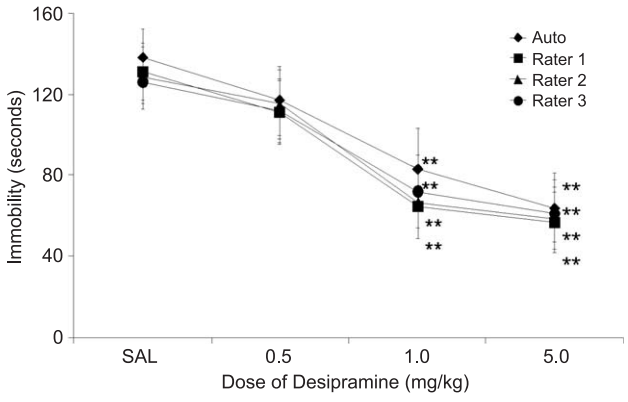

Fig. 2. Effect of desipramine on TST immobility values (mean ± S.E.M.) recorded by an automated TST device and three experienced raters (all authors; identified by numbers). The number of animals in each group was the following: saline (15); 0.5 mg/kg (12); 1 mg/kg (12); 5 mg/kg (11). ** $P < .01$ from corresponding saline-treated values.

Table 2
Interrater correlations (r values derived from linear regression) for TST immobility values determined by the automated TST device and three trained raters for 50 individual mice treated with different doses of desipramine or saline

Rater	1	2	3	Auto
1	–	–	–	–
2	.96	–	–	–
3	.95	.95	–	–
Auto	.97	.97	.96	–

significant effects in the TST. Table 2 shows the interrater correlations of immobility values for 50 individual mice scored by three trained raters and the automated TST. All of the correlations between the human raters and the automated device were significant ($P < .001$), with the individual correlations ranging from .96–.97. This range was similar to the range of correlations among the human raters (.95–.96; $P < .001$).

4. Discussion

The present study validated automated versions for measuring the FST and TST in the mouse by examining dose–response curves for the antidepressant drug desipramine and by comparing the similarity of scores from experienced human raters with the automated devices. Although this article is primarily methodological, the findings have practical utility for labs that use the mouse FST or TST. The SMART II system from San Diego Instruments is utilized as a behavioral tracking device for locomotor activity or swimming but has not been used previously as an automated mouse FST device as described in this manuscript. The key feature of this software that enables use in the FST is the ability to calculate the duration of time when targeted movement is less than a specified speed from repeated samplings of location. Thus, a criterion for immobility can be objectively specified according to distance and time using the device. The software is capable of tracking up to 16 animals at one time, although we have successfully conducted only four swim tests simultaneously (data not shown). The system is also capable of tracking mice of any coat color: we have easily tracked albino (CD-1), brown (DBA/2), and black (C57BL/6) mouse strains.

Two other automated mouse FST devices have been reported in the literature. The first used a frame grabber and image analysis software to determine the change in the area of pixels occupied by a mouse's image at 240-ms intervals (Sanchez and Meier, 1997). The mouse was considered immobile if the change in movement during the sampling interval was less than 700 pixels. The authors stated that this criterion was selected from preliminary comparisons between manually and automated assessments of dose–response studies for imipramine. This method is similar in principle to the one described in this manuscript because a criterion for immobility was objectively defined in distance and time and scored automatically. The other method used an activity-monitoring system from Japan, termed Supermex, which monitors radiated body heat of mice over time across multiple zones of the water surface (Masuo et al., 1997) and was used in a recent genetic study (Yoshikawa et al., 2002). It was unclear from the publication, however, how the sensors for this device determined scores for immobility.

Several automated commercial versions of the FST for use with rats have also been reported (Shimazoe et al., 1987; De Pablo et al., 1989; Hedou et al., 2001) using different types of sensors to monitor movement or locomotion. However, none of these devices allow criteria for determining immobility to be specified directly and objectively. In addition, there are mechanistic advantages to the direct observations from videotape that determine the form of active behaviors (e.g., swimming versus climbing behavior) that displace behavioral immobility in the rat FST when antidepressants are active (Lucki, 1997).

One of the early advantages of the development of the TST was that a custom automated system which used a strain gauge to measure the mouse's movements (Itematic-TST from Item-Labo; Le Kremlin-Bicetre, France), was validated by Steru et al. (1987). By comparison with the FST, the present validation indicated that measures of immobility tended to be more similar among raters, and between raters and the automated device in the TST. A number of companies now offer similar automated tail suspension systems, all of which measure continuous digital output converted from a strain gauge device (Med Associates; Hamilton-Kinder, Poway, CA; Neuroscience, Tokyo, Japan). In this manuscript, we successfully validated the Med Associates automated TST device against manual scoring. A complete system is available where a total of eight mice could be run simultaneously, although fewer boxes could be sufficient throughput for most academic laboratories.

Although automated systems typically involve a large initial cost, a cost–benefit analysis reveals that the overall advantage of automated systems can be substantial. This is especially true for laboratories screening a large number of animals. Rating immobility for the 71 mice in the FST portion of this study required approximately 12 h of scoring videotapes per person to rate the entire data set. Likewise, to rate immobility for the 50 mice in the TST portion of this study required approximately 8 h to score each data set. In general, using the automated versions of these tests more than doubled the speed of data scoring compared to manual ratings from videotapes, allowing for necessary breaks to prevent fatigue. Some groups might reduce the time of manual scoring by rating the experiments live but sacrifice the reliability provided by a video record. The labor-saving value of the automated system for any group depends on their work load. For industry laboratories that might screen 4000 animals per year in a drug discovery program, we estimated

that at least 600 person-hours (or 4 months of full-time work) could be saved from automated scoring. Greater savings might accrue to laboratories with high-throughput programs requiring greater screening demands. These savings would certainly offset the cost of the automated device in less than 1 year of use. For smaller academic laboratories conducting occasional pharmacological screening studies or genetic studies, the device would still accrue savings from labor, although the savings would accrue from more long-term use.

Other substantial scientific benefits can also be attributed to automated scoring systems whose value may be more difficult to assess than overall cost savings. The systems are innately blind to the identity of the treatment group, although animal handlers could still influence experimental outcome if they know the identity of the treatment groups. In addition, the stability of scoring criteria would remain constant over time. This could be important to laboratories where different investigators run the same tests or after long-term use by different personnel that have worked in the laboratory. These characteristics would be difficult to achieve with human raters.

The development of automated versions of the FST and TST for the mouse does not diminish other sources of laboratory variations with these tests. For example, cylinder size, water depth, and temperature remain important methodological variables that alter behavioral immobility and vary between laboratories (Sunal et al., 1994). Although Porsolt originally employed small cylinders or beakers for tests with mice (Porsolt et al., 1977a,b), larger cylinders can avoid the detection of false positives by nonantidepressant drugs (Sunal et al., 1994). Although most laboratories use the testing procedure originally suggested by Porsolt (6-min test with only 2–6 min scored), other variations involving multiple sessions have also been used. Finally, automated versions of these tests do not eliminate the need to evaluate changes in locomotor activity as a contributing factor to apparent antidepressant-like effects from novel drugs.

In conclusion, we have used a novel automated method for measuring the mouse FST using a commercially available video tracking device and validated scores for the effects of desipramine against those from three experienced human raters. In addition, an existing automated TST device was similarly validated for the effects of desipramine and compared with the FST. Applications for reliable, high-throughput versions of these systems could include screening chemical libraries for compounds with antidepressant activity and also screening large groups of behavioral mutants for genetic contributions to stress-mediated behaviors.

Acknowledgements

This research was supported by USPHS grants MH 01465 and MH 48152. The authors are grateful to Dr. Richard Butcher and to San Diego Instruments for providing a prototype SMART II video tracking system for supporting this research.

References

Borsini F, Meli A. Is the forced swimming test a suitable model for revealing antidepressant activity. Psychopharmacology 1988;94:147–60.

Bourin M, Hascoet M, Colombel M-C, Redrobe JP, Baker GB. Differential effects of clonidine, lithium, and quinine in the forced swimming test in mice for antidepressants: possible roles of serotonergic systems. Eur Neuropsychopharmacol 1996;6:231–6.

Cesana R, Ceci A, Ciprandi C, Borsini F. Mesulergine antagonism towards the fluoxetine anti-immobility effects in the forced swimming test in mice. J Pharm Pharmacol 1993;45:473–5.

Cryan JF, Markou A, Lucki I. Assessing antidepressant activity in rodents: recent developments and future needs. Trends Pharmacol Sci 2002; 23:238–45.

Da-Rocha MA, Puech AJ, Thiebot MH. Influence of anxiolytic drugs on the effects of specific serotonin uptake. J Psychopharmacol 1997; 11:211–8.

David DJP, Renard CE, Jolliet P, Hascoet M, Bourin M. Antidepressant-like effects in various mice strains in the forced swimming test. Psychopharmacology 2003;166:373–82.

De Pablo JM, Parra A, Segoia S, Guillamon A. Learned immobility explains the behavior of rats in the forced swimming test. Physiol Behav 1989;46:229–37.

Detke MJ, Rickels M, Lucki I. Active behaviors in the rat forced swimming test differentially produced by serotonergic and noradrenergic antidepressants. Psychopharmacology 1995;121:66–72.

Hedou G, Pryce C, Di Iorio L, Heidbreder CA, Feldon J. An automated analysis of rat behavior in the forced swim test. Pharmacol Biochem Behav 2001;70:65–76.

Koe BK, Wiessman A, Welch WM, Browne RG. Sertraline, 1S, 4S-*n*-methyl-4-(3,4-dichlorophenyl)-1,2,3,4-tetrahydro-1-naphthylamine, a new uptake inhibitor with selectivity for serotonin. J Pharmacol Exp Ther 1983;226:686–700.

Kulkarni SK, Mehta AK. Purine nucleoside-mediated immobility in mice-reversal by antidepressants. Psychopharmacology 1985;85:460–3.

Lucki I. The forced swimming test as a model for core and component behavioral effects of antidepressant drugs. Behav Pharmacol 1997; 8:523–32.

Lucki I, Dalvi A, Mayorga AJ. Sensitivity to the effects of pharmacologically selective antidepressants in different strains of mice. Psychopharmacology 2001;155:315–22.

Masuo Y, Matsumoto Y, Morita S, Noguchi J. A novel method for counting spontaneous motor activity in the rat. Brain Res Protoc 1997;1:321–6.

Nixon MK, Hascoet M, Bourin M, Colombel MC. Additive effects of lithium and antidepressants in the forced swimming test: further evidence for involvement of the serotonergic system. Psychopharmacology 1994;115:59–64.

Perrault GH, Morel E, Zivkovic B, Sanger DJ. Activity of litoxetine and other serotonin uptake inhibitors in the tail suspension test in mice. Pharmacol Biochem Behav 1992;42:45–7.

Porsolt RD, Bertin A, Jalfre M. Behavioral despair in mice: a primary screening test for antidepressants. Arch Int Pharmacodyn Ther 1977a; 229:327–36.

Porsolt RD, Le Pichon M, Jalfre M. Depression: a new animal model sensitive to antidepressant treatments. Nature 1977b;266:730–2.

Redrobe JP, MacSweeney CP, Bourin M. The role of 5-HT1A and 5-HT1B receptors in antidepressant drug actions in the mouse forced swimming test. Eur J Pharmacol 1996;318:213–20.

Sanchez C, Meier E. Behavioral profiles of SSRIs in animal models of depression, anxiety and aggression—are they all alike? Psychopharmacology 1997;129:197–205.

Shimazoe T, Shibata S, Ueki S. A new forced swimming test for the

evaluation of antidepressants in rats by recording vibration of a water tank. J Pharmacobio-Dyn 1987;10:639–43.

Steru L, Chermat R, Thierry B, Simon P. The tail suspension test: a new method for screening antidepressants in mice. Psychopharmacology 1985;85:367–70.

Steru L, Chermat R, Thierry B, Mico JA, Lenegre A, Steru M, et al. The automated tail suspension test: a computerized device which differentiates psychotropic drugs. Prog Neuropsychopharmacol Biol Psychiatry 1987;11:659–71.

Sunal R, Gumusel B, Kayaalp SO. Effect of changes in swimming area on results of ''behavioral despair test''. Pharmacol Biochem Behav 1994;49:891–6.

Teste JF, Martin I, Rinjard P. Electrotherapy in mice: dopaminergic and noradrenergic effects in the tail suspension test. Fundam Clin Pharmacol 1990;4:39–47.

Teste JF, Pelsey-Johann I, Decelle T, Boulu RG. Anti-immobility activity of different antidepressant drugs using the tail suspension test in normal or reserpinized mice. Fundam Clin Pharmacol 1993;7:219–26.

Yoshikawa T, Watanabe A, Ishitsuka Y, Nakaya A, Nakatani N. Identification of multiple genetic loci linked to the propensity for "behavioral despair" in mice. Genome Res 2002;12:357–66.