

Frage- und Antwort-Dokument des
Umweltbundesamtes zur neuen Richtlinie
OECD 209 „Activated Sludge, Respiration
Inhibition Test (Carbon and Ammonium
Oxidation) (22.07.2010)

Seit Inkrafttreten der neuen Version der OECD 209 gab es verschiedene Anfragen von Antragstellern und Prüfinstituten an das Umweltbundesamt bezüglich der Interpretation der Richtlinie und der daraus resultierenden Durchführung der Studien. Einige Passagen in der neuen Richtlinie bedürfen einer Überarbeitung, da sie überflüssig, wissenschaftlich nicht sinnvoll oder auf Dauer ungenügend erklärt sind. Das Umweltbundesamt hatte während der Kommentierungsphase des OECD Prüfrichtlinienentwurfs viele dieser Punkte kritisch kommentiert, aber die deutschen Kommentare wurden z.T. nicht akzeptiert. Eine evtl. erneute Revision der OECD 209 ist z.Z. nicht absehbar.

Um Antragstellern und Prüfinstituten die Sicherheit zu geben, dass die von ihnen durchgeführten Studien vom Umweltbundesamt für die Umweltrisikobewertung in allen Vollzügen (REACH, Pflanzenschutz, Biozide, Arzneimittel) anerkannt werden, werden in diesem Dokument Antworten auf die häufigsten Fragen gegeben.

Es wird dargestellt, wie das Umweltbundesamt die neue Richtlinie interpretiert, wie Studien durchgeführt werden müssen und wann von einigen Vorgaben der Richtlinie abgewichen werden kann.

Ob diese Studien dann auch von den Bewertungsbehörden anderer Länder akzeptiert werden (insbesondere bei EU-Verfahren), sollten Antragsteller zuvor in Erfahrung bringen.

Das Dokument wird bei Bedarf aktualisiert, um weitere häufige Fragen zur Prüfrichtlinie oder neue Erkenntnisse zu integrieren.

Eine aktuelle Version des Dokuments ist unter <http://reach-info.de/pruefrichtlinien.htm> abrufbar.

Anfragen zur Prüfrichtlinie und Anmerkungen zum vorliegenden Dokument können per E-Mail an Lars.Hohndorf@uba.de erfolgen.

Ist die Bestimmung der Nitrifikation obligatorisch?

Die Untersuchung der Hemmung der Nitrifikation wurde aus der DIN EN ISO 8192:2007 übernommen. Hintergrund ist, dass bei Messung des Gesamtsauerstoffverbrauchs in einzelnen Fällen Dosis-Wirkungs-Kurven einen biphasischen bzw. gestörten Verlauf zeigten oder EC50-Werte zu niedrig waren. In diesen Fällen nitrifizierte der verwendete Belebtschlamm deutlich und die Testsubstanz hatte auf die Ammoniumoxidation einen größeren Effekt als auf die heterotrophe Kohlenstoffoxidation.

Paragraph 2 der Richtlinie enthält Informationen zum Hintergrund, warum und wann der Einfluss auf die Nitrifikation gemessen werden sollte.

Die Nitrifikation muss nicht obligatorisch gemessen werden, sondern nur dann, wenn erwartet wird, dass der verwendete Belebtschlamm deutlich nitrifiziert. Das kann in einem Vortest mit einem Nitrifikationsinhibitor untersucht werden (siehe auch Paragraph 2, 4, 6, 30, 38 der Richtlinie).

Die Nitrifikation muss auch dann nicht gemessen werden, wenn z.B. in einem Vortest mit der Testsubstanz keine Hemmung des Gesamtsauerstoffverbrauchs gemessen wird.

Paragraph 4 ‚For most purposes, the method to assess the effect on organic carbon oxidation processes alone is adequate. (...)‘

Paragraph 30 ‚... Normally, the measurement of total oxygen uptake inhibition should be adequate. ...‘

Welche Anzahl an Kontrollen, Konzentrationen und Replikaten muss getestet werden und wie ergab sich diese Forderung in der Richtlinie?

Die Forderung und Diskussion einer höheren Anzahl an Konzentrationen und Replikaten kam erst während der Kommentierungsphase des Prüfrichtlinienentwurfs auf. Die neue Richtlinie OECD 209 sollte statistisch an die vorhandenen Ökotoxizitätstests angepasst werden, um die ermittelten NOEC- und EC_x-Werte abzusichern.

In Paragraph 5 ‚Principle of the Test‘ im letzten Satz steht, dass der Test normalerweise genutzt wird, um eine EC_x (z.B. EC₅₀) der Testsubstanz **und/ oder** die ‚no-observed effect concentration‘ (NOEC) zu bestimmen. Im Paragraphen 41 der Richtlinie wird dann aber nur ein Ansatz für die kombinierte Bestimmung einer NOEC **und** EC_x vorgeschlagen. Dieser Vorschlag beinhaltet 6 Kontrollen und 5 Konzentrationen mit je 5 Replikaten.

Es ist bekannt, dass für die statistische Ermittlung von NOEC und EC_x unterschiedliche Ansätze erforderlich sind. Beide Ansätze waren während der Diskussion des Entwurfs enthalten, wodurch die statistische Anleitung kompliziert und schwer verständlich wurde. Um zu vermeiden, dass beide Ansätze parallel getestet werden, wurde mit Hilfe eines Statistikers der Kompromiss gefunden, mit nur einem Ansatz NOEC und EC_x statistisch ausreichend sicher zu ermitteln.

Für den kombinierten Ansatz ist die Anzahl an Kontrollen, Konzentrationen und Replikaten entsprechend den Vorgaben der Richtlinie zu wählen. Der Ansatz kann aber genau so verwendet werden, wenn nur NOEC oder EC₅₀ einzeln bestimmt werden. D.h., das Testdesign muss nicht von vorn herein auf einen bestimmten Ansatz festgelegt werden.

Andererseits kann aus dem Paragraphen 41 geschlussfolgert werden, dass wenn nur eine NOEC **oder** eine EC₅₀ bestimmt werden soll, auch von diesem Ansatz abgewichen werden kann, solange eine ausreichende statistische Sicherheit durch die notwendige Anzahl an Kontrollen, Testkonzentrationen und Replikaten gewährleistet ist. D.h. für die Bestimmung der NOEC i.d.R. weniger Konzentrationen aber mehr Replikate, für die Bestimmung der EC₅₀ eher mehr Konzentrationen und weniger Replikate.

Entsprechende Studien werden dann vom Umweltbundesamt anerkannt.

Zum Beispiel stellt ein EC_x-Ansatz mit 5 Konzentrationen (Faktor $\leq 3,2$), jeweils 3 Replikaten und möglichst 6 Kontrollen einen geeigneten Minimalansatz dar, muss aber gewährleisten, dass der gewünschte Wirkungsbereich über die Konzentrationen abgedeckt ist.

Muss in jedem Fall eine NOEC bestimmt werden und sind auch Alternativen wie EC₁₀ verwendbar?

Für die EU gilt, dass grundsätzlich immer der EC_x-Ansatz zu verwenden ist. Auch NOEC-Alternativen wie EC₁₀ sind akzeptabel. Ob in einem außereuropäischen Land eine NOEC notwendig ist, sollte der Antragsteller vor Durchführung der Studie bei der dort zuständigen Behörde in Erfahrung bringen, um das richtige Testdesign zu wählen.

Erfolgt keine signifikante Hemmung bis zur höchsten getesteten Konzentration, wird diese Konzentration als NOEC betrachtet (NOEC \geq höchste Testkonzentration).

Ist es möglich, einen Limit-Test durchzuführen und wenn ja mit welcher Konzentration?

Die Richtlinie ermöglicht, einen Limit-Test mit 6 Kontrollen und einer Testkonzentration mit 3 Replikaten durchzuführen. Dieser kann zum Beispiel bei 1000 mg/L Testkonzentration durchgeführt werden, allerdings wäre bei im Test gemessenen Effekten ein erneuter vollständiger Test mit einer Konzentrationsreihe nötig. Die Richtlinie lässt also einen Spielraum für die Wahl der Testkonzentration im Limit-Test, sie sollte aber möglichst hoch gewählt werden (PEC/PNEC Kläranlage). Nach bisheriger Bewertungspraxis sind Limit-Tests mit 100 mg/L Testkonzentration zu empfehlen.

Gelten für alle Vollzüge die gleichen Anforderungen und Kriterien?

Für alle Vollzüge (REACH, Pflanzenschutz, Biozide, Arzneimittel) gelten die gleichen Anforderungen und Kriterien.

Können Ergebnisse zur Nitrifikationshemmung einer Testsubstanz im Boden auf die Kläranlage übertragen werden?

Testergebnisse aus den Soil-Microflora-Tests zur Hemmung der Nitrifikation durch die entsprechende Testsubstanz sind nicht aussagekräftig für die Vorhersage einer Wirkung auf die Nitrifikation in der Kläranlage.

Wann ist eine Analytik erforderlich?

In der Richtlinie finden sich deutliche Informationen dazu, wann und wie eine Analytik erfolgen soll.

Eine Analytik ist nur dann erforderlich, wenn die nominale Testkonzentration in den Testgefäßen nicht bekannt ist. Das ist z.B. der Fall, wenn Stammlösungen von Testsubstanzen mit Konzentrationen oberhalb der Wasserlöslichkeit angesetzt werden und ein Teil der eingewogenen Testsubstanz ungelöst verbleibt. In diesem Fall ist nicht bekannt, wie viel der Testsubstanz tatsächlich in das Testgefäß übertragen wurde und eine Konzentrationsbestimmung ist erforderlich. Eine Analytik sollte dann der Einfachheit halber immer vor Zugabe des Inokulums erfolgen.

Grundsätzlich wird für alle Testsubstanzen insbesondere für problematische Stoffe (schlecht wasserlöslich, adsorbierend, hydrolysierend, flüchtig) die Direkteinwaage in das Testgefäß empfohlen. Dann ist ein Bezug auf die eingewogene nominale Konzentration möglich, unabhängig davon, ob die gesamte Testsubstanz gelöst vorliegt, da dieser Ansatz am wahrscheinlichsten die Situation in der Kläranlage abbildet. Zudem ist bekannt, dass auch Effekte oberhalb der Wasserlöslichkeit möglich sind.

Wie ist mit flüchtigen Testsubstanzen zu verfahren?

Die in der Richtlinie enthaltenen entsprechenden Formulierungen erfordern in Einzelfällen eine Analytik bei instabilen oder flüchtigen Testsubstanzen. Diese Forderungen erscheinen nicht immer sinnvoll.

Auch bei instabilen und flüchtigen Verbindungen sollten eine Direkteinwaage und ein Bezug auf die Nominalkonzentration erfolgen. Eine aufwendige Analytik ist dann nicht erforderlich. Die Testsubstanzen sollten unter möglichst schonenden Bedingungen als letztes direkt vor Versuchsbeginn in die Testgefäße gegeben werden. Das Testdesign sollte so gestaltet werden, dass die Verflüchtigung möglichst gering ist bei gleichzeitig ausreichender Sauerstoffversorgung. Da auch unter den realen Bedingungen in Abwasserleitungen und Kläranlagen Verflüchtigung erfolgt, sollte das Testdesign möglichst den „worst case“ bezüglich der verfügbaren Konzentration im Testgefäß darstellen, so dass davon ausgegangen werden kann, dass die Verflüchtigung in Abwasserleitungen und in der Kläranlage der im Laborexperiment entspricht oder sogar höher ist.

Eine Testdauer von 3h ist auch bei flüchtigen Substanzen der Mindeststandard und erfasst auch die Effekte in den ersten 30 min.

Darüber hinaus wurde überlegt, Triggerwerte für kritische Henry-Konstanten, bei deren Überschreitung ein modifiziertes Testdesign erforderlich wäre, als Empfehlung in das vorliegende F&A-Dokument zu integrieren. Erste Untersuchungen eines Prüfinstituts dazu zeigen aber, dass es nicht möglich ist, die ermittelten Konstanten repräsentativ auf andere Labore, anderes Testequipment und andere Testsubstanzen zu übertragen.

Das neue Validitätskriterium des Sauerstoffverbrauchs der Kontrollen kann in der Praxis wahrscheinlich nicht mit allen Belebtschlämmen eingehalten werden. Basiert es auf wissenschaftlichen Notwendigkeiten?

Das Validitätskriterium des Sauerstoffverbrauchs der Kontrollen wurde der DIN EN ISO 8192:2007 entnommen und ist seit 1995 Bestandteil dieser Richtlinie. Deutschland hatte hierzu während der Kommentierungsphase des OECD Prüfrichtlinienentwurfs Ende 2009/Anfang 2010 kritisch kommentiert aber die deutschen Kommentare wurden nicht akzeptiert.

Dem UBA vorliegende ältere Prüfberichte zeigen, dass einige der von den Prüfinstituten verwendeten Belebtschlämme das Kriterium leicht erfüllen, während andere es nicht erreichen. Die Erfahrungen von Prüfinstituten zeigen, dass die Respirationsraten der Belebtschlämme verschiedener Kläranlagen deutlich streuen. Darüber hinaus können jahreszeitliche Schwankungen gemessen werden. Ein Einfluss auf die Sensitivität der Belebtschlämme konnte dabei bisher aber nicht beobachtet werden.

Da eine zu geringe Eigenatmung des Belebtschlammes genau so ungünstig ist wie eine zu hohe, erscheint ein Validitätskriterium dahingehend aber wissenschaftlich sinnvoll, um atypische Belebtschlämme gestörter Kläranlagen bzw. besonders belasteter Kläranlagen zu identifizieren.

Aus Sicht des UBA sollte dieses Kriterium an realistische Werte bzw. einen Wertebereich angepasst werden, wenn der Großteil der verwendeten Belebtschlämme zeigt, dass das Validitätskriterium nicht erfüllt wird. Es wird geprüft, in wie weit das Kriterium auch in anderen Ländern und Regionen nicht eingehalten wird, um dann eine entsprechende Änderung der OECD 209 vorzuschlagen.

Bis dahin wird das Umweltbundesamt im Einzelfall prüfen, ob auch solche Studien anerkannt werden können, bei denen das Validitätskriterium unterschritten ist, insofern die Sensitivität des Belebtschlammes gegenüber der Referenzsubstanz im gültigen Bereich liegt und die Studie in allen anderen Punkten valide und plausibel ist. Eine plausible Begründung des Prüfinstituts für die Abweichung vom Validitätskriterium muss im Prüfbericht enthalten sein.

Ob diese Studien dann auch von den Bewertungsbehörden anderer Länder akzeptiert werden (insbesondere bei EU-Verfahren), sollten Antragsteller zuvor in Erfahrung bringen.

In der neuen Richtlinie soll der Variationskoeffizient der Sauerstoffverbrauchsraten in den Kontrollen am Ende des Tests nicht mehr als 30 % betragen. Ist dieser Wert nicht viel zu groß?

In der Tat erscheint dieser Wert sehr groß und Studien mit einem Variationskoeffizienten von 30 % in den Kontrollen könnten schwer auszuwerten sein. Allerdings orientierte man sich wahrscheinlich bei der Entwicklung des Prüfrichtlinienentwurfs an Werten aus den Ökotoxizitätstests (z.B. OECD 201 35 % section-by-section, OECD 211 \leq 25 %).

Es wird empfohlen, bei der Präparation der Testgefäße 3 Kontrollen am Anfang und 3 Kontrollen am Ende der Messreihe anzuordnen und zum Beispiel über einen t-Test mit den logarithmierten Daten (Normalverteilung) zu prüfen, ob sich deren Mittelwerte signifikant

unterscheiden (was dem alten Validitätskriterium gerecht werden würde). Anschließend können alle Kontrollen „gepoolt“ und der Variationskoeffizient berechnet werden.

Das neue Validitätskriterium ist nicht zu vergleichen mit dem bisherigen, bei dem die absolute Abweichung zwischen der Kontrolle am Anfang und der Kontrolle am Ende der Messreihe nicht größer als 15 % sein durfte.

Ist Paragraph 34 der Richtlinie so zu verstehen, dass bei einer parallelen/ simultanen Präparation und Messung der Testgefäße 2 Kontrollen ausreichend sind?

Nein, auch bei einer parallelen/ simultanen Präparation und Messung der Testgefäße sind aus statistischen Gründen insgesamt mindestens 6 Kontrollen erforderlich. Bei einer simultanen Messung der Testgefäße in Gruppen (batch) sind je Gruppe mindestens 2 Kontrollen erforderlich, so dass je nach Anzahl der Gruppen insgesamt mehr als 6 Kontrollen erforderlich sein können. Aufgrund der großen Anzahl an Testgefäßen wird eine serielle Präparation und Messung der Testgefäße eher die Regel sein.

In diesem Zusammenhang ist zu erwähnen, dass der erste Satz des Paragraphen 34 falsch formuliert ist. Es muss heißen: „Blank controls (F_B) have to be prepared at the beginning and end of the **test series** in tests in which the test beakers are set up sequentially at intervals.“

Warum müssen die Ergebnisse über eine so große Anzahl an Kontrollen, Testkonzentrationen und Replikaten statistisch abgesichert werden, wenn die Angabe der Ergebnisse als „order of magnitude“ möglich ist?

Die Angabe der Ergebnisse als „order of magnitude“ war im Prüfrichtlinienentwurf als Option enthalten. Auch die kritische Kommentierung durch das UBA konnte diese Option nicht vollständig entfernen. Da für die Risikobewertung einzelne Werte und nicht Wertebereiche erforderlich sind, ist die Angabe als „order of magnitude“ nur noch zusätzlich möglich. Anderenfalls wäre auch der statistische Aufwand nicht gerechtfertigt.

Muss die Exposition des Belebtschlammes gegenüber der Testsubstanz immer 3 h betragen?

Die Exposition des Belebtschlammes muss mindestens 3 h betragen. Eine Messung nach 30 min kann nur zusätzlich erfolgen. Dem UBA vorliegende Prüfberichte, bei denen nach 30 min und nach 3 h gemessen wurde, zeigen in einzelnen Fällen, dass es deutliche Unterschiede in den ermittelten Effektwerten geben kann. 30 min sind zudem nicht repräsentativ für die Verweilzeiten in den Kläranlagen.

Studien, die ausschließlich mit einer Expositionszeit von < 3 h durchgeführt wurden, werden vom Umweltbundesamt nicht mehr anerkannt.

Je nach erwarteter Wirkung einer Testsubstanz kann es in Einzelfällen auch angebracht sein, längere Expositionszeiten zu wählen, um sich den realen Verweilzeiten in den Kläranlagen anzunähern und evtl. später auftretende Effekte zu erfassen.