Organic Process Research & Development **2008,** *12,* 956–966

Development and Scale-Up of Three Consecutive Continuous Reactions for Production of 6-Hydroxybuspirone

Thomas L. LaPorte,* Mourad Hamedi, Jeffrey S. DePue, Lifen Shen, Daniel Watson, and Daniel Hsieh

Bristol-Myers Squibb, Process Research and Development, One Squibb Drive, New Brunswick, New Jersey 08903, U.S.A.

Abstract:

This paper describes the development of a continuous, high yielding, and scalable enolization, oxidation, and quench process for the hydroxylation of the azapirone psychtropic agent buspirone to afford 6-hydroxybuspirone (6-hydroxy-8-[4-(4-pyrimidin-2-yl-piperazin-1-yl)-butyl]-8-aza-spiro[4.5]decane-7,9-dione). Two feed streams were reacted continuously using an in-line static mixer followed by oxidation in a continuous flow trickle-bed reactor. The laboratory reactor operation was demonstrated at steady state for over 40 h. The process was scaled up using both volumetric (enolization) and numbering-up (oxidation) scale-up strategies. A pilot-plant reactor was developed and successfully implemented in a three-batch campaign (47 kg input per batch).

Introduction

The batch process to prepare 6-hydroxybuspirone from buspirone was scaled up and implemented in a three-batch campaign (10−14 kg input per batch) as previously reported.[1] Although successful, several issues remained, particularly with respect to further scale-up. Oxidation reaction times at the pilot-plant scale (16−24 h) were much longer than those experienced in the laboratory (about 8 h). This was not surprising since the batch reaction operates in the mass transfer limited regime (even at −70 °C) and is subject to the mass transfer characteristics of the reactor. Another issue of concern for scale-up in the pilot plant and in manufacturing was the need for a cryogenic-capable (batch reaction at −70 °C) large-scale reactor. In addition, from a risk perspective, an entire batch may be compromised due to an overcharge of the base or loss of cooling capability at −70 °C. Another major concern was process safety. The batch process was operated by sparging air into the batch and sweeping the head space of the reactor with nitrogen. The larger scale makes uniform purging of the headspace more difficult and failure of potentially greater impact.

Due to these concerns, a continuous process was considered. A continuous process has three main advantages compared to the corresponding batch process. These advantages are related to safety, quality, and economics.[2] Enhanced safety results from smaller reactors with lower levels of active hazardous materials present at any given time in the process.[3] In addition, improved mass and heat transfer are characteristic of the smaller geometries. This translates to better control and possible mitigation of many scale-up effects. In addition, this processing mode allows operation at higher temperatures for short periods for processing of otherwise unstable intermediates.[4] Reactions under cryogenic conditions may give way to more moderate temperature conditions. Continuous processes are characterized by multiple feed tanks with segregated starting material, intermediates, and reagents. The feed solutions should be homogeneous and stable for extended periods of time. Typically, the completed reaction mixture flows into a quench tank or is quenched continuously and collected in a tank or drums. Plug flow reactors (for homogeneous reaction mixtures) are the reactors of choice because of their inherent narrow residence time distribution.[5] Most reactors consist of an in-line mixer (microreactor, static mixer, etc.) and a subsequent extension of residence time to complete the reaction.[6] Heat exchangers and custom-made jacketed tubing (tube-in-tube or shell-in-tube) are often used for the extension of residence time.[7,8] The merits of converting a batch process to a continuous process need to be evaluated on a case-by-case basis. This mode of processing was applied to the production of 6-hydroxybuspirone.

The evolution of the continuous process will be described in this paper from feasibility studies to the development of a laboratory unit reactor. Process analytical technology (PAT) was applied to the startup of the process to ensure the quality of the product and to minimize waste production. Finally, the process was scaled up and implemented in a pilot plant for a large-scale campaign to provide the active pharmaceutical ingredient (API) for clinical trials, formulation development, and drug safety evaluation.

* Author for correspondence. E-mail: thomas.laporte@bms.com.

(1) Watson, D. J.; Dowdy, E. D.; DePue, J. S.; Kotnis, A. S.; Leung, S.; O'Reilly, B. C. Development of a safe and scalable oxidation process for the preparation of 6-Hydroxybuspirone: Application of in-line monitoring for process ruggedness and product quality. *Org. Process Res. Dev.* **2004**, *8*, 616.

(2) LaPorte, T. L.; Wang, C. Continuous processes for the production of pharmaceutical intermediates and active pharmaceutical ingredients. *Curr. Opin. Drug Discovery Dev.* **2007**, *10* (6), 738–745.

(3) Proctor, L. D.; Warr, A. J. Development of a continuous process for the industrial generation of diazomethane. *Org. Process Res. Dev.* **2002**, *6*, 884–892.

(4) Roberge, D. M.; Bieler, N.; Thalmann, M. Microreactor technology and continuous processes in the fine chemical and pharmaceutical industries. *PharmaChem* **2006**, 14–17.

(5) Brechtelsbauer, C.; Ricard, F. Reaction engineering evaluation and utilization of static mixer technology for the synthesis of pharmaceuticals. *Org. Process Res. Dev.* **2001**, *5*, 646–651.

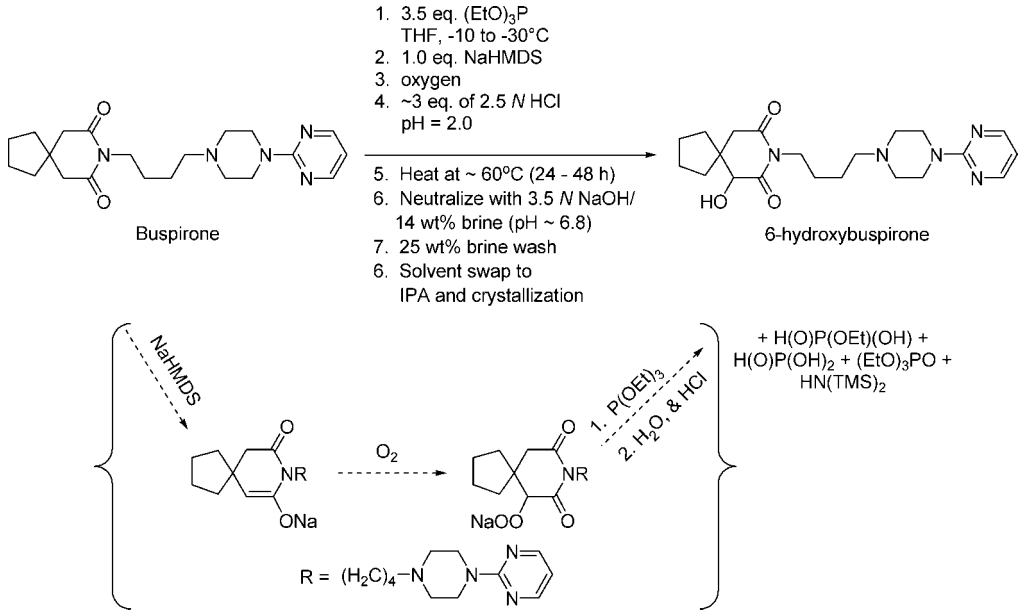
(6) Choe, J.; Kim, Y.; Song, K. H. Continuous synthesis of an intermediate of quinolone antibiotic drug using static mixers. *Org. Process Res. Dev.* **2003**, *7*, 187–190.

(7) Wakami, H.; Yoshida, J. Grignard exchange reaction using a microflow system: from bench to pilot plant. *Org. Process Res. Dev.* **2005**, *9*, 787–791.

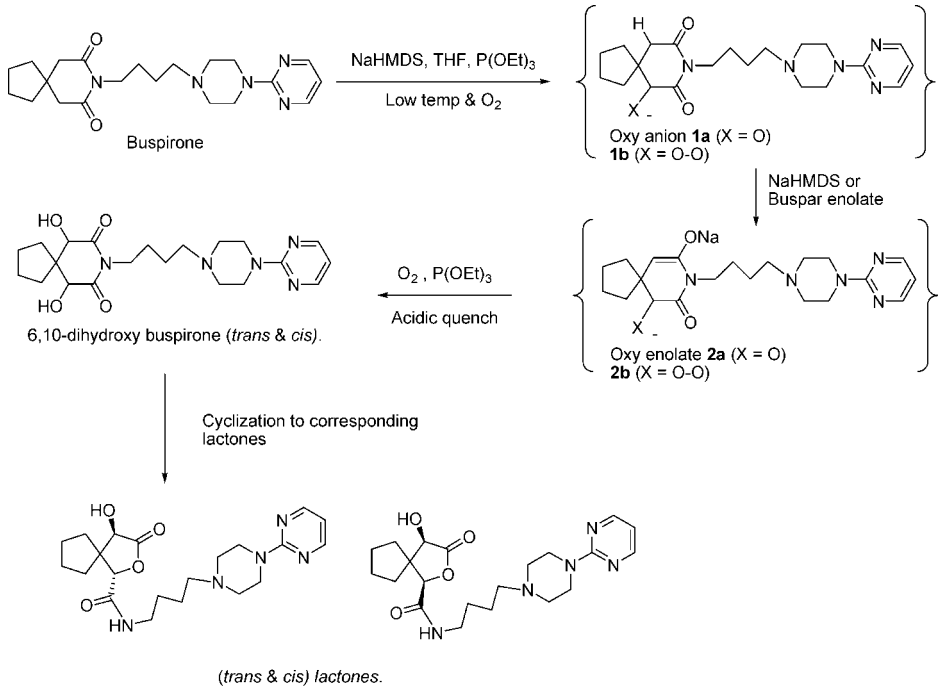
(8) Iwasaki, T.; Kawano, N.; Yoshida, J. Radical polymerization using microflow system: numbering-up of microreactors and continuous operation. *Org. Process Res. Dev.* **2006**, *10*, 1126–1131.

10.1021/op800079u CCC: $40.75
Published on Web 08/23/2008

Scheme 1. **Preparation of 6-hydroxybuspirone**



Scheme 2. **Generation of diol impurities**



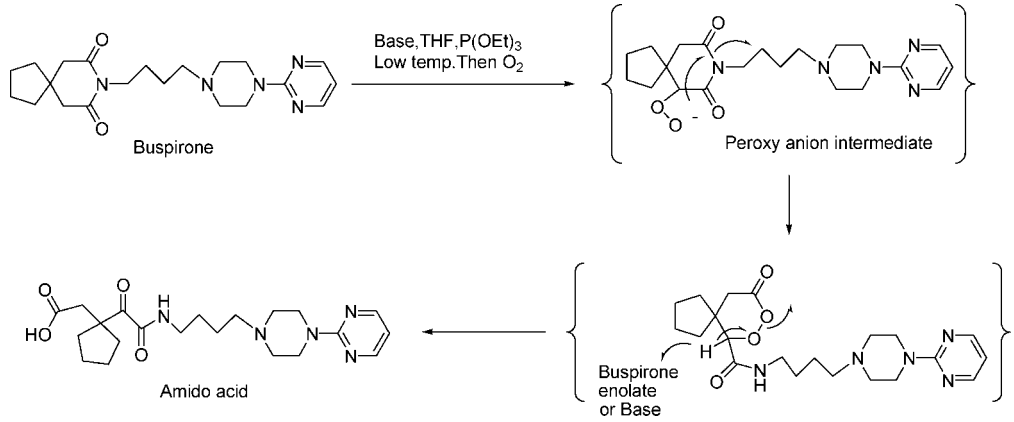
Results and Discussion

Process Chemistry. The batch process involves the generation of the sodium enolate of buspirone with sodium bis(trimethylsilyamide) in tetrahydrofuran at or below −70 °C in the presence of an excess of triethylphosphite (3.5 equiv) (Scheme 1). The enolate generation is monitored *in situ* via FTIR, and after enolate generation is completed, subsurface sparging with air is commenced and maintained until the oxidation is completed.

It is imperative to carefully control the base to buspirone stoichiometry, since buspirone has two readily enolizable protons, and excess base results in the formation of diol impurities (Scheme 2). There is a direct correlation between the amount of excess base used with the level of diol impurities formed. In addition, even if a stoichiometric amount of base is employed, significant levels of diol impurities are produced if oxygen (from air) is introduced prior to complete enolization. Introduction of air prior to complete enolization may result in residual unreacted sodium bis(trimethylsilylamide) and unoxidized buspirone enolate. This may lead to the deprotonation of the intermediate 6-oxy anion (**1**), which can then be further oxidized to the diol impurities. It should be emphasized that we do not know if intermediate (**1**) exists as the peroxide anion (**1a**) or if, in fact, it exists as the reduced form 6-hydroxy anion (**1b**).

In order to efficiently and rapidly reduce the intermediate hydroperoxy anion, it is necessary to conduct the oxidation in the presence of an excess of the triethylphosphite, since the

Scheme 3. **Amido acid impurity formation in the absence of triethylphosphite**



intermediate peroxy anion can intramolecularly rearrange to the amido acid (Scheme 3). When the oxidation is conducted without triethylphosphite, a 1:1 mixture of starting material and the amido acid are isolated. As the charge of triethylphosphite is increased, the byproduct is correspondingly reduced. A systematic study revealed that the use of less than 3.0 equiv of triethylphosphite resulted in unacceptably high levels of the amido acid and other polar impurities. Using between 3.5 and 5.0 equiv of triethylphosphite suppressed generation of the amido acid and the other impurities to levels wherein they could be effectively removed during workup and isolation of 6-hydroxybuspirone.

Chemical Hazard Analysis. The batch process to prepare 6-hydroxybuspirone involves feeding air to a solution of deprotonated buspirone in THF. The deprotonation step has a slight exotherm of 41.9 kJ/mol and a calculated adiabatic temperature rise of only 4.2 °C. In contrast, the oxidation reaction has a moderate exotherm of 685 kJ/mol and a calculated adiabatic temperature rise of 68 °C. Since a flammable solvent and air are utilized in the process, care must be taken to inert vapor spaces and minimize the possibility of static discharge particularly in regions operated above the solvent's flash point. Dilution of oxygen-rich process streams with nitrogen is required. The process streams are considered thermally stable on the basis of calorimetry tests with the advanced reactive system screening tool (ARSST).[9]

Continuous Oxidation Feasibility Study. Continuous oxidation (a gas−liquid reaction) was considered the most challenging aspect of the process development effort. A single- and a two-stage microreactors were used to study the feasibility of a continuous oxidation process (Figure 1). In general, microreactors provide a small hold-up volume and excellent heat transfer characteristics.[10] They have also been used for multiphasic reactions involving liquids, gases and catalyst particles.[11] Although not ideally suited for gas/liquid reactions, the small reaction scale was amenable to our screening studies. For our studies, we used a CPC CYTOS microreactor.[12] These reactors have been used for numerous reactions including pharmaceutical applications.[13,14] In order to simplify the studies further, an enolate solution (deprotected buspirone at −70 °C) was used as one feed (3 mL/min) and oxygen (0.3 L/min) was used as the second feed. Coolant was recirculated to the microreactor jackets at a temperature of −10 °C. With a single stage microreactor, a conversion of 65 to 70% was obtained with a residence time of 2 to 3 min. A two-stage microreactor system led to an 85−92% conversion with a total residence time of 5 to 6 min. These preliminary results demonstrated that the reaction could be operated at elevated temperatures for short reaction times. The productivity of this nonoptimized laboratory system was about 300 g of 6-hydroxybuspirone per day. On the basis of the productivity of the microreactor and the projected demand for production, the microreactor was determined not to be practical for scale-up to pilot or manufacturing scale. A laboratory unit reactor with substantially higher productivity was needed for a scalable continuous process.

Scalable Laboratory Unit Oxidation Reactor. The objective of our laboratory studies was to develop a unit reactor with reasonable productivity which could then be scaled up or numbered-up by operating multiple units in parallel to achieve productivity targets. The enolate formation was expected to be

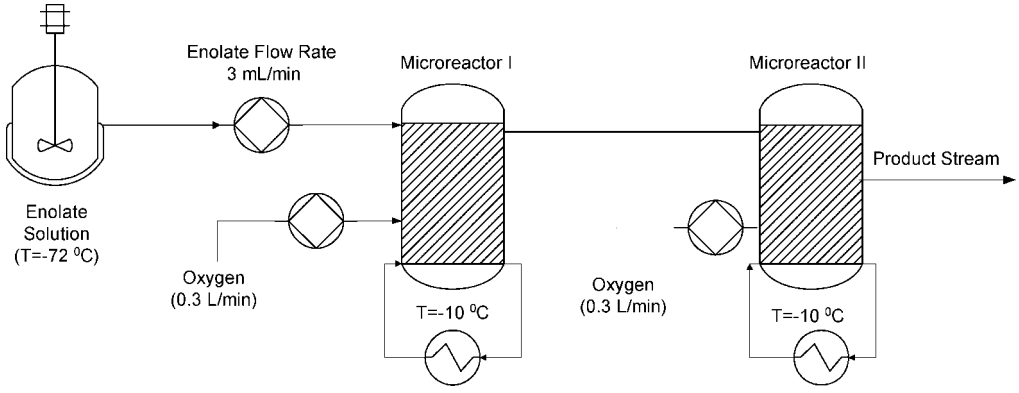
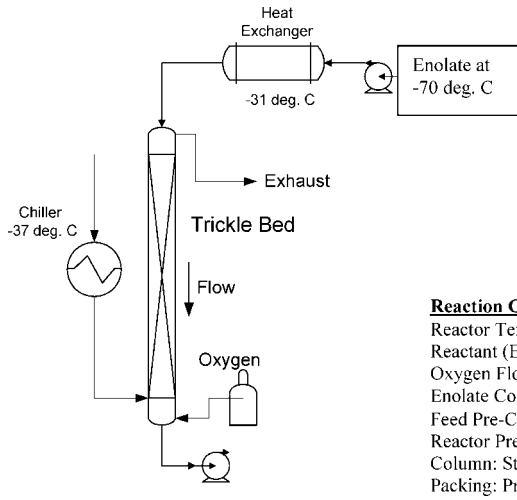


Figure 1. **The preparation of 6-hydroxybuspirone can be completed via a continuous process using at least two microreactors.**



Reaction Conditions
Reactor Temperature: -37°C
Reactant (Enolate) Feed: 39 – 45 mL/min
Oxygen Flow: 1 L/min
Enolate Concentration: 10 mL THF/ g buspirone
Feed Pre-Chiller Heat Exchanger: -31°C
Reactor Pressure: < 1 psig
Column: Stainless Steel, 60 in. x 7/8 in.
Packing: Pro-Pak, Stainless Steel, 0.24 in. x 0.24 in

Figure 2. **Continuous Oxidizing Reactor. The enolate solution was prepared and held at −70 °C. It was fed first to a heat exchanger prior to entering the trickle-bed reactor.**

easily scalable since it was a homogeneous liquid-phase reaction. The key challenge was scale-up of the oxidation reaction. We first tested a tubular plug flow reactor with the two-phase reaction stream (gas/liquid). Low conversions and quality were obtained compared to the microreactor. In general, a gas/liquid two-phase continuous reaction is not effectively fitted to a plug flow reactor scheme. Other reactors considered but not tested include a bubble column reactor, a membrane reactor, and a packed bed reactor. A conventional trickle-bed reactor with associated high mass transfer rates was an attractive choice for gas/liquid reactions. Several constraints led to early key design and process decisions. First, since this was a laboratory reactor, there was a practical limit to the length. For instance, the reactor had to fit into a laboratory walk-in hood when oriented in a vertical position utilizing gravity to enable trickle flow. Further, the exothermic nature of the reaction limited the diameter of the reactor tube in order to maintain a high surface area to volume ratio for heat transfer. A one-pass configuration was also desired to minimize complexity and to maintain a narrow residence time distribution.

Our prototype reactor was a 1 in. stainless steel (316 L) tube with an internal diameter of 0.875 in.. The total length of the tube varied up to a maximum of 60 in. by connecting two smaller reactors end to end. The tube(s) was jacketed to affect temperature control. The tube was tested with several commercially available packing materials including glass Raschig rings, stainless steel Heli-Pak and stainless steel Pro-Pak. Early screening experiments showed that the Pro-Pak packing gave the best results in terms of conversion and purities. As a result, all subsequent studies utilized the Pro-Pak packing.

In order to facilitate multiple reaction studies, the sodium enolate of buspirone (with triethylphosphite present) was generated as a batch solution at −70 °C (Figure 2). This preformed enolate solution was pumped through a heat exchanger to rapidly bring the temperature to about −31 °C after which it flowed into the oxidizing column while a countercurrent flow of oxygen was introduced from the bottom of the oxidizing column. The reactor was cooled by circulating coolant (−37 °C) through the external jacket. Operating parameters that were examined included: (1) flow rates for the enolate solution, (2) oxygen flow rates, 3) oxygen pressure, (4) heat transfer rates, and (5) operating temperature within the oxidation column. From our initial laboratory runs, we found that increasing the column length and the flow rate of the enolate solution resulted in higher conversion to 6-hydroxybuspirone and lower levels of impurities in the product stream exiting the oxidation reactor. However, as stated before, there was a practical limit to increases in reactor performance as a result of this strategy.

Considering the exothermic nature of the oxidation reaction, temperature control was a challenging and critical parameter for the process. The column was equipped with an external jacket through which a coolant was circulated in order to control the temperature within the reactor. The bulk of the oxidation reaction occurs in the first portion of the column. Consequently, the greatest heat load occurs in the first portion of the column. In contrast, the latter portion of the column has the lowest heat load associated with a lower level of reaction conversion as it approaches completion.

Counter-current operation of feed stream and gas flow was preferred over cocurrent operation because of higher reaction conversions and easier separation of the two flow streams. In order to obtain reasonable conversions with a balance between minimizing reactor length while maximizing throughput, pure oxygen gas was chosen over air as the oxygen source. The reactor was also tested with air at higher pressures to match

(9) ARSST (Fauske & Associates, LLC): Advanced Reactive System Screening Tool allows calorimetry studies at small scale (∼10 mL) under adiabatic conditions. Uses include thermal stability analysis, material compatibility testing, kinetic parameter evaluation, heat of mixing, heat of reaction, and determination of the self-accelerating decomposition temperature.
(10) Mason, B. P.; Price, K. E.; Steinbacher, J. L.; Bogdan, A. R.; McQuade, D. T. Greener approaches to organic synthesis using microreactor technology. *Chem. Rev.* **2007**, *107*, 2300–2318.
(11) Tadepalli, S.; Halder, R.; Lawal, A. Catalytic hydrogenation of *o*-nitroanisole in a microreactor: Reactor performance and kinetic studies. *Chem. Eng. Sci.* **2007**, *62*, 2663–2678.
(12) CYTOS microreactor is a product of CPC (Cellular Process Chemistry). The reactor consists of microstructured stacked plates configured with a cooling/heating capability. There are two input feeds and one output stream.
(13) Panke, G.; Schwalbe, T.; Stirner, W.; Taghavi-Moghadam, S.; Wille, G. A practical approach of continuous processing to high energetic nitration reactions in microreactors. *Synthesis* **2003**, *18*, 2827–2830.
(14) Schwalbe, T.; Kursawe, A.; Sommer, J. Application report on operating Cellular Process Chemistry plants in fine chemical and contract manufacturing industries. *Chem. Eng. Technol.* **2005**, *28*, 408–419.

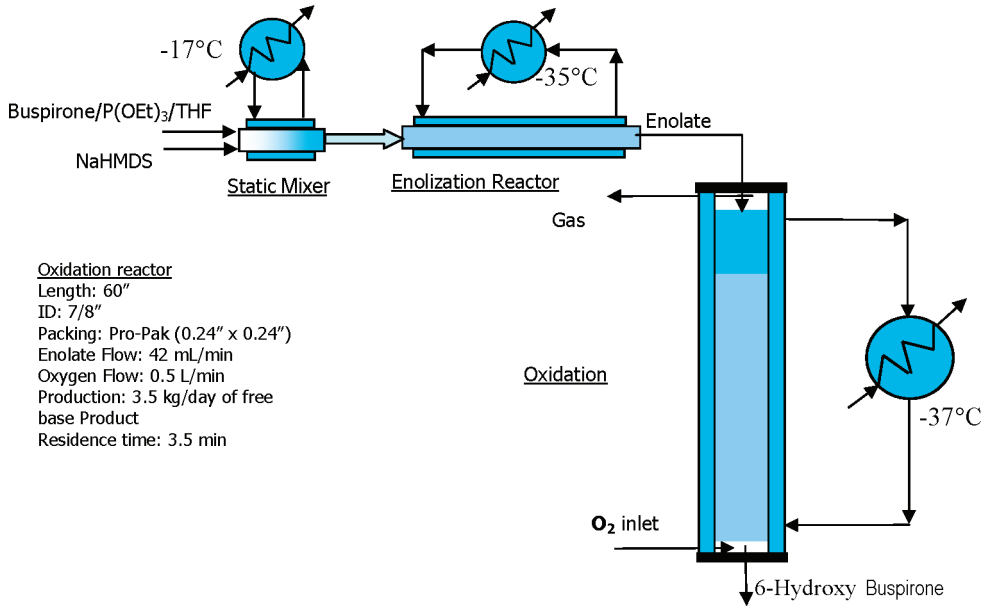


Figure 3. **Schematic of Continuous Enolization and Oxidation of Buspirone. Two stable feeds were prepared and held at room temperature.**

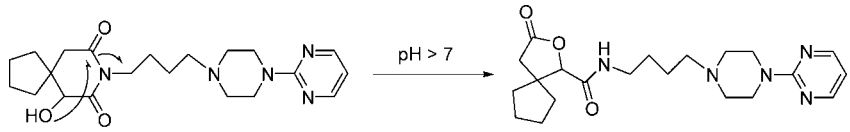
the partial pressures employed with atmospheric oxygen. Although higher conversions (compared to atmospheric air) were obtained, there was no process safety advantage to using high pressure air and the reactor pressure control was more difficult. As the oxygen flow rate was varied between 0.5 and 2.0 L/min (3.1 to 12.4 equiv of oxygen), there was no effect on conversion or impurity profile. Increasing the deprotonated buspirone solution feed rate decreased the mean residence time and the conversion. The liquid holdup in the reactor (60 in. long) was about 150 mL at an optimum enolate feed flow rate of 42 mL/min corresponding to a residence time of 3.6 min.

Continuous Enolization and Oxidation. After the optimal design and operating conditions for the oxidation had been established, we sought to implement a continuous enolization as opposed to generating the enolate via a batch process. As with any continuous process, two or more stable (preferably at room temperature) feed solutions were required for a viable process. In addition, the reagent(s) and starting material needed to be segregated. One feed consisted of buspirone, triethylphosphite and tetrahydrofuran. The second feed solution was sodium bis(trimethylsilylamide) dissolved in tetrahydrofuran. This was a standard strategy to separate the starting material (or intermediate) from the reagent. After several rounds of equipment design and experimentation, the equipment setup shown in Figure 3 was developed. The two feed solutions were pumped into a jacketed static mixer with a jacket coolant temperature of about −17 °C. After exiting the static mixer, the stream (partially enolized buspirone solution) flowed into a second jacketed (−35 °C) reactor which extended the residence time and allowed for complete enolization of buspirone prior to entering the oxidation reactor. In the oxidation reactor, the buspirone enolate was oxidized to 6-hydroxybuspirone via a counter current flow of oxygen. The residence time for the enolization of buspirone was approximately one minute and the residence time for the oxidation of the enolate was about 3.5 to 4 min. In contrast, the batch process required one to two hours for the enolization step and eight to twenty-four hours for the oxidation step. The difference for the enolization was process related and due to the time required for the temperature controlled slow addition of base to the buspirone solution. However, for the oxidation reaction, we attributed the dramatic difference in reaction time to a much higher oxygen mass transfer rate due to a higher gas−liquid mass transfer coefficient within the trickle-bed reactor and a higher oxygen driving force from using pure oxygen instead of air (higher partial pressure of oxygen in equilibrium with the liquid phase).

For our coupled continuous process (i.e., continuous enolization and oxidation), we carefully controlled the equivalents of base by adjusting the base feed pump rate. During a given run we would start the buspirone feed pump and set it at a fixed flow rate. Simultaneously, the base feed pump was started and its flow rate adjusted until an acceptable product stream purity profile was obtained. Typically, the ratio of buspirone feed solution to base feed solution was 4.9:1. Hence, a few percent changes in the base feed rate did not significantly affect the total flow rate. We analyzed samples of the product stream (post oxidation column) via HPLC. Using this protocol we could carefully adjust (*in situ*) the base/substrate ratio during startup which was particularly important as the base concentration and water content could vary from batch to batch. After standardizing a startup protocol, we were able to conduct a few small-scale laboratory preparatory runs using 150−300 g input of buspirone, wherein we obtained high-quality 6-hydroxybuspirone both as a crude product stream and upon isolation.

Multikilo Production. Having optimized our reactor design and operating conditions, the laboratory unit reactor was operated for an extended time period (40 h) to produce material to be isolated in our glass plant (kilo laboratory). This run was necessary to provide material in support of phase I clinical trials and to demonstrate the feasibility and steady-state operation of the continuous process. It is important to note that the laboratory unit reactor was not scaled up for this multikilo demonstration.

Scheme 4. **Formation of the lactone of 6-hydroxybuspirone**



However, feed tanks, collection vessels, and workup equipment were scaled up to meet the capacity demands of this demonstration run.

One vessel (feed tank #1) was charged with buspirone (6.8 kg), triethylphosphite (10.3 kg), and THF (60.5 kg). A second vessel (feed tank #2) was charged with sodium bis(trimethylsilylamide) in THF (17 kg, ~18.8 L). The two feed streams (fed by pumps) were first mixed in a jacketed static mixer unit (−15 °C jacket temperature) which was followed by a plug flow reactor (−35 °C jacket temperature) resulting in a one minute residence time. The enolate solution then flowed into the oxidation reactor where a counter current flow of oxygen oxidized the buspirone enolate to 6-hydroxybuspirone. As the product stream exited the oxidation column it was pumped into a stirred mixture of MTBE and 1.0 N HCl (aq) as a reverse quench. The initial pH was low and rose during the quench but did not exceed 6.5 (pH). Careful pH control was necessary, since 6-hydroxybuspirone can readily cyclize to its corresponding lactone particularly at basic pH (Scheme 4). The addition of MTBE to the quenched mixture facilitated the separation of the organic and aqueous phases, which further reduced the amount of lactonization since the lactonization of 6-hydroxybuspirone had been found to be slower in the separated organic phase. When the quench operation was conducted in the absence of MTBE, five to seven area percent lactone (measured via HPLC) would be formed during the continuous oxidation run of several days.

In-process results from the oxidation were excellent (88 to 90 area % 6-hydroxybuspirone). Figure 4 displays the impurity and product profiles for the duration of the 48 h run. Only slight variations were observed reflecting the stability of the process at steady state. The product stream contained typical levels of the unreacted buspirone (7%) and polar, early eluting byproducts (3%).

Process Analytical Technology: FTIR for Process Startup. In the batch process for 6-Hydroxybuspirone,[1] *in situ* FTIR was used to monitor the level of both the buspirone and enolate in solution during deprotonation. Since the deprotonation was rapid, the charge of base could be adjusted nearly real-time to minimize excess levels. This allowed titration of the correct amount of base for generation of the enolate prior to oxidation in the batch process. This methodology was then adapted to the continuous process by utilizing a flow cell for the FTIR probe (React-IR DiComp probe) in-line with the continuous enolate formation. For process startup, the initial base feed rate was calculated from the molarity of the base and the reaction stoichiometry. The actual feed rate utilized was adjusted (at a fixed starting material flow rate) on the basis of the feedback from the in-line FTIR. The strategy was to start the base feed rate at slightly lower than the calculated or estimated flow rate. The flow rate was then systematically increased at 1% increments. With each incremental increase, a corresponding decrease of the buspirone starting material signal was observed. If no decrease was observed, this implied that all of the buspirone had been converted to enolate and the base feed rate was reduced approximately 1−3% so as to keep a slight undercharge of base and reduce the probability of formation of the diol impurity. This startup strategy worked well because the enolization process was relatively fast and reached equilibrium within several minutes. Once the optimal feed rates were

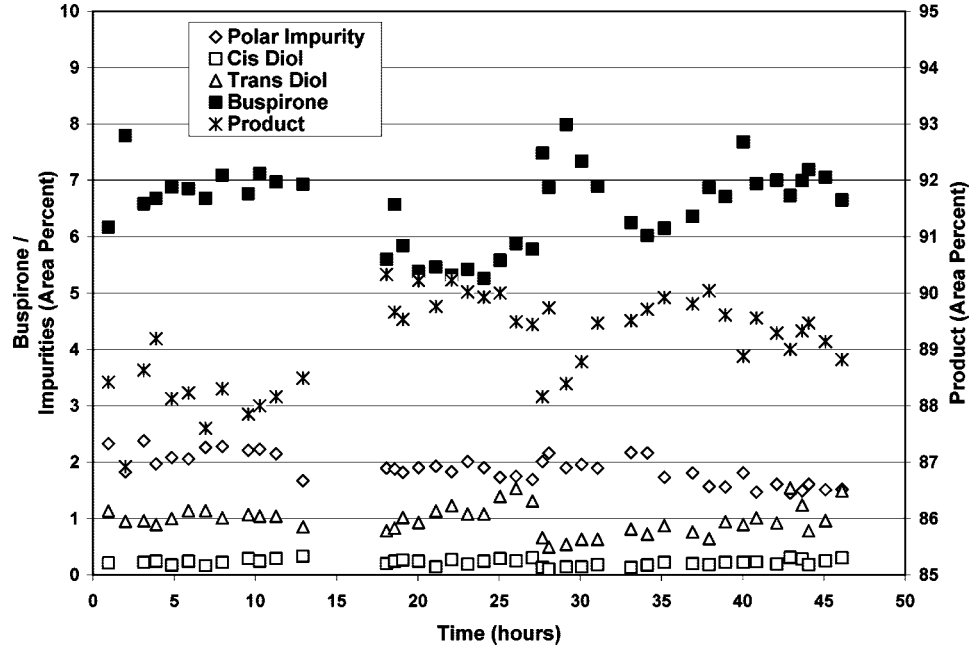


Figure 4. **The results of a multikilogram continuous enolization/oxidation process based on HPLC analysis is shown here.**

identified (as demonstrated via acceptable HPLC profiles obtained from sampling), the oxidation column was allowed to re-equilibrate (~3 residence times or about 10 min) for the new enolate composition. A sample was taken from the outlet of the oxidation column and checked by HPLC analysis. The results from this sample confirmed the proper ratio of base and buspirone and that the resultant product was of the proper quality.

This executed startup strategy helped ensure the quality of product during the startup operation. Another advantage of this startup strategy was to minimize losses of product due to extended startup times. During the startup procedure, the product stream went to waste. The product stream was not collected until a steady state was achieved and the product met the in-process control (IPC) criterion. In general, in-process controls (which elicit an action) are quality specifications built into the process to ensure the quality of the process. Once the IPC criterion was met, the product stream was diverted to the quench tank and collected for subsequent workup. By shortening the startup time, minimal product was wasted. The FTIR was also used to monitor the enolization during steady state operation.

Scale-Up of Continuous Enolization. The equipment for the enolization portion of the continuous process was composed of a static mixer to mix the two feed streams (buspirone and base) and a downstream heat exchanger which had sufficient volume to extend the reaction time for a given flow rate allowing complete enolate formation. Concerning scale-up of this setup, the operating range of a static mixer is fairly large. Moreover, a doubling of the diameter of a static mixer increases its capacity by a factor of 4. Consequently, minimal scale-up is necessary for a 10−100-fold capacity increase. In general, it is important that a heat exchanger meets the heat duty of the reaction; however, this was not a concern for the enolization reaction since it was weakly exothermic.

Numbering-Up of Oxidation Reaction. In general there are two approaches to scaling up a continuous process: (i) increasing the reactor volume while adjusting the stream flows to maintain a constant residence time and (ii) numbering-up the reactor. For this process, scaling up by volume was explored first. Increasing the column diameter results in lowering the heat transfer efficiency of the column and thus increases the overall temperature in the column. In order to determine the reactor diameter limitation, we built two oxidation columns (1 in. and 2 in. tubes) with different inner diameters and equipped the columns with thermocouples measuring the temperature of the reaction stream at the center of the reactors. The inner diameters of these two columns were 0.875 in. and 1.875 in., respectively, corresponding to a volumetric scale-up factor of 4.6. The wall thickness of both columns was 1/16 of an inch.

Our laboratory unit reactor (1 in. tube) was operated with a flow rate of 42 mL/min with a jacket temperature of −37 °C. The steady-state temperature profiles along the column are shown in Figure 5. Two operating regimes are evident from the figure. The first regime shows an increase in temperature traveling down the column indicating that the heat generated by the reaction was initially greater than the cooling capability of the reactor. The second regime shows a decrease in the reaction stream temperature as the heat of reaction drops below

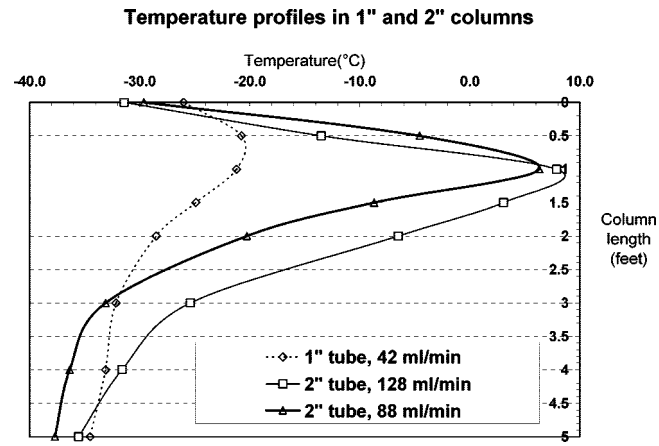


Figure 5. **Temperature profile along the oxidation reaction column. The curve with the open diamond symbols (◇) was obtained in the 1 in. column (7/8" i.d.) with a flow rate of 42 mL/min; the curves with the open triangle (Δ) and open square (□) symbols were obtained using the 2 in. column (1 7/8" i.d.) with flow rates of 88 and 128 mL/min, respectively.**

the cooling capability of the column. Another factor is that, as the reaction stream temperature increases, the oxygen solubility decreases further, lowering the mass transfer rate and thus the reaction rate and corresponding exotherm.

Scaling up to the larger diameter column (2 in. tube) should in theory allow a flow rate of 193 mL/min (maintaining a constant linear velocity). Considering the lower cooling efficiency due to scale-up (lower surface area to volume ratio for heat transfer), we tested the larger unit at modest flow rates of 88 and 128 mL/min. Tests with this column showed a dramatic increase in the center line temperature at the hottest point from −20 °C (1 in. column) to +6 °C while operating at 88 mL/min. A similar result was seen at 128 mL/min with a peak temperature of more than 8 °C. The temperature profiles here illustrate the limitations in a scale-up based on an increase in the column diameter. These higher center-line temperatures led to higher levels of impurities in the product stream. Increasing the column length would be one possible option for a scale-up as the heat transfer efficiency of the column would be maintained. However, that type of scale-up is not practical due to height limitations.

We judged that the best scale-up option for a reaction of this type carried out in a trickle-bed reactor was to number-up the reactor by integrating many reaction columns in one common shell. Adopting this strategy makes the scale-up quite straightforward since the heat and mass transfer characteristics inside the reactor are kept unchanged. Only the coolant flow rate has to be adjusted in order to maintain the same heat transfer capacity on the shell side.

The prototype reactor was tested by comparing its performance when one, two, or three of its tubes were used. The process setup is essentially the same as that shown in Figure 3. The key difference is that the enolate stream was split into two or three streams of equal flow rates using a combination of rotameters and metering valves. The test run was carried out by passing the enolate solution through one oxidation column with a total flow rate of 42 mL/min (standard operation). The system was shut down, restarted, and run with an enolate flow rate of 42 mL/min through each of the two designated columns

Table 1. **Proof of concept for numbering-up the oxidation reactor as shown by HPLC analyses**

			impurities		
mode	product AP	start. mat. AP	*trans*-diol AP	3.5 min AP	*cis*-diol AP
1 tube	88.5	6.85	0.98	2.18	0.16
2 tubes	88.1	6.55	1.17	2.25	0.18
3 tubes	91.4	5.34	0.85	1.17	0.16

in the multitube reactor. Similarly, a third run was then carried out by tripling the total original flow rates and running the same enolate flow rate through each of the three oxidation columns. Samples were taken from each run to assess the performance. Slight adjustments were made in the base flow rate to meet impurity specifications during each run. Table 1 summarizes the results that were obtained. The performance was based on the level of product, residual starting material, and three key impurities. The results for the three runs were similar with the only differences attributed to variations in the base feed rate (actual equivalents). This result indicated to us that numbering-up of the trickle-bed oxidation reactor would be a successful strategy.

The flow division of the enolate stream relied upon establishing an equal pressure drop (about 0.5−2 psi) across each stream, utilizing a metering valve. In addition, an in-line rotameter allowed visual verification of each flow rate. In general, the oxidation step could handle a flow rate variation of up to 10% (based on quality of product). The setup allowed adjustments of the metering valves if needed during processing. After flow division of the enolate stream, each stream was cooled individually through a multitube heat exchanger to −25 to −30 °C prior to entry to the trickle bed. Each stream was then directed approximately 1−2 cm above the top of the packing in each column. After traversing the length of the reactor, the reaction streams would recombine in a common funnel region below the columns. Oxygen entered the reactor through this common region. The oxygen would flow up the reactor through each column with distribution dependent upon an equal pressure drop (very low) across the length of each packed column. This pressure drop was mainly attributable to consumption of oxygen by the oxidation reaction up the length of the column. The reactor was operated at a slightly positive pressure (about 1 psig) with the oxygen flow rate in excess of the stoichiometric required amount for complete reaction. As a result, the process was not sensitive to oxygen flow in each column.

Steady-State Operation in the Pilot Plant. In order to produce at least 75 kg of 6-hydroxybuspirone for clinical supplies, a pilot-plant trickle-bed reactor with four oxidation columns ("Quad" reactor) was constructed (Figures 6 and 7). The scaled-up enolate reactor, four-tube oxidation column (Quad reactor), and the continuous quench stirred tank reactor were implemented in the pilot plant for this campaign. In order to monitor the steady-state operation of the continuous process, periodic samples were taken from the Quad reactor output with the corresponding HPLC analyses shown in Figure 8. All samples that were taken met in-process control criteria throughout processing. The typical product quality in terms of HPLC area percent was 88−90% and comparable to the 48-h single

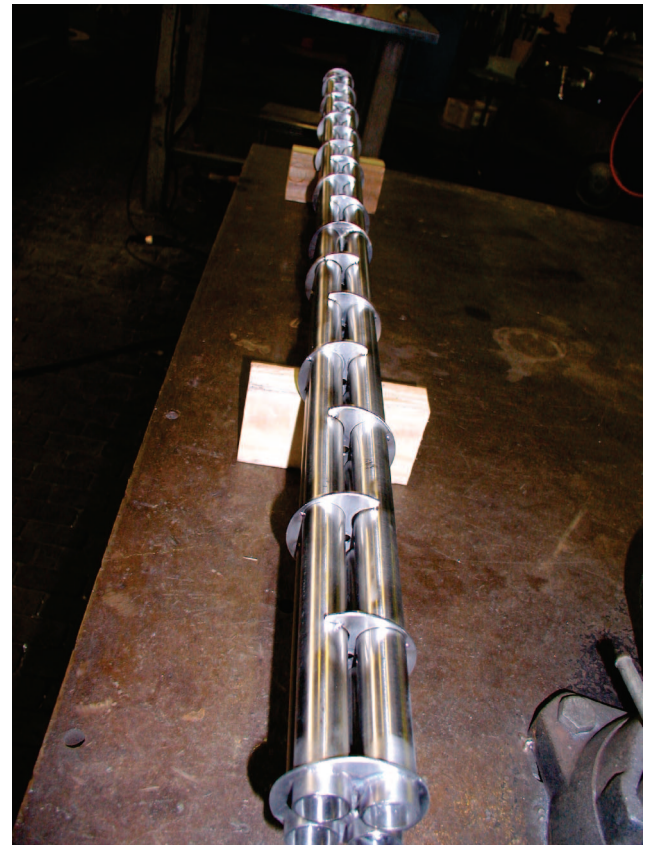


Figure 6. **Internal reaction tubes and extra-tubular baffles for the Quad four-tube trickle-bed reactor are shown. For each tube, diameter is 1 in., and the length is 72 in.**

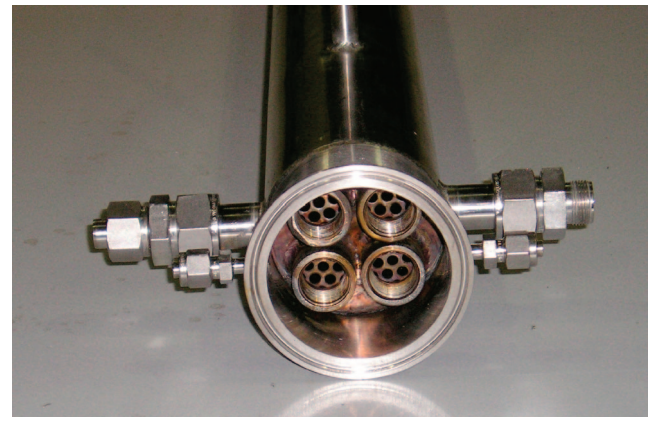


Figure 7. **Bottom of the Quad oxidation reactor is shown. Note, the grating holds the packing in each column. All columns terminate into a common region. The oxygen is fed to this common region and then traverses up each column.**

oxidation column run in the laboratory. The process was run for about 50 h without interruption and then was shut down for the weekend. A shutdown protocol was executed, and the system was held solvent wet and inert. Also, the two feed solutions were also held at room temperature under inerted conditions. The quenched product and feed solutions were stable at 20 °C. At the beginning of the following week, the process was resumed and completed. This also demonstrated how the process could be stopped and restarted at any time while minimizing any losses since only a small fraction of starting material/product was in-process at any given moment. Product

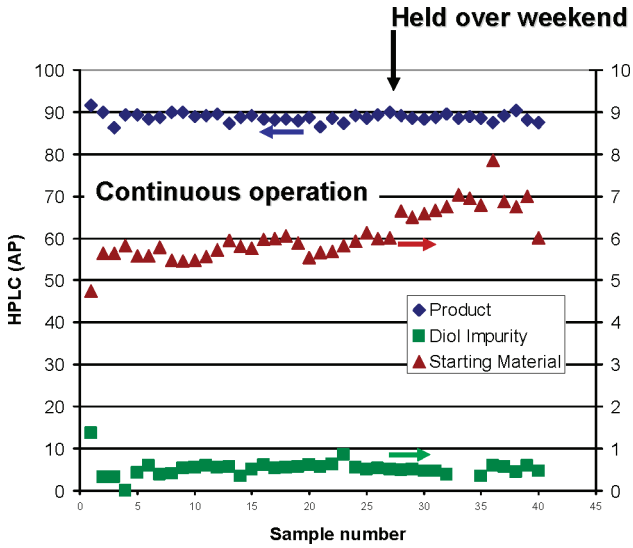


Figure 8. **Total run time was 72 h with 50+ h of continuous operation without interruption. The process was restarted and completed after a weekend hold.**

quality was similar throughout the run. The level of remaining starting material was only about 1% higher after restart. This was most likely due to a drift in one of the operating parameters or a slight variation in one of the feeds such as solvent evaporation. This could have been compensated by increasing the base feed rate. However, this was not done because the level of starting material in the product stream met IPC criteria. The total run time for the entire batch was 72 h. Total startup time was about 2.5 h. A total of 754 kg of quenched product solution were collected, containing 41.4 kg of product (83% of theoretical). The product was later isolated as described previously.[1] About 2.1 kg of product was in the startup waste drum (about 5% of theoretical batch). The impact of startup waste should be minimized as batches are operated for longer periods of time.

The Quad reactor processed 15.8 kg of buspirone each day. This was a 4 times numbering-up of the laboratory unit reactor. Because of the design, further numbering-up is not limited. The key to the performance of the Quad reactor and the larger multitube trickle-bed reactors is the equal flow distribution of the enolate stream. As with all numbering-up scenarios, the best flow distributor would be automated with feedback control of the individual resultant streams. Our pressure drop flow distributor using needle valves was adequate for our pilot-plant batches. However, an automated system would be recommended for manufacturing.

Conclusions

Conversion of a batch process to a fully continuous process was demonstrated for the production of 6-hydroxybuspirone. A much faster reaction was enabled due to higher operating temperatures and enhanced mass transfer in the continuous process. Widely variable batch reaction times (from 8 to 24 h) were supplanted by continuous reaction times of less than 4 min. The low coolant operating temperature was increased from −80 °C (or lower) to −38 °C with a significant increase in cooling efficiency (higher cooling surface area to volume ratio). The continuous process was inherently safer with a much smaller reaction volume (about 3 orders of magnitude) of a

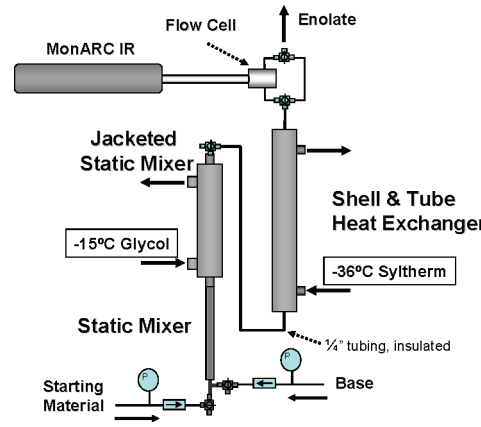


Figure 9. **The first continuous reactor is for enolate formation. The first static mixer is not jacketed, while the second one is jacketed (cooled by −15 °C glycol). The feeds are mixed and then cooled.**

flammable solvent in the presence of oxygen. In addition, only a small fraction of starting material or product was at risk (due to equipment failure or human error) at any time.

The continuous process developed in the laboratory proved to be a scalable process in the pilot plant. The scalability was due to the "numbering-up" mode of operation for the trickle-bed oxidation reactor. Laboratory operating parameters were replicated for each oxidation column operated in parallel. Heat and mass transfer characteristics were identical in each oxidation column. Steady-state operation was demonstrated in the pilot plant for over 50 h. Operating times were only limited by the amount of the feed solution available for processing. The use of process analytical technology was demonstrated to reduce startup time and allow an *in situ* control for determination of the appropriate base feed rate.

Three continuous reactors were operated in succession for production of a stable quenched product solution. Over 100 kg of API was produced during the pilot-plant campaign (three batches) using this setup. Moreover, the quality of the product produced by the continuous process was comparable to that of the batch process. Additional "scale-up" could be readily achieved along with extended operation times at the current pilot-plant "scale".

Experimental Section

Pilot-Plant Setup. The Quad reactor has two-zone cooling (for added flexibility) with baffles in the shell for greater efficiency. In the pilot-plant scale-up, we utilized the main reactor (1000 L) in the process suite as the starting material feed tank. A portable tank (400 L) was used as the base feed tank. The continuous enolate reactor (Figure 9) and Quad trickle-bed reactor (Figure 10) were arranged on a common skid with associated heat exchangers. For the pilot plant, we utilized a two-temperature mixing zone for enolization. The two feed streams proceeded into the first static mixer with their flows being diametrically opposed as they entered a 3/8 in. "T". The "T" was connected to a half-inch tube static mixer (Chemineer, 27 elements) which was not jacketed. This was connected to a second static mixer (same model as the first) which had a stainless steel jacket with inlet and outlet ports for coolant. The jacket coolant temperature was approximately −15 °C. We have

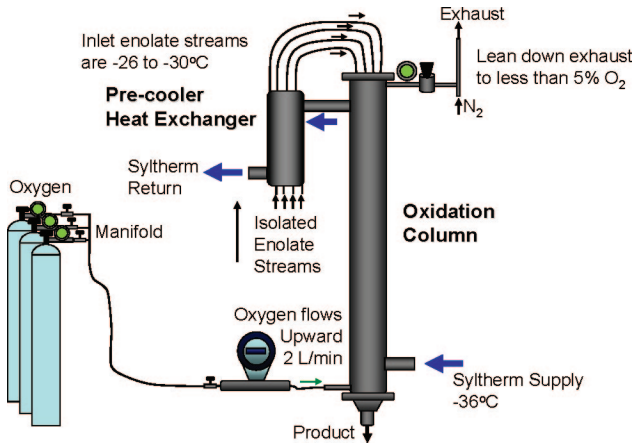


Figure 10. **Quad trickle-bed oxidation reactor: 3 in. outer pipe has internal baffles. The coolant is isolated from the inner tubes (4 × 1 in. tubes with packing). The packing is Pro-pak packing: 0.25 in. stainless steel. The column utilizes counter-current flow (liquid/gas). Enolate flows into the top of the reactor, and product combines at the bottom. The residence time is less than 4 min. The column operates at a pressure of 0.5 psig.**

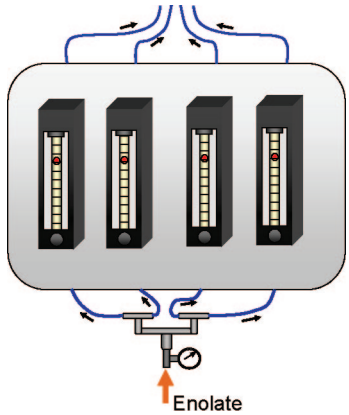


Figure 11. **Rotameters were used to evenly divide four enolate streams. Each channel was set to a pressure drop of 0.5 to 2 psi.**

found that the sodium bis(trimethylsilyl)amide can suffer from precipitation if it is cooled too quickly. This setup allows mixing at room temperature only for a few seconds prior to temperature reduction. The total residence time for the two mixers was about 8 s. The two-stage mixing was followed by a heat exchanger to ensure completion of the reaction. The total residence time was approximately one minute. The output stream from the enolate reactor was split into four equal streams using a needle valve and rotameter assembly (Figure 11). These four streams fed the Quad trickle-bed reactor. Glycol (−15 °C) was used to cool the enolate reactor. A chiller containing Syltherm (−37 °C) was used to cool the Quad reactor and the two associated heat exchangers. These were connected serially from jacket to jacket.

The four product streams from each packed column were recombined in a common funnel region at the bottom of the Quad reactor. The product then flowed (by gravity) to a level control chamber. This allowed a "liquid seal" to be maintained on the reactor outlet, ensuring that no oxygen gas flowed out the bottom of the reactor (Figure 12). The level control system utilized a differential pressure monitor and controller. The level output signal controlled a valve on the outlet of a piston pump

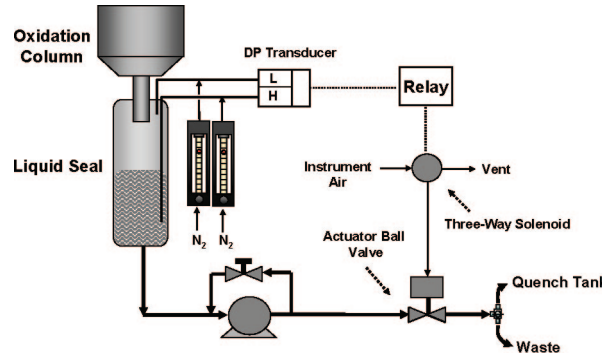


Figure 12. **A "liquid seal" was used to eliminate gas leakage from the bottom of the reactor. The level control was automated.**

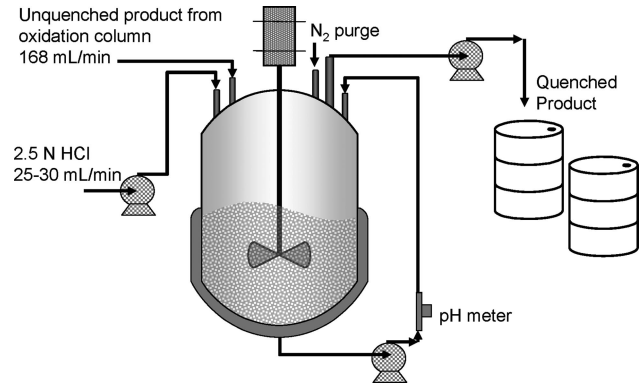


Figure 13. **The quench is the third continuous reaction that was implemented in the pilot plant. A 60-L Hastelloy reactor was used with the volume maintained constant at 25 L. A nitrogen purge was maintained in the headspace, and the pH was controlled between 1.0 and 2.5 by adjusting the continuous flow rate of 2.5 N HCl.**

which was used to pump the product stream to the quench tank (Figure 13). When the valve was closed, the pump recycled the product stream to its inlet. As a result, the liquid level in the control chamber would rise. A high level in the level control chamber would signal the valve to open, and the level would then fall as the product was pumped to the quench tank. This controller used high and low limits along with a preset dead-band for automatic control.

A small portable Hastelloy tank (60 L) was used as a continuous quench tank (with pH control). The product stream from the oxidation reactor was pumped into the quench at about 168 mL/min. The volume in the quench tank was maintained at a constant level of 25 L by fixing the height of a dip tube at the desired level. The dip tube was connected to a diaphragm pump which pumped continuously at a rate higher than that of the sum of input flows. The quenched product stream was diverted to a product drum. The pH in the quench tank was controlled between 1.0 and 2.5 by the continuous addition of 2.5 N HCl. This range provided robustness since a pH as high as 6.9 was determined to have no deleterious effects. The pH in the tank was monitored using an external pH loop and pump. The contents of the reactor were continuously agitated and the head space was swept with a continuous nitrogen flow. A condenser was present to minimize solvent losses. The residence time of the quench tank was about 2 h. This was not a design criterion but was chosen on the basis of available equipment

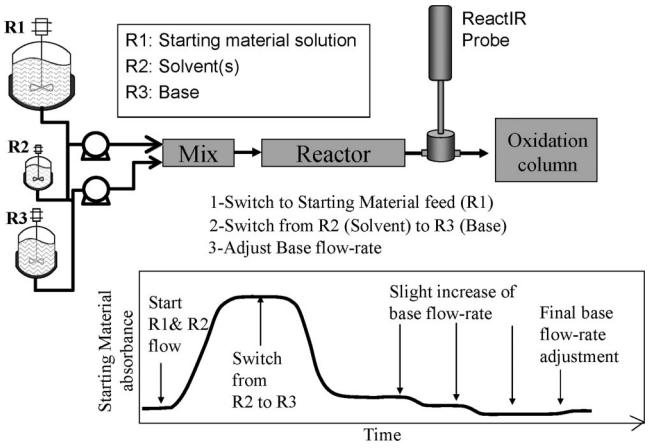


Figure 14. **Visualization of the startup procedure utilizing ReactIR. Initially the product stream is diverted to a waste drum. When product quality specs are met, flow is diverted to the quench tank.**

sizes. This larger tank size (larger than required) did dampen any possible pH swing due to pumping deviations.

As with any continuous process, flow control is of utmost importance in terms of quality and steady-state operation. For this process, precise control was needed to ensure the proper equivalents of base relative to the amount of starting material. We utilized magnetically coupled gear pumps that generated pulseless flow with a deviation of 1% or less. The pump skid utilized mass flow meters with feedback control of their individually set flow rates. The oxygen flow to the column used a mass flow meter and a manual needle valve for flow control.

Process safety is an important design consideration built into the process. Unreacted oxygen was vented from the top of the Quad reactor via a pressure regulator (set at 1 psig). The excess oxygen was diluted to less than 5% with nitrogen and sent to the plant's thermal oxidizer. The packed trickle-bed reactor also acted as a condenser (−37 °C) to minimize the release of tetrahydrofuran in the exhaust gas. The liquid holdup of the Quad reactor was approximately 600 mL. An emergency shutdown protocol was in place in the event of a process mishap. A two-way valve was in place to enable switching from oxygen to nitrogen (feed to the reactor) at any time. Two oxygen cylinders (outside the processing suite) were connected to a manifold to ensure continuous oxygen service. The temperature of the process stream throughout the column was below the corresponding flash point of the mixture.

Process Startup. The startup procedure was devised to minimize the wasting of starting material while achieving a steady state in an efficient manner. This incorporated the use of FTIR to identify and set the proper ratio and pumping rate of the base feed. A solvent (tetrahydrofuran) was first pumped through both feed lines (starting material feed and base feed) to allow thermal equilibration with all the equipment and to make sure all equipment was working properly (Figure 14). Thermocouples were located throughout the process lines to ensure target temperatures were achieved. Once the system was operating within normal parameters, the startup procedure was advanced. The starting material feed supply was switched from solvent to starting material. This was allowed to pump for about 5 to 10 min or until a steady starting material signal was obtained on the in-line FTIR. At that point, the gas supply to the Quad reactor was switched from nitrogen to oxygen. Also, the base feeding line was switched from solvent to base (NaHMDS in THF). This initiated the enolization of the buspirone. The progress was monitored by the FTIR as seen by the sharp drop in the starting material signal. When the FTIR readings were stable, the base was adjusted according to the procedure described in the previous section (Process Analytical Technology: FTIR for Startup). During this time, the oxidation of the enolate was proceeding in the Quad reactor. The output from the column was diverted to a waste drum. Samples were taken from the Quad reactor output and analyzed by HPLC. Once our IPC criteria were met, the Quad reactor output was directed to the quench tank for product collection. Samples were taken periodically to monitor and assess the state of the reactions. As long as pumping of the two feeds was controlled as expected and temperature control was satisfactory, product quality and yield were assured. The pH of the quench tank was controlled (pH of 1.0−2.5), and the output pump was activated to maintain a constant volume in the tank. Quenched and stable product was collected in drums for the entire batch.

The buspirone (starting material) feed solution consisted of 47.5 kg of buspirone free base, 72 kg of triethylphosphite (98%), and 422 kg of tetrahydrofuran. The base feed solution consisted of 138 kg of commercially purchased 1 M sodium bis(trimethylsilyl)amide in tetrahydrofuran. We used ultrahigh purity grade oxygen from cylinders. The acid feed solution was made from 142 kg of water and 45 kg of 37 wt % HCl.

The buspirone feed solution flow rate was set at 127 g/min. The base feed flow rate was set at 26 g/min initially. This was adjusted during startup on the basis of in-line FTIR analysis and HPLC analysis. In general, this value was within 3% of the steady-state operating feed rate. The oxygen feed was 2 L/min based on standard temperature and pressure. The flow of nitrogen to lean the exhaust of the oxidation reactor was 32 L/min. The oxidation reactor outlet pump was set to at least 200 mL/min. The quench tank outlet pump was set to at least 300 mL/min. The acid feed pump was set to approximately 25−30 mL/min. This was adjusted during the run to maintain our quench pH between 1.0 and 2.5.

Acknowledgment

We thank Eric Dowdy for his insightful discussions and John Shabaker for his assistance with the pilot-plant FTIR. We thank the Operations team and particularly Joe Corsentino, Thomas Mitchell, and Maria Gando during the setup and implementation of the pilot-plant process. We also thank the analytical support of Robert Falana and Qinggang Wang during the long batches in the plant. Special thanks to Tony Kukulski and Walter Suchowiecki who built our prototype reactors. Also, thanks to San Kiang, Yeung Chan, Simon Leung, and Frank Okuniewicz for their advice on the project. Finally, we would like to thank Jean Tom, Robert Waltermire, David Kronenthal, and Scott Jones for their feedback and advice on this manuscript.

Received for review March 31, 2008.

OP800079U