



# Kombination von On-Chip-Synthese einer fokussierten kombinatorischen Bibliothek mit computergestützter Vorhersage der biologischen Aktivität enthüllt Imidazopyridine als GPCR-Liganden\*\*

Michael Reutlinger, Tiago Rodrigues, Petra Schneider und Gisbert Schneider\*

**Abstract:** Am Beispiel der Ugi-Dreikomponentenreaktion stellen wir ein schnelles und effizientes Verfahren für die Kopplung von On-Chip-Mikroflusssynthesen mit einer neuen Methode zur Vorhersage biologischer Targets vor, um neue Liganden-Protein-Beziehungen zu entdecken. Wir konnten auf diese Weise eine GPCR-modulierende, kombinatorisch zugängliche Verbindungsklasse identifizieren. Diese effizienten Liganden binden mit antagonistischen Eigenschaften an die humanen Adenosin- $A_{1/2B}$ - und adrenergen  $\alpha_{1A/B}$ -Rezeptoren. Die Integration von Mikrofluidiksystemen für die chemische Synthese mit computergestützten Targetvorhersagen ist ein vielversprechendes Verfahren zur raschen Erstellung fokussierter Substanzbibliotheken mit hohen Trefferraten.

Die oftmals kurzen Laufzeiten von Wirkstofffindungs-Projekten zur Identifizierung chemischer Leitstrukturen werden häufig von Hochdurchsatz-Screeningkampagnen unterstützt. Die dafür verwendeten Substanzsammlungen profitieren auch von kombinatorischen Bibliotheken mit Leitstruktur- und Wirkstoffeigenschaften.<sup>[1]</sup> Während es bereits zahlreiche Synthesevorschriften gibt, wäre es wünschenswert, zusätzlich auch eine verlässliche Vorhersage der möglichen biologischen Ziele (makromolekulare Targets) für diese Moleküle zu erhalten, um Target-fokussierte Bibliotheken erstellen zu können. Am Beispiel der Ugi-Mehrkomponentenreaktion,<sup>[2]</sup> deren Robustheit bereits mit der Synthese von Modellsubstanzen und Wirkstoffkandidaten gezeigt wurde,<sup>[3,4]</sup> stellen wir ein schnelles und effizientes Verfahren für die Kopplung von On-Chip-Mikroflusssynthesen mit einer neuen Methode zur Vorhersage biologischer Targets vor, um neue Liganden-Protein-Beziehungen zu entdecken. Wir konnten auf diese Weise eine GPCR-modulierende, kombinatorisch zugängliche Verbindungsklasse identifizieren (GPCR = G-Proteingekoppelter Rezeptor). Diese effizienten Liganden binden mit antagonistischen Eigenschaften an die humanen Adenosin- $A_{1/2B}$ - und adrenergen  $\alpha_{1A/B}$ -Rezeptoren. Unsere Ergeb-

nisse zeigen, dass die Integration von Mikrofluidiksystemen für die chemische Synthese mit computergestützten Targetvorhersagen vielversprechend für die rasche Erstellung fokussierter Substanzbibliotheken mit hohen Trefferraten ist.

Imidazopyridine können in Bezug auf ihre Aktivität für verschiedene Zielproteine als privilegierte Substruktur angesehen werden.<sup>[5–8]</sup> Zwar wurde die Herstellung dieser Verbindungsklasse mithilfe der Ugi-Dreikomponentenreaktion bereits beschrieben,<sup>[5–7]</sup> allerdings konnte bisher keine Maßstabsvergrößerung der kombinatorischen Ansätze erreicht werden. Daher konzentrierten sich unsere ersten Versuche darauf, zunächst eine robuste Synthese in einem Mikrofluidikreaktor zu entwickeln, ausgestattet mit pulsationsfreien Niederdruck-Spritzenpumpen, der im kontinuierlichen Betrieb beliebige Mengen der entsprechenden Substanzen synthetisieren kann. Der Aufbau umfasst ein 3-2-Wege-Magnetventil, das die automatische Füllung und iterative Abgabe der Reaktanten ermöglicht. Das Amin und das Benzaldehyd wurden gemeinsam mit der Perchlorsäure in Ethanol gelöst; die Isocyanatverbindung wurde getrennt eingeleitet. Die Konzentrationen der Stammlösungen wurden so angepasst, dass die jeweils notwendige Endkonzentration im Mikroreaktor erreicht wurde. Als Primärreaktor wurde ein DeanFlow-Chip aus Borsilicatglas, mit einem Volumen von 5  $\mu\text{L}$  und einer Zick-Zack-Mischzone, verwendet (Abbildung 1 A). Als Alternative setzten wir einen KombiMix-

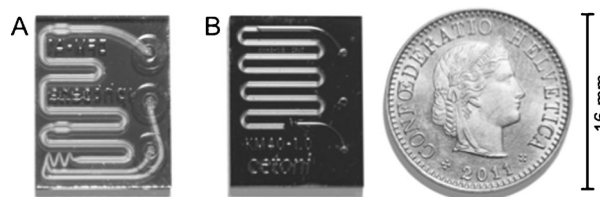
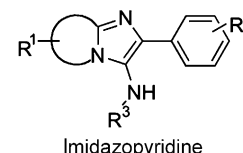


Abbildung 1. DeanFlow- (A) und KombiMix-Mikroreaktoren (B).

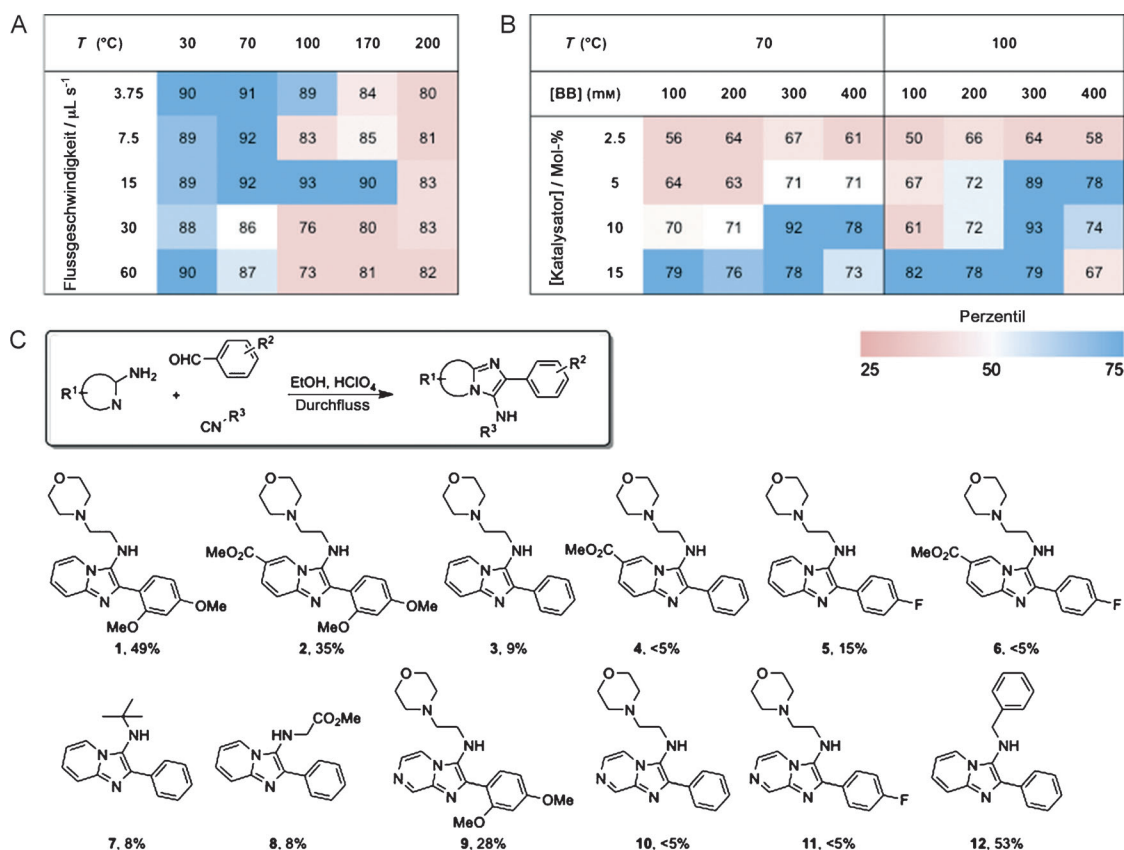
Chip mit einem Reaktionsvolumen von 13  $\mu\text{L}$  ein (Abbildung 1 B). Die Synthesevorschrift wurde mithilfe der Cetoni-Q<sub>mix</sub>-Elements-Software zur Automatisierung des gesamten Prozesses implementiert, inklusive aller erforderlichen Waschschritte.

Zur Optimierung der Reaktionsbedingungen wurde 1 sowohl durch klassische als auch durch automatisierte Synthese hergestellt. Die Umsätze wurden aus den  $^1\text{H}$ -NMR-Spektren abgeleitet (Abbildung 2 A,B). Als Erstes suchten

[\*] M. Reutlinger, Dr. T. Rodrigues, Dr. P. Schneider, Prof. Dr. G. Schneider  
Eidgenössische Technische Hochschule (ETH)  
Departement Chemie und Angewandte Biowissenschaften  
Wolfgang-Pauli-Strasse 10, 8093 Zürich (Schweiz)  
E-Mail: gisbert.schneider@pharma.ethz.ch  
Dr. P. Schneider, Prof. Dr. G. Schneider  
inSili.com GmbH  
Segantiniesteig 3, 8049 Zürich (Schweiz)

[\*\*] Diese Arbeit wurde finanziell von der OPO Stiftung (Zürich) unterstützt. GPCR = G-Protein-gekoppelter Rezeptor.

Hintergrundinformationen zu diesem Beitrag sind im WWW unter <http://dx.doi.org/10.1002/ange.201307786> zu finden.



**Abbildung 2.** Mikroflusssynthese von Imidazopyridinderivaten: A) Bestimmung der optimalen Flussgeschwindigkeit und Temperatur ( $T$ ) bei konstanten Konzentrationen von Katalysator (10 Mol-%) und Reaktantbausteinen (BB; 0.3 M); B) Bestimmung der optimalen Katalysator- und BB-Konzentrationen bei fester Flussgeschwindigkeit ( $15 \mu\text{L s}^{-1}$ ) und Temperatur (70 und  $100^\circ\text{C}$ ); C) die synthetisierte fokussierte Substanzbibliothek mit erzielten Ausbeuten.

wir eine optimale Flussgeschwindigkeit und Reaktionstemperatur unter Verwendung von 10 Mol-% Katalysator und einer Reaktantenkonzentration von 0.3 M, wie bereits früher beschrieben.<sup>[6]</sup> Im Allgemeinen funktionierte die Reaktion bei niedrigeren Temperaturen (30 und  $70^\circ\text{C}$ ) deutlich besser als bei 170 und  $200^\circ\text{C}$ . Im Rundkolben bei Raumtemperatur sowie bei  $30^\circ\text{C}$  erhielten wir Umsätze von 73 bzw. 80 %. Diese Befunde verdeutlichen die Nützlichkeit des Mikroreaktors, denn es werden sowohl der Umsatz erhöht als auch die Reaktionsdauer verkürzt. Reaktionen bei höheren Flussgeschwindigkeiten (30 und  $60 \mu\text{L s}^{-1}$ ) schnitten deutlich schlechter ab bei Geschwindigkeiten von 3.75 und  $7.5 \mu\text{L s}^{-1}$ , vermutlich wegen der geringeren Verweilzeit im Mikroreaktorchip. Die höchsten Umsätze ergaben sich bei mittleren Temperaturen (70 und  $100^\circ\text{C}$ ). Bei  $70^\circ\text{C}$  war der Umsatz beinahe unabhängig von der Flussgeschwindigkeit, während bei  $100^\circ\text{C}$  eine Flussgeschwindigkeit von  $15 \mu\text{L s}^{-1}$  bevorzugt war.

Nachdem die beste Kombination aus Temperatur und Flussgeschwindigkeit gefunden worden war, suchten wir nach den optimalen Konzentrationen von Katalysator und Reaktanten, die sich zu 10 Mol-% bzw. 0.3 M ergaben (Abbildung 2B). Die erreichten Umsätze sind vergleichbar mit den unter Mikrowellenbestrahlung erzielten (94 bzw. 93 %).<sup>[6]</sup> Dabei ist zu beachten, dass die Temperatur in der Durchflussmethode niedriger ( $100^\circ\text{C}$  im Mikroreaktor gegenüber

$170^\circ\text{C}$  unter Mikrowellenbestrahlung) und auch die Reaktionszeit kürzer war (0.3 s im Mikroreaktor gegenüber 15 min unter Mikrowellenbestrahlung). Als Letztes wurden die optimierten Reaktionsbedingungen in den beiden Mikroreaktionschips verglichen. Während für **1** im DeanFlow-Chip ein Umsatz von 93 % erzielt wurde, konnte im KombiMix nur ein Umsatz von 88 % beobachtet werden.

Nach der Reaktionsoptimierung synthetisierten wir eine kleine fokussierte Bibliothek von Imidazopyridinen **1–12** (Abbildung 2C) im DeanFlow-Reaktorchip. Die potenziellen biologischen Targets wurden mithilfe eines Gaußprozess-Regressionsmodells auf Basis bekannter Liganden für 469 Zielproteine, die in der ChEMBL-Datenbank (Version 14)<sup>[9]</sup> annotiert sind, vorhergesagt. Das neu entwickelte ComputermodeLL berechnet für Molekülstrukturen einen Aktivitätswert ( $p\text{Affinity}$ ) für jedes Target, was von ähnlichen Programmen zur Targetvorhersage bisher nicht geleistet wird.<sup>[10]</sup> Um aussagekräftige, nicht-triviale Vorhersagen zu erhalten, wurde die Mahalanobis-Distanz (MD) jedes vorhergesagten Wertes zu den Vorhersagen einer großen Sammlung zufällig ausgewählter Substanzen berechnet. Hierbei berücksichtigten wir nur diejenigen Zielproteine, für welche die vorhergesagte Güte  $p\text{Affinity} > 5.5$  mit  $\text{MD} > 0.5$  betrug. Mithilfe dieser leichten Einschränkung wurden im Mittel 18 Zielproteine pro Molekül erhalten. Dieser Wert liegt höher als bei anderen theoretische Betrachtungen und experimentellen

**Tabelle 1:** Zusammenfassung der Ergebnisse für **1, 3, 5, 7, 8, 9** und **12**.

	Zielprotein	vorhergesagte pAffinity	Mahalanobis- Distanz	exp. pK <sub>i</sub> oder % Bindung	LE <sup>[a]</sup>	LLE <sup>[b]</sup>	SILE <sup>[c]</sup>
<b>1</b>	$\alpha_{1A}$ <sup>[d]</sup> /PDE10A <sup>[e]</sup>	5.7/5.7	0.7/0.8	< 4/ < 4	–	–	–
<b>3</b>	$\alpha_{1B}$ <sup>[f]</sup>	6.2	2.4	5.6	0.33	3.46	3.04
<b>5</b>	$\alpha_{1A}/A_{2B}$ <sup>[g]</sup>	5.8/6.5	0.7/2.4	5.4/5.2	0.30/0.29	3.07/2.86	2.87/2.76
<b>7</b>	$\alpha_{1B}$	6.1	2.0	5.7	0.40	1.74	3.23
<b>8</b>	A <sub>1</sub> <sup>[h]</sup>	5.7	3.2	> 80% <sup>[i]</sup>	–	–	–
<b>9</b>	A <sub>2B</sub> /PDE10A	6.4/5.8	2.6/1.7	< 4/ < 4	–	–	–
<b>12</b>	A <sub>1</sub>	6.0	3.3	> 80% <sup>[i]</sup>	–	–	–

[a] Ligandeneffizienz. [b] Lipophile Ligandeneffizienz. [c] Größenunabhängige Ligandeneffizienz.

[d] Adrenerger  $\alpha_{1A}$ -Rezeptor. [e] Phosphodiesterase 10A. [f] Adrenerger  $\alpha_{1B}$ -Rezeptor. [g] Adenosin-A<sub>2B</sub>-Rezeptor. [h] Adenosin-A<sub>1</sub>-Rezeptor. [i] Radioligandentest. Aktivitätswerte sind gemittelt ( $n = 2$ ).

Ergebnissen, die zwischen 2 und 10 Targets pro Substanz annehmen, vermutlich weil die von uns gewählte Affinitätsschwelle relativ niedrig liegt.<sup>[11]</sup> Werden die Kriterien jedoch auf  $pAffinity > 6$  und  $MD > 1$  verschärft, verbleiben im Mittel nur noch vier Targets für jedes Imidazopyridinderivat. Insgesamt 41 Targets erfüllten diese Einschränkungen und wurden in der Folge genauer untersucht. Für diese Zielproteine ergab das Modell gute Werte bei der Kreuzvalidierung von  $Q^2 = 0.68 \pm 0.10$ ,  $MAE = 0.65 \pm 0.11$  und  $BEDROC = 0.67 \pm 0.15$  (alle Werte Mittelwert  $\pm$  Standardabweichung).<sup>[12]</sup>

Es wurden schlussendlich fünf Zielproteine auf Grundlage der Mehrheitsvorhersagen für die gesamte kombinatorische Bibliothek ausgewählt, auch mit Blick auf ihre pharmazeutische Bedeutung und die Verfügbarkeit der biologischen Tests. Die vorhergesagten Affinitätswerte ( $pAffinity$ ) liegen im mikromolaren Bereich (Tabelle 1), wobei ihre berechnete Varianz recht hoch war. Diese Beobachtung unterstreicht die potenzielle Neuheit des chemischen Grundgerüsts gegenüber den bekannten Liganden aus der ChEMBL-Datenbank. Tatsächlich wurden unseres Wissens bisher keine ähnlichen Imidazopyridinderivate als Liganden für den Adenosin- oder adrenergen Rezeptor veröffentlicht.<sup>[20]</sup>

Obwohl wir für alle synthetisierten Substanzen mögliche Targets vorhergesagt hatten, testeten wir nur diejenigen Verbindungen, für welche die Aktivitätsvorhersagen am robustesten waren. Für eines der vorhergesagten Zielproteine, die Phosphatidylinosit-3-Kinase (PI3K), wurde erst kürzlich eine Aktivität für Imidazopyridinstrukturen nachgewiesen,<sup>[6]</sup> was unsere Vorhersage unterstreicht. Als Machbarkeitsnachweis haben wir daher eine Reihe von GPCRs als Zielmoleküle untersucht, um neue Aktivitätsinseln im chemischen Raum zu entdecken. In einem kompetitiven Radioligandentest, in welchem die direkte Ligand-Rezeptor-Bindung untersucht wurde, und in zellbasierten Tests konnten 71 % der Moleküle wie vorhergesagt als aktiv bestätigt werden (Tabelle 1).

Die Substanzen **3** und **7** zeigten antagonistische  $K_i$ -Werte von 2 bzw. 3  $\mu M$  am adrenergen  $\alpha_{1B}$ -Rezeptor. Molekül **5** erzielte vergleichbare antagonistische Aktivität am adrenergen  $\alpha_{1A}$ - und am Adenosin-A<sub>2B</sub>-Rezeptor. Für **8** und **12** wurde eine potente direkte Aktivität für den A<sub>1</sub>-Rezeptor bestätigt (84 bzw. 89 % Bindung bei 100  $\mu M$ ), allerdings zeigten beide Moleküle keine Aktivität im zellulären Test. Um die vollständigen Selektivitätsprofile der Substanzen zu untersuchen,

sind weitergehende Tests in einem umfangreichen GPCR-Reihentest erforderlich.

Etliche Qualitätsindizes wurden in der Vergangenheit vorgeschlagen, um die Priorisierung von aktiven Substanzen in der Wirkstoffforschung zu ermöglichen.<sup>[13]</sup> Nach diesen Indizes sind unsere Moleküle als Leitstrukturen anzusehen (Tabelle 1). Substanz **7** ist beispielsweise nur mäßig „dekoriert“ und weist eine hohe Ligandeneffizienz auf ( $LE = 0.40$ ;  $SILE = 3.23$ ), die eine Entwicklung als  $\alpha_{1A}$ -Re-

zeptor-Antagonist rechtfertigten würde. **3** hingegen verfügt über eine bessere Balance zwischen Affinität und berechnetem  $\log P(o/w)$  ( $LLE = 3.46$  gegenüber 1.74), bei gleichzeitig etwas schwächerer Ligandeneffizienz als **7**.

Eine weitere wichtige Erkenntnis aus unserer Studie ist zudem, dass die hier präsentierten Leitstrukturen strukturell unähnlich zu den nächsten Nachbarn aus den Trainingsdaten sind (Tanimoto-Strukturähnlichkeit = 0.16–0.30; Tabelle S1 der Hintergrundinformationen) und die Moleküle daher wahrscheinlich nicht mit einer klassischen ligandenbasierten Ähnlichkeitssuche ausgewählt worden wären.

Unser Ansatz zum Entwurf fokussierter kombinatorischer Molekülbibliotheken hat schnell und effizient zur Identifikation eines molekularen Gerüsts mit Aktivität an vier GPCRs geführt. Die Ergebnisse verdeutlichen, dass das Imidazopyridin-Grundgerüst ein privilegiertes Strukturmotiv ist, und zeigen einen Weg zur Integration von neuen Technologien in die Wirkstofffindung und chemische Biologie auf, gezeigt am Beispiel der On-Chip-Synthese und computergestützten Vorhersage von biologischen Targets. Im Licht jüngster Fortschritte in der Lab-on-a-Chip-Technologie<sup>[14]</sup> kann man sich einen vollautomatischen Moleküldesign-Roboter vorstellen, der die computergestützte Targetvorhersage, Auswahl von Reaktanten für die Mikrodurchflusssynthese und die Testung der Moleküle integriert.

## Experimentelles

Computergestützte Rechnungen: Zum Trainieren des gaußschen Prozessmodells<sup>[15]</sup> wurde die ChEMBL-Datenbank (Version 14) mit 1213242 verschiedenen Molekülen mit 10129256 Bioaktivitäten für 9003 Zielproteine verwendet.<sup>[9]</sup> Zielproteine mit weniger als 200 eingetragenen humanen Bioaktivitäten wurden aussortiert. Alle eingetragenen Aktivitäten wurden standardisiert zu  $pAffinity = -\log_{10}(Aktivität)$ . Der finale Datensatz enthielt 209293 Moleküle mit insgesamt 431313 Bioaktivitäten für 469 humane Targets. Die weitere Verarbeitung der Daten wurde mithilfe von Python (<http://www.python.org>) und Knime v.2.6.0<sup>[16]</sup> durchgeführt. Molekülstrukturen wurden mithilfe der „Wash“-Funktion in MOE 2012.10 vereinheitlicht (The Chemical Computing Group Inc.);  $\log P(o/w)$  wurde mit MOE berechnet. Es wurden zwei unterschiedliche molekulare Repräsentationen verwendet: ein topologischer Pharmakophordeskriptor (CATS2, 0–9 Bindungen, Typen-sensitive Skalierung)<sup>[6]</sup> und ein ECFP-ähnlicher binärer topologischer Fingerabdruck (Morgan fingerprint,  $Radius = 4$ , 2048 bit; RDKit: <http://www.rdkit.org>).<sup>[17]</sup> Die Modelle zur Eigenschaftsvorhersage wurden in Matlab R2012b (The

MathWorks Inc.) unter Verwendung des GPML-Moduls v3.1 (<http://www.gaussianprocess.org>) entwickelt. Die Qualität der Vorhersagen wurde mithilfe einer 10-fach stratifizierten Kreuzvalidierung beurteilt (kreuzvalidierter quadratischer Korrelationskoeffizient,  $Q^2$ ; mittlerer absoluter Fehler, MAE). Die Boltzmann-verstärkte Unterscheidung des ROC (BEDROC;  $\alpha = 56$ , die ersten 3 % tragen zu 80 % der Bewertung bei) wurde verwendet, um die Qualität der frühen Anreicherung der Vorhersagen zu beurteilen.<sup>[18]</sup> Es wurde eine konservative Abschätzung für pAffinity verwendet:  $\text{prediction} = \mu_* - \sigma_*^2$ , wobei  $\mu_*$  die mittlere Vorhersage des Modells ist und  $\sigma_*^2$  die vorhergesagte Varianz. Um signifikante von zufälligen Vorhersagen zu unterscheiden, wurde die Mahalanobis-Distanz für alle Aktivitätsvorhersagen berechnet:  $\text{MD}(\text{prediction}) = (\text{prediction} - \mu_r)/\sigma_r$ , wo  $\mu_r$  und  $\sigma_r$  die mittlere und die Standardabweichung einer randomisierten Vorhersageverteilung sind. Als Hintergrundverteilung haben wir 50000 Moleküle zufällig aus der ChemDB<sup>[19]</sup> ausgewählt.

Synthese: Stammlösungen der Reaktanten wurden in Ethanol angesetzt. Die Amin- und die Aldehydlösungen wurden vorgemischt und mit der Perchlorsäure versetzt. Zwei unabhängige Spritzenpumpen förderten sowohl die Amin-Aldehyd-Perchlorsäure-Mischung als auch die Ethanollösung des Isocyanats mit der entsprechenden Flussgeschwindigkeit. Die Reaktionskammer, die den Mikroreaktorchip enthält, wurde auf die jeweils gewünschte Temperatur vorgeheizt und das Rohprodukt in einem Gefäß aufgefangen. Das Rohprodukt wurde mithilfe präparativer HPLC (Acetonitril/Wasser jeweils mit 0.1 % Ameisensäure) bei Lösungsmittelgradienten von 30–95 % oder 5–50 % Acetonitril über 16 min aufgereinigt. Das Mikroflusssystem und die Q<sub>mix</sub>-Elements-Software stammen von Cetoni. Die Mikrowellen-initiierten Synthesen wurden, wie schon früher beschrieben, mit einem Biotage-Initiator in 1–2-mL-Gefäßen durchgeführt.<sup>[6]</sup>

Testung: Die Aktivitätsmessungen wurden von der Firma Cerep als Auftragsarbeit durchgeführt. Für ausführlichere Angaben verweisen wir auf die bereitgestellten Hintergrundinformationen.

Eingegangen am 4. September 2013

Online veröffentlicht am 26. November 2013

**Stichwörter:** Computerchemie · Kombinatorische Chemie · Mehrkomponenten-Reaktionen · Mikroflusssynthese · Wirkstoff-Forschung

- [1] a) R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer, G. S. Sittampalam, *Nat. Rev. Drug Discovery* **2011**, *10*, 188–195; b) M. C. Bryan, C. D. Hein, H. Gao, X. Xia, H. Eastwood, B. A. Bruenner, S. W. Louie, E. M. Doherty, *ACS Comb. Sci.* **2013**, *15*, 503–511.
- [2] I. Ugi, *Angew. Chem.* **1962**, *74*, 9–22; *Angew. Chem. Int. Ed. Engl.* **1962**, *1*, 8–21.
- [3] B. Beck, S. Srivastava, K. Khoury, E. Herdtweck, A. Dömling, *Mol. Diversity* **2010**, *14*, 479–491.
- [4] C. Kalinski, H. Lemoine, J. Schmidt, C. Burdack, J. Kolb, M. Umkehrer, G. Ross, *Synthesis* **2008**, 4007–4011.
- [5] H. Bienaymé, K. Bouzid, *Angew. Chem.* **1998**, *110*, 2349–2352; *Angew. Chem. Int. Ed.* **1998**, *37*, 2234–2237.
- [6] M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues, G. Schneider, *Mol. Inf.* **2013**, *32*, 133–138.
- [7] M. Hieke, C. B. Rödl, J. M. Wisniewska, E. Buscató, H. Stark, M. Schubert-Zsilavecz, D. Steinhilber, B. Hofmann, E. Proschak, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 1969–1975.
- [8] a) S. Meister, D. M. Plouffe, K. L. Kuhen, G. M. C. Bonamy, T. Wu, S. W. Barnes, S. E. Bopp, R. Borboa, A. T. Bright, J. Che, S. Cohen, N. V. Dharia, K. Gagaring, M. Gettayacamin, P. Gordon, T. Groessl, N. Kato, M. C. S. Lee, C. W. McNamara, D. A. Fidock, A. Nagle, T.-g. Nam, W. Richmond, J. Roland, M. Rottmann, B. Zhou, P. Froissard, R. J. Glynn, D. Mazier, A. Chatterjee, T. T. Diagana, E. A. Winzeler, *Science* **2011**, *334*, 1372–1377; b) T. Wu, A. Nagle, K. Kuhen, K. Gagaring, R. Borboa, C. Francek, Z. Chen, D. Plouffe, A. Goh, S. B. Lakshminarayana, J. Wu, H. Q. Ang, P. Zeng, M. L. Kang, W. Tan, M. Tan, N. Ye, X. Lin, C. Caldwell, J. Ek, S. Skolnik, F. Liu, J. Wang, J. Chang, C. Li, T. Hollenbeck, T. Tuntland, J. Isbell, C. Fischli, R. Brun, M. Rottmann, V. Dartois, T. Keller, T. Diagana, E. Winzeler, R. Glynn, D. C. Tully, A. K. Chatterjee, *J. Med. Chem.* **2011**, *54*, 5116–5130; c) A. T. Baviskar, C. Madaan, R. Preet, P. Mohapatra, V. Jain, A. Agarwal, S. K. Guchhait, C. N. Kundu, U. C. Banerjee, P. V. Bharath, *J. Med. Chem.* **2011**, *54*, 5013–5030; d) J. M. Wisniewska, C. B. Rödl, A. S. Kahnt, E. Buscato, S. Ulrich, Y. Tanrikulu, J. Achenbach, F. Rörsch, S. Grösch, G. Schneider, J. Cinatl, Jr., E. Proschak, D. Steinhilber, B. Hofmann, *Biochem. Pharmacol.* **2012**, *83*, 228–240; e) S. Nordhoff, H. Deppe, U. Abel, A. Feurer, I. Ott, G. Metz, EP1974729 A1, **2008**; f) B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, R. S. L. Chang, V. J. Lotti, D. J. Cerino, T. B. Chen, P. J. Kling, K. A. Kunkel, J. P. Springer, J. Hirshfield, *J. Med. Chem.* **1988**, *31*, 2235–2246.
- [9] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.* **2011**, *40*, D1100–D1107.
- [10] a) E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet, L. Urban, *Nature* **2012**, *486*, 361–367; b) J. Besnard, G. F. Ruda, V. Setola, K. Abecassis, R. M. Rodriguez, X. P. Huang, S. Norval, M. F. Sassano, A. I. Shin, L. A. Webster, F. R. Simeons, L. Stojanovski, A. Prat, N. G. Seidah, D. B. Constam, G. R. Bickerton, K. D. Read, W. C. Wetsel, I. H. Gilbert, B. L. Roth, A. L. Hopkins, *Nature* **2012**, *492*, 215–220.
- [11] a) A. A. Antolín, J. Mestres, *Therapeutic Targets*, Wiley, New York, S. 309–326; b) Y. Hu, J. Bajorath, *AAPS J.* **2013**, *15*, 808–815.
- [12] a) A. Tropsha, *Mol. Inf.* **2010**, *29*, 476–488; b) A. Nicholls, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 239–255; c) K. Roy, I. Mitra, *Comb. Chem. High Throughput Screening* **2011**, *14*, 450–474.
- [13] a) M. M. Hann, G. M. Keserü, *Nat. Rev. Drug Discovery* **2012**, *11*, 355–365; b) G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, *Nat. Chem.* **2012**, *4*, 90–98.
- [14] a) P. N. Nge, C. I. Rogers, A. T. Woolley, *Chem. Rev.* **2013**, *113*, 2550–2583; b) D. Lombardi, P. S. Dittrich, *Expert Opin. Drug Discovery* **2010**, *5*, 1081–1094; c) T. Tsukahara, K. Mawatari, T. Kitamori, *Chem. Soc. Rev.* **2010**, *39*, 1000–1013.
- [15] a) C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, **2006**; b) M. Rupp, T. Schroeter, R. Steri, H. Zettl, E. Proschak, K. Hansen, O. Rau, O. Schwarz, L. Müller-Kuhr, M. Schubert-Zsilavecz, K. R. Müller, G. Schneider, *ChemMedChem* **2010**, *5*, 191–194.
- [16] M. R. Berthold, N. Cebon, F. Dill, T. Gabriel, T. Koetter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL)*, Springer, Berlin, **2007**, S. 319–326.
- [17] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- [18] J. F. Truchon, C. I. Bayly, *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
- [19] J. H. Chen, E. Linstead, S. J. Swamidass, D. Wang, P. Baldi, *Bioinformatics* **2007**, *23*, 2348–2351.
- [20] Chemical Abstracts Service, SciFinder, <https://scifinder.cas.org/>, Columbus, OH, USA.