

KNOBLE: Ein wissensbasierter Ansatz zu Design und Synthese einfacher herstellbarer Liganden als Ausgangspunkt für eine strukturbasierte Leitstrukturoptimierung**

Christof Gerlach, Martin Münzel, Bernhard Baum, Hans-Dieter Gerber, Tobias Craan, Wibke E. Diederich und Gerhard Klebe*

Im Gedenken an Jörg Stürzebecher

Die erfolgreiche Entwicklung von effektiven Hochdurchsatz-Technologien für die biologische Testung von Substanzen hatte in der letzten Dekade zur Folge, dass die Bereitstellung ausreichender Mengen an niedermolekularen organischen Testverbindungen zum Flaschenhals der Wirkstoff-Forschung wurde. Methoden der kombinatorischen Chemie, mit denen sowohl möglichst diverse als auch fokussierte Substanzbibliotheken erzeugt werden können, haben sich daraufhin als wichtiges Kettenglied in der Wirkstoffentwicklung etabliert. Die ursprüngliche Hoffnung, durch die Bereitstellung einer Vielzahl von möglichst diversen Ligandbibliotheken schneller zu neuen potenziellen Treffern für eine Leitstrukturentwicklung zu kommen, hat sich jedoch nicht erfüllt.^[1-3]

Deshalb ist man dazu übergegangen, fokussierte Ligandbibliotheken zu entwerfen. Als Ausgangspunkt dient dazu ein privilegierter Zentralbaustein, der ein passendes Wechselwirkungsmuster zur Adressierung des konservierten Bindungsmotivs einer Proteinfamilie enthält.^[4] Versehen mit geeigneten Substituenten und Seitenketten sollen dessen Derivate anschließend die gewünschte biologische Aktivität gegenüber ausgewählten Mitgliedern einer Proteinfamilie erzielen.^[5-8]

Hier stellen wir KNOBLE (KNOWledge-Based Ligand Enumeration) als neue Strategie zum strukturbasierten Aufbau einer fokussierten kombinatorischen Bibliothek vor. Bei jedem Arbeitsschritt wird die synthetische Zugänglichkeit der Reaktanten und der Produkte berücksichtigt. Zur Überprüfung des Konzepts werden einzelne Moleküle aus der Bibliothek synthetisiert und anschließend biologisch getestet. Das Ziel dieser Methode ist die Identifizierung erster Leitstrukturen. Diese sollen mit geringem Aufwand synthetisier-

bar sein und über eine ausreichende Affinität zum Zielprotein verfügen, sodass eine Röntgenstrukturbestimmung im Komplex mit dem Protein möglich wird. Die so erhaltene Struktur des Protein-Ligand-Komplexes soll dann als Ausgangspunkt für die weitere strukturbasierte Optimierung der entdeckten Leitstruktur dienen.

Zum Aufbau einer virtuellen kombinatorischen Bibliothek wird ein Zentralbaustein ausgewählt, der über entsprechende Linkergruppen für die Einführung geeigneter Substituenten verfügen muss. Alle für die Synthese infrage kommenden Bausteine werden herausgesucht und in einer Computersimulation zu den kompletten Liganden zusammengesetzt. Dabei erfolgt die Auswahl der Seitenketten anhand der Gestalt und der physikochemischen Eigenschaften der Proteinbindetasche des Zielproteins. Proteinbindetaschen, insbesondere in Serinproteasen, können in klar definierte Subtaschen unterteilt werden, die dann getrennt voneinander untersucht werden können. Nach Identifizierung und Extraktion einer Subtasche können mit der Cavbase-Methode Pseudozentren für den Vergleich zugewiesen werden. Anschließend kann diese Beschreibung der Subtaschen für die Entdeckung ähnlicher Taschen mit Cavbase herangezogen werden.^[10-13]

Cavbase detektiert automatisch die Bindetaschen und beschreibt ihre physikochemischen Eigenschaften anhand von fünf generischen Deskriptoren. Die molekulare Erkennung wird durch die Bildung von Wasserstoffbrücken und hydrophoben Kontakten bestimmt. Über 130000 Proteinbindetaschen sind in der Cavbase-Datenbank gespeichert und über die Verknüpfung mit der Datenbank Relibase zugänglich.^[14] Beide Datenbanken beruhen auf Strukturinformationen zu Protein-Ligand-Komplexen aus der Proteindatenbank (PDB).^[15] Cavbase vergleicht alle gespeicherten Bindetaschen mit der zuvor definierten Suchanfrage. Als Ergebnis wird eine Bewertungsliste mit absteigender Ähnlichkeit zur gesuchten Bindetasche erhalten.

Bezüglich des Designs einer für die Zielstruktur spezifischen Molekülbibliothek liefert Cavbase nicht nur ähnliche Subtaschen, sondern auch die Bausteine von Liganden, die diese Bindetaschen besetzen. Die chemischen Formeln der Ligandbausteine aus den gefundenen Taschen werden extrahiert und für eine Substruktursuche in Chemikalienkatalogen verwendet. Es werden mögliche Ausgangsverbindungen herausgesucht, die diese Ligandbausteine enthalten und reaktive Gruppen tragen, die für chemische Umsetzungen verwendet

[*] Dr. C. Gerlach, M. Münzel, B. Baum, H.-D. Gerber, T. Craan, Dr. W. E. Diederich, Prof. Dr. G. Klebe
Institut für Pharmazeutische Chemie
Philipps-Universität Marburg
Marbacher Weg 6, 35032 Marburg (Deutschland)
Fax: (+49) 6421-282-8994
E-Mail: klebe@mail.uni-marburg.de

[**] Wir danken dem DFG-Graduiertenkolleg „Proteinfunktion auf atomarer Ebene“ (C.G.) und der DPhG „Stiftung zur Förderung des wissenschaftlichen Nachwuchses“ (W.E.D.) für die finanzielle Unterstützung.

 Hintergrundinformationen zu diesem Beitrag sind im WWW unter <http://www.angewandte.de> zu finden oder können beim Autor angefordert werden.

werden können. Sie werden ausgewertet und dienen zum Aufstellen eines allgemeinen Syntheseschemas zum Aufbau einer Molekülbibliothek mit den gewünschten Ligandbausteinen nach den Prinzipien der kombinatorischen Chemie.

Wir haben diese neue Strategie am Beispiel der Serinproteasen überprüft. Enzyme dieser Klasse spalten Polypeptidketten nach unterschiedlichen Erkennungsmotiven. Während der Spaltung entsteht ein kovalent verknüpftes Acyl-enzym, und nach dem Prozess verbleibt der N-terminale Teil der Peptidkette an der Protease. Deshalb haben Serinproteasen auf der ungestrichenen Seite ausgeprägte Subtaschen, um die Peptidkette des Substrats nahe der Spaltungsstelle zu binden. Zusätzlich wird die Bindung durch Wasserstoffbrücken zu dem Peptidrückgrat verstärkt.

Als Referenzprotein soll im Folgenden Thrombin dienen. Thrombin zeichnet sich durch drei klar definierte Subtaschen aus, in denen das Substrat zur Spaltung der Peptidbindung fixiert wird. Die Suchanfragen in der Datenbank Cavbase ergaben für jede Subtasche ca. 4000 Treffer mit ähnlichen Taschen. Darunter wurden für die Thrombin-S1-Subtasche 646 und die S3-Subtasche 550 Bindetaschen von anderen Vertretern der Serinprotease-Familie gefunden. Für die S2-Subtasche fanden sich nur 274 Einträge aus anderen Serinproteasen, was auf eine Selektivitätsbestimmende Besonderheit der Serinprotease Thrombin in diesem Bereich zurückzuführen ist. Die S2-Subtasche wird räumlich stark durch die 60er-Schleife begrenzt.

Bei genauerer Analyse der Ähnlichkeitsbewertungen für die S3-Subtasche fällt der hydrophobe Charakter auf, der durch einen Tryptophanrest am Boden der Tasche bedingt ist (z.B. Acrosin, Cathepsin G, Faktor VIIa, Faktor IXa, Faktor XIa, Chymase, Trypsin, Faktor Xa, C1s-Elastase, Tryptase, t-Plasminogen-Aktivator, Protein C und Granzym A). Für die S1-Tasche ergeben sich Treffer in folgenden Serinproteasen: Cathepsin G, Faktor VIIa, Faktor IXa, Faktor XIa, Trypsin, Tryptase, Protein C, Faktor Xa, t-Plasminogen-Aktivator, Enteropeptidase und Urokinase.

Aus den vielen als ähnlich entdeckten Taschen werden die darin gebundenen Ligandbausteine extrahiert. Dabei wird durch den Vergleich mit der Thrombinsubtasche darauf geachtet, dass nur solche Ligandbausteine entnommen werden, die ein ähnliches Wechselwirkungsmuster in Thrombin eingehen könnten. Bei der Cavbase-Suche für die S2-Tasche im Bereich der 60er-Schleife wird die spezifische Gestalt dieser Tasche deutlich: Unter den ersten 100 ähnlichen Taschen befinden sich 98 Taschen, die aus Proteinstrukturen anderer Thrombin-Komplexe stammen.

Ligandbausteine mit einem Isopropylrest, einem aliphatischen und heteroaliphatischen Ring (wie einem Pyrrolidinring) oder Bausteine mit aromatischen Ringen kommen sehr häufig vor (Abbildung 1).

Charakteristisch für die S1-Subtasche ist die Seitenkette von Aspartat 189, die hochkonserviert in eine durch hydrophobe Aminosäuren gebildete Umgebung eingebettet ist. Durch den Bindetaschenvergleich fallen Ligandbausteine mit gleichen physikochemischen Wechselwirkungseigenschaften auf, wobei das genaue Ausmaß an Ähnlichkeit in der Aminosäuresequenz des umgebenden Proteins keine Rolle spielt. Interessanterweise werden unter den besetzenden Ligand-

bausteinen nicht nur basische Reste gefunden, die die Aspartatseitenkette adressieren, sondern auch halogensubstituierte Phenylringe, die in Richtung einer Tyrosinseitenkette zeigen (Abbildung 1). Am deutlichsten variieren die chemischen Eigenschaften der Ligandbausteine für die S3-Tasche. Im Einzelnen werden für diese Subtasche Vorschläge für Reste aus Liganden gefunden, die sehr unterschiedlichen Vertretern der Serinprotease-Familie entstammen (Abbildung 1).

Hier zeigt sich die Stärke des vorgestellten Ansatzes, da nicht nur die bekannten Thrombininhibitoren zur Erstellung der kombinatorischen Bibliothek herausgesucht werden. Auch Ligandbausteine aus Proteinen, die ähnliche Subtaschen enthalten wie Thrombin und ähnliche Wechselwirkungen ausbilden, werden gefunden. Ein „Subtaschen-Hopping“ zu anderen Vertretern der Serinprotease-Familie wird möglich.

Auf die Auswahl eines geeigneten Ligandbausteins zur Adressierung der jeweiligen Subtasche folgte eine Suche im Sigma-Aldrich-Katalog mithilfe des Programms JME.^[16] Die Ausgangsverbindungen, die aus Suchanfragen für die verschiedenen Subtaschen hervorgingen, sind in den Hintergrundinformationen tabellarisch aufgeführt. Bei der Auswahl geeigneter Ausgangsverbindungen wurde darauf geachtet, dass eine allgemeine Synthesevorschrift erstellt werden kann, die jeweils die gleichen funktionellen Gruppen zur Verknüpfung der einzelnen Fragmente verwendet. Hier haben wir uns für Ester-, Amid- und Sulfonamid-Verknüpfungen und Reaktionen über nucleophile Substitutionen entschieden. Unter Einbeziehung synthesetechnischer Aspekte wurden die drei verschiedenen Bausteine (*S*)-Prolinol, (*S*)-2-Hydroxymethylpiperidin und (*S*)-Valinol für die Adressierung der S2-Subbindetasche als Zentralfragmente ausgewählt (Schema 1).

Als eine weitere Modifikation der P2-Bausteine wurde zwischen P2 und P3 ein Glycinrest eingeführt. Durch diese Gruppe soll zum einen eine Wechselwirkung mit dem Bereich der nichtspezifischen Peptiderkennung in der Protease ermöglicht werden, zum anderen kann die S3/S4-Tasche besser ausgefüllt werden. Die Verknüpfung des Zentralfragments mit den Bausteinen für die S1-Subtasche soll über eine Esterbildung erfolgen. Unter Verwendung von sekundären oder tertiären Aminen werden die verschiedenen Grundgerüste über eine Amid- oder Sulfonamidbindung bzw. über eine Substitutionsreaktion um die Bausteine zur Besetzung der S3-Bindetasche erweitert. In Schema 1 sind alle Carbonsäurederivate aufgelistet, die als Bausteine für die S1-Subtasche, ebenso wie die Carbonsäure-, Sulfonsäure- und Chlorderivate, die zur Adressierung der S3-Tasche dienten. Die einzelnen Bausteine wurden zunächst im Computer zusammengesetzt (Schema 2) und mithilfe des kombinatorischen Moduls von FlexX in das Protein gedockt.^[17,18] Zwei Bibliotheken mit jeweils 507 möglichen Molekülen – mit und ohne Glycinrest – wurden erzeugt. Für jedes während des kombinatorischen Dockings zusammengesetzte Molekül wurden die zehn besten Bindungsgeometrien gespeichert, die anhand der FlexX-Bewertungsfunktion selektiert worden waren. Die erhaltenen Bindungsgeometrien der beiden auf diese Weise erstellten kombinatorischen Bibliotheken wurden anschlie-

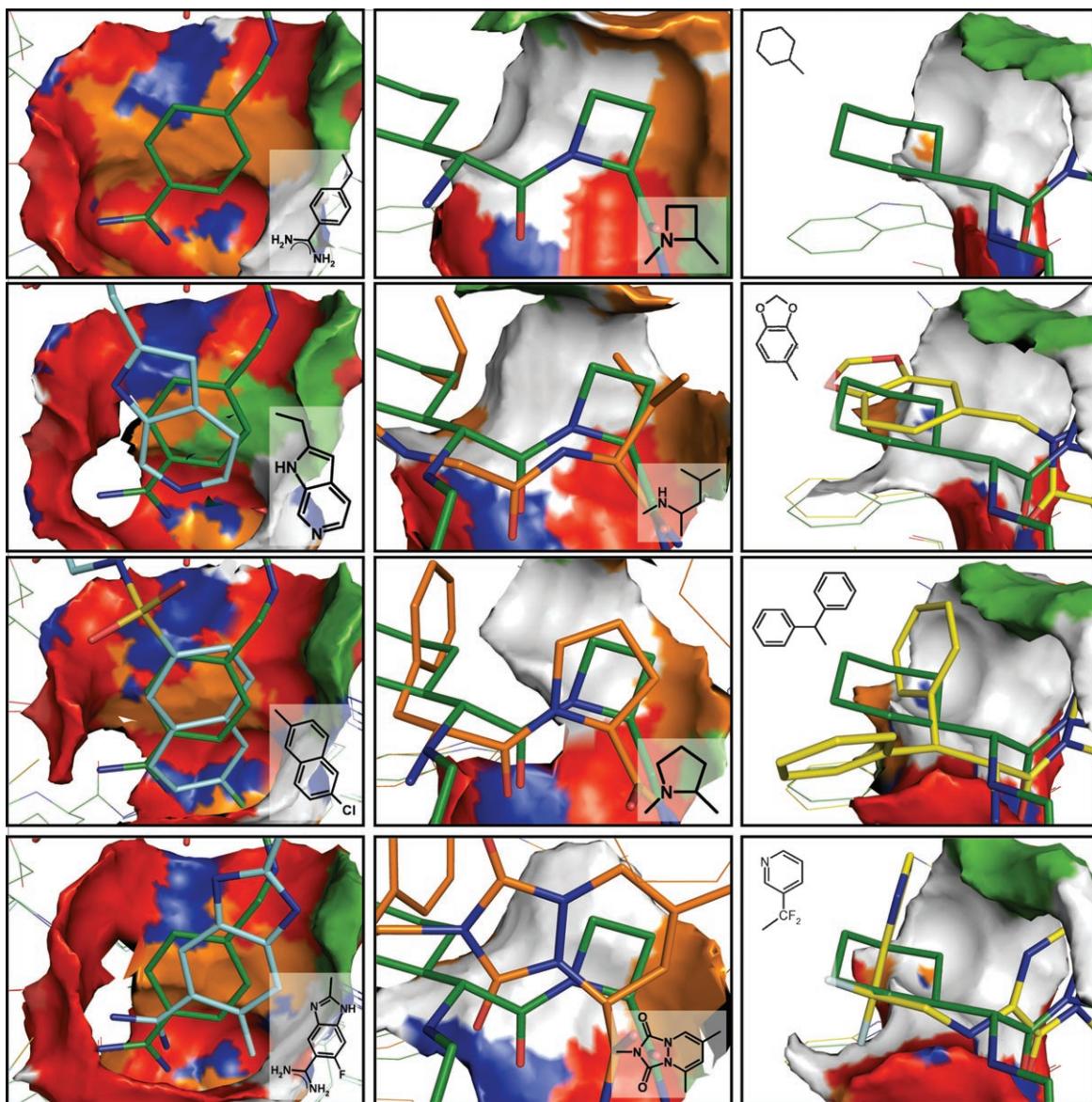
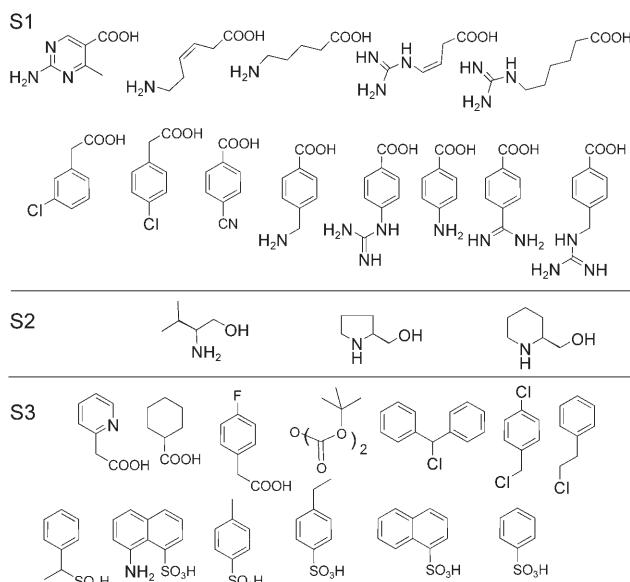


Abbildung 1. Ligandbausteine, die durch eine Subtaschensuche identifiziert wurden. Zum Vergleich ist die kristallographisch bestimmte Bindungsgeometrie von Melagatran (grün, PDB 1k22) in Thrombin^[9] (erste Reihe) für die S1- (links), S2- (Mitte) und S3-Tasche (rechts) mit drei Beispielen von Ligandbausteinen in ähnlichen Subtaschen anderer Proteine dargestellt (zweite bis vierte Reihe). Die dort angetroffenen Proteinoberflächen haben sehr ähnliche physikochemische Eigenschaften (blau: H-Brücken-Donor, rot: H-Brücken-Akzeptor, grün: Donor-Akzeptor-Funktionalität, weiß und orange: Bereiche für Wechselwirkungen mit aliphatischen bzw. aromatischen Gruppen). In der ersten Reihe sind die physikochemischen Eigenschaften der Taschen im Melagatran-Komplex gezeigt. Die folgenden Reihen weisen auf physikochemische Eigenschaften in sehr ähnlichen Subtaschen anderer Proteine. Die Bindungsgeometrie (hellblau, orange, gelb) und Strukturformeln der Ligandbausteine, die diese Subtaschen besetzen, sind ebenfalls angegeben.

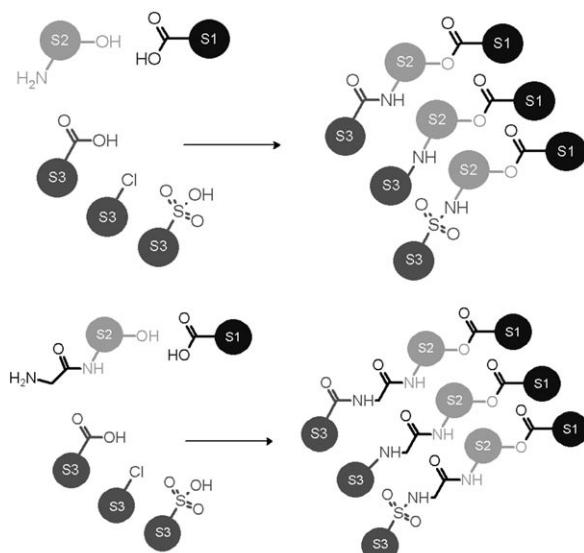
ßend mit DrugScoreCSD nachbewertet.^[19] Dazu wurden sie visuell überprüft, wobei sich die einzelnen Atom-Atom-Kontaktbeiträge zur Gesamtbewertung mithilfe einer Computergraphik veranschaulichen ließen.^[20]

Wegen der einfacheren Synthese haben wir aus der Liste am besten bewerteter Bibliotheksmitglieder die 3-Chlorbenzyl- und 4-Cyanphenyl-Bausteine zur Adressierung der S1-Tasche ausgewählt (Fragmente mit Amidino- oder Guanidinogruppen erfordern mehrere zusätzliche Syntheseschritte beim Aufbau der Liganden). Für die Auswahl der P2- und P3-Bausteine spielten ebenfalls ihre Zugänglichkeit und die

Reaktivität der gewünschten funktionellen Gruppen die entscheidende Rolle. Die letztlich für die S3-Tasche selektierten Bausteine, 2-Pyridylessigsäure, *p*-Fluorbenzoësäure und die *tert*-Butoxycarbonyl-Gruppe, fanden sich häufig als Bestandteile in den durch FlexX und DrugScoreCSD am besten bewerteten Liganden. Die Verbindungen **2–6** wurden synthetisiert und anhand eines Enzymkinetik-Assays^[21] als mikromolare Hemmstoffe identifiziert (siehe Schema 3 und die Hintergrundinformationen). Anschließend gelang es, **5** in Thrombinkristalle eindiffundieren zu lassen und eine Struktur des Komplexes mit diesem Liganden zu bestimmen (Ab-



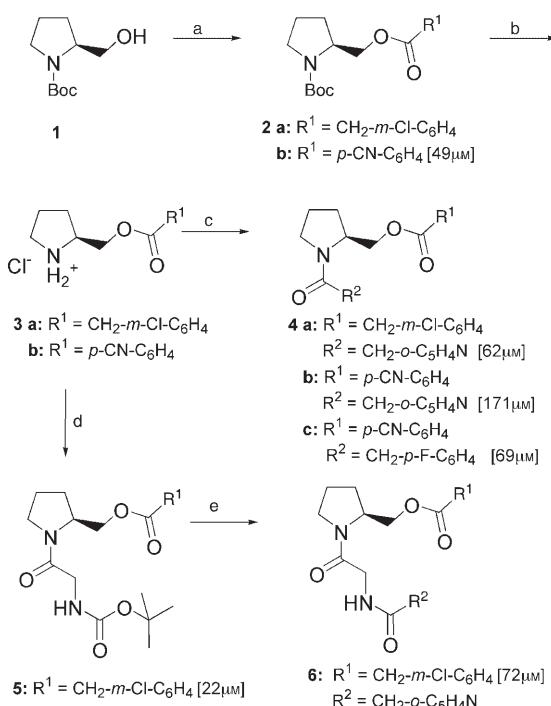
Schema 1. Strukturformeln der Ligandbausteine für die verschiedenen Subtaschen.



Schema 2. Zusammensetzung der virtuellen kombinatorischen Bibliothek mithilfe von FlexX^C.

bildung 2). Diese Struktur kann als Ausgangspunkt für eine weitere strukturbasierte Optimierung der neuen Leitstruktur dienen.

Thrombin ist eine etablierte Zielstruktur. Zieht man einen Vergleich zum bekannten Komplex mit Melagatran, einem potentiellen Thrombininhibitor, drängen sich die nächsten Optimierungsschritte förmlich auf: Zunächst erscheint die Platzierung des Ester-Sauerstoffatoms nicht ideal. Es hat keine Donorfunktion, um eine Wasserstoffbrücke zu der benachbarten Serinseitenkette einzugehen. Melagatran dagegen trägt in vergleichbarer Position ein Amidstickstoffatom für eine solche Wechselwirkung. Weiterhin fällt die Esterverknüpfung zwischen P1- und P2-Rest zu lang aus und sollte um



Schema 3. Synthese der Verbindungen 2–6. Die K_i -Werte für 2b, 4a, 4b, 4c, 5 und 6 sind in eckigen Klammern angegeben. Reagenzien und Bedingungen sind in den Hintergrundinformationen aufgeführt.

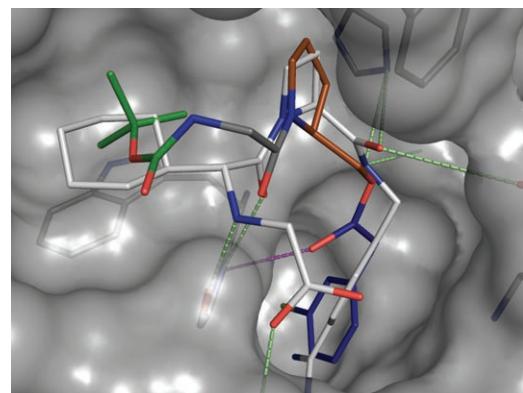


Abbildung 2. Überlagerung der Kristallstrukturen von 5 (P1: 3-Chlorbenzyl-Rest in blau; P2: Prolinol in braun; Glycinrest in dunkelgrau; P3: *tert*-Butylrest in grün) und Melagatran (PDB-Nr. 1k22). Melagatran bildet mit seiner P1-P2-Amidbindung Wasserstoffbrücken zu Ser195 und Glu192 in Thrombin aus (hellgrün). Weiterhin kommt es zu einer doppelten Wasserstoffbrücke zwischen Melagatran und Gly216. Im Fall von Verbindung 5 wird der Aufbau eines optimalen Wasserstoffbrückennetzwerks zwischen Protein und Ligand durch die Wechselwirkung zwischen dem Ester-Carbonylsauerstoffatom und Gly216 NH (violett) verhindert.

eine Einheit verkürzt werden. In Verbindung 5 drängt sich das Ester-Carbonylsauerstoffatom regelrecht in eine Wechselwirkung zur NH-Gruppe von Gly216 hinein. Dies hat zur Folge, dass die Wechselwirkung des eigentlich für diesen Zweck vorgesehenen Glycin-Fragments in 5 unter Bildung einer doppelten, parallelen Wasserstoffbrücke zum Protein-

rückgrat ausbleibt. Melagatran hingegen baut dieses typische Wasserstoffbrückennetzwerk auf. Daraus lassen sich die folgenden Syntheseschritte ableiten: Die viergliedrige Esterverknüpfung sollte durch eine dreigliedrige Amidverknüpfung ersetzt werden. Dies sollte dann auch das substratähnliche Bindungsmotiv über den Glycinbaustein zu der Aminosäure Gly216 des Thrombins ermöglichen.

In der vorliegenden Studie wurde ein neues wissenschaftliches Verfahren zum Design kombinatorischer Ligandbibliotheken überprüft. Das Vorgehen verknüpft die Auswahl geeigneter Bausteine mit Strukturinformationen aus anderen Protein-Ligand-Komplexen, die in ihren ähnlichen Subtaschen Ligandfragmente mit vergleichbaren physikochemischen Eigenschaften aufnehmen. Durch die Extraktion dieser Ligandbausteine und ihr kombinatorisches Zusammensetzen mit anschließendem Docking wird ein direktes Inhibitordesign in der Proteinbindetasche ermöglicht. Ein analoger Subtaschenvergleich mit Cavbase führte bereits im Fall der SARS-Protease zu Vorschlägen für mögliche Seitenkettenvariationen mit nichtnatürlichen Aminosäuren in peptidomimetischen Inhibitoren.^[12]

Computermethoden wird häufig vorgeworfen, sie würden beim Design erster Leitstrukturen viel zu komplizierte Moleküle vorschlagen, die nur mit hohem Aufwand zu synthetisieren sind. Mit dem hier vorgestellten Konzept wird versucht, gleich von Beginn an die Machbarkeit einer Synthese in das Design einzubringen. Computergestütztes Design und kombinatorisches Docking führen zur Selektion vielversprechender Kandidaten. Dabei baut die Synthesestrategie auf einfachen Umsetzungen auf und verwendet ausschließlich kommerziell erhältliche Ausgangsverbindungen.

Bei der weiteren Optimierung der so gefundenen Leitstrukturen sollte dann allerdings auf die metabolische Stabilität geachtet werden. Die Esterbindung lässt sich zwar einfach aufbauen, sie muss aber ersetzt werden, da ihre Spaltung durch Esterasen sicherlich sehr schnell zu einem Wirkungsverlust der Liganden führen würde. Das beschriebene Verfahren hat allerdings den Vorteil, dass das Design der Verbindungen vollständig auf Bindungsgeometrien beruht, die in experimentell bestimmten Kristallstrukturen nachgewiesen wurden.

Eingegangen am 24. Juli 2007
Online veröffentlicht am 23. Oktober 2007

Stichwörter: Kombinatorische Chemie · Ligandendesign · Thrombin · Wirkstoffdesign

- [1] E. J. Martin, R. E. Critchlow, *J. Comb. Chem.* **1999**, *1*, 32.
- [2] A. R. Leach, M. M. Hann, *Drug Discovery Today* **2000**, *5*, 326.
- [3] C. G. Wermuth, *Drug Discovery Today* **2006**, *11*, 160.
- [4] G. Müller, *Drug Discovery Today* **2003**, *8*, 681.
- [5] H. J. Böhm, D. W. Banner, L. Weber, *J. Comput.-Aided Mol. Des.* **1999**, *13*, 51.
- [6] K. Illgen, T. Enderle, C. Broger, L. Weber, *Chem. Biol.* **2000**, *7*, 433.
- [7] G. Schneider, M. L. Lee, M. Stahl, P. Schneider, *J. Comput.-Aided Mol. Des.* **2000**, *14*, 487.
- [8] L. Weber, S. Wallbaum, C. Broger, K. Gubernator, *Angew. Chem.* **1995**, *107*, 2452; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2280.
- [9] F. Dullweber, M. T. Stubbs, D. Musil, J. Stürzebecher, G. Klebe, *J. Mol. Biol.* **2001**, *313*, 593.
- [10] S. Schmitt, D. Kuhn, G. Klebe, *J. Mol. Biol.* **2002**, *323*, 387.
- [11] N. Weskamp, D. Kuhn, E. Hüllermeier, G. Klebe, *Bioinformatics* **2004**, *20*, 1522.
- [12] S. I. Al-Gharabli, S. T. A. Shah, S. Weik, M. F. Schmidt, J. R. Mestres, D. Kuhn, G. Klebe, R. Hilgenfeld, J. Rademann, *ChemBioChem* **2006**, *7*, 1048.
- [13] D. Kuhn, N. Weskamp, S. Schmitt, E. Hüllermeier, G. Klebe, *J. Mol. Biol.* **2006**, *359*, 1023.
- [14] M. Hendlich, A. Bergner, J. Günther, G. Klebe, *J. Mol. Biol.* **2003**, *326*, 607.
- [15] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.
- [16] P. Ertl, Novartis Institutes for BioMedical Research Basel, Schweiz, **2004**.
- [17] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, *J. Mol. Biol.* **1996**, *261*, 470.
- [18] M. Rarey, T. Lengauer, *Perspect. Drug Discovery Des.* **2000**, *20*, 63.
- [19] H. F. Velec, H. Gohlke, G. Klebe, *J. Med. Chem.* **2005**, *48*, 6296.
- [20] P. Block, N. Weskamp, A. Wolf, G. Klebe, *Proteins Struct. Funct. Genet.* **2007**, *68*, 170.
- [21] J. Stürzebecher, U. Stürzebecher, H. Vieweg, G. Wagner, J. Hauptmann, F. Markwardt, *Thromb. Res.* **1989**, *54*, 245.