Combinatorial receptor finding—large and random *vs.* small and focused libraries

Carsten Schmuck* and Peter Wich

Received (in Montpellier, France) 18th April 2006, Accepted 10th July 2006
First published as an Advance Article on the web 23rd August 2006
DOI: 10.1039/b605731b

Whereas in the classical combinatorial approach mainly large but completely random libraries are used, nowadays also the use of small but designed libraries is coming into focus. We discuss the pros and cons of these two different approaches, using examples from literature work and our own studies showing how combinatorial libraries are applied in supramolecular chemistry *e.g.* to identify artificial receptors for peptide binding in aqueous solvent.

1. Introduction

The use of combinatorial chemistry has fundamentally changed the pace and scope of scientific research in some areas. The introduction of synthetic peptide libraries has proven that combinatorial chemistry is a powerful tool for the generation of libraries with immense molecular diversity. But the hype as a new tool mainly in the pharmaceutical industry at the beginning of the 1990s has slowed down in recent years. The original hope that the screening of large libraries that contain millions of compounds would produce many new drug candidates has not been fulfilled with complete satisfaction. Nevertheless, combinatorial chemistry has established itself as a powerful tool—among others—in chemistry even though it is not the magic bullet initially anticipated by some. Based on the initial failures, mainly in pharmaceutical research, combinatorial chemistry is currently changing once more. Besides the large but random libraries initially employed also the use of small but focused libraries are increasingly used to address specific problems in various fields of research. This article will describe first some general aspects of combinatorial chemistry and then discuss the pros and cons of these two approaches, the use of large and random *vs.* small and focused libraries. We will mainly concentrate on examples from our own area of research, the finding of artificial peptide receptors with the help of combinatorial libraries.

Carsten Schmuck is Professor of Organic Chemistry at the University of Würzburg. He studied chemistry and obtained his PhD at the Ruhr-University in Bochum with a work in physical organic chemistry. After a postdoctorate at Columbia University in New York with Ron Breslow, he started his independent academic career at the University of Cologne where he finished his Habilitation in 2001 before moving to his current position in Würzburg in 2002. His main reseach interests are supramolecular and bioorganic chemistry with a special emphasis on molecular recognition events in aqueous solvents. His group designs new artificial hosts for amino acids, oligopeptides or carbohydrates using both rational approaches as well as combinatorial chemistry. The Schmuck group is also working in the field of organocatalysis, self-assembling materials based on zwitterions, and the development of sensor materials.

Prof. Dr. Carsten Schmuck, Dipl.-Chem. Peter Wich, Institut für Organische Chemie, Universität Würzburg, Am Hubland, 97074 Würzburg, Germany. E-mail: schmuck@chemie.uni-wuerzburg.de; Fax: +49 931 8884625; Tel: +49 931 8885326

Combinatorial chemistry

The general concept of combinatorial libraries, as first developed for peptides,[1] involves the generation of all possible sequence permutations for a peptide of a given length in connection with a subsequent screening and selection process that enables the identification of unique highly "active" peptides in the presence of "inactive" peptides, *i.e.* in terms of binding activity to a certain target. However, the field of applications for combinatorial chemistry has expanded since then to include proteins,[2] synthetic oligomers,[3] small molecules[4] or oligosaccharides.[5]

Split and mix approach

The majority of immobilized libraries reported use a resin as solid support and the *split and mix* method,[6] generating *one-*

Peter Wich studied chemistry in Würzburg and obtained his diploma in 2005 working under the guidance of Prof. Schmuck studying the selectivity of artificial receptors for biologically relevant peptide targets using combinatorial methods. He is currently working on his dissertation.

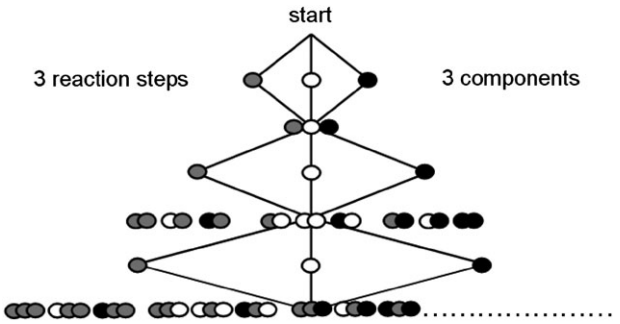


Fig. 1 Schematic mechanism of the split and mix approach. Example with three steps and three components resulting in $3 \times 3 \times 3 = 27$ different product combinations.

bead–one-compound libraries.[7] Combinatorial libraries, prepared *via* this approach contain only one single library member, which in the case of a peptide library corresponds to one sequence of amino acids on a single 100 μm diameter bead. The result of such a split synthesis is a collection of beads each of which contain one specific peptide sequence consisting of every possible combination of every amino acid used in the synthesis. Fig. 1 illustrates the fast and easy synthesis of a combinatorial library with 27 different products from three different building blocks in only three steps requiring only nine individual reactions. One problem of this approach is however that it is only applicable to the synthesis of sequenceable oligomers.

These libraries are then typically screened in a solid-phase assay, which is based on the selection of positively reacting resin beads, followed by analysis of the compounds attached to the "active" beads. In this way, active library members can be qualitatively identified, for example by some colour change (*vide infra*). However, the main problem of such one-bead–one-compound libraries is the low loading of the resin. One bead carries only approx. 100 pmol ($\sim 10^{13}$ molecules) of each library member. While this quantity is adequate for modern Edman sequencing of small peptides, it is generally too small for any other spectroscopic technique or analytical method. Hence, the identity of the active library members as seen in those screening assays is difficult to establish. This is especially a problem for libraries of non-peptides which can not be analyzed by Edman degradation. Therefore, unique chemical tags to encode the structure of the library compound on each resin bead were developed,[8] which can then be decoded by sequencing or some other analysis, for example, gas chromatography or high performance liquid chromatography (HPLC). In principle, almost any type of molecule can be used as a tag. However, there are practical limitations because tags need to be chemically inert and reliably analyzed on femtomolar scales from a single bead. But even then, the problem remains, that the screening of the library and the subsequent decoding is tedious and challenging and in most cases provides only indirect answers on the identity of the library member under study.

Dynamic combinatorial libraries

This traditional combinatorial chemistry approach is based on large libraries of prefabricated molecules synthesized by the essentially irreversible building and breaking of covalent bonds. Hence, the composition of the library is determined after its synthesis and before any screening or further study is performed. In contrast to this covalent approach, dynamic combinatorial chemistry relies on the reversible connection of building blocks to give access to libraries whose composition is not yet fixed but can change in response to its surroundings. If bond formation is reversible, the library can rearrange, governed by the thermodynamics of the whole molecular ensemble.[9] For example, if a target molecule is added to such a dynamic library, the composition of the library will re-equilibrate until the new thermodynamic minimum is reached. This can lead to an amplification of host molecules with high affinity to the added target molecule. Hence, the presence of a guest molecule induces the formation of an appropriate host molecule within the library.

However, there are several limitations such as the need for sufficient solubility of every single library member, the lack of suitable reversible chemistry and the low level of reaction control. For example, nowadays most dynamic libraries are based on reversible chemistry such as imine or disulfide formation. However, this of course limits the scope of molecules that can be used to build a dynamic combinatorial library. Furthermore, it was recently shown with both theroretical simulations[10] as well as in experiments,[11] that it is not necessarily the host with the highest affinity to a given target that is amplified the most within the library. Statistical reasons can lead to the amplification of hosts with lower affinity. Nevertheless, even though the research in the field of dynamic combinatorial chemistry is still in an early stage, this technique makes a promising addition to the set of combinatorial methods. The next few years and the advent of more versatile reversible chemistry will surely reveal the full applicability of this approach.

2. Combinatorial libraries in supramolecular chemistry

The original purpose of combinatorial chemistry has evolved in recent years into broad fields of applications as diverse as material science,[12] catalyst development,[13] and biochemistry to identify the substrates of novel enzymes.[14] It opens also the way in the widespread area of supramolecular chemistry. We use combinatorial methods in this context to find new receptors that are capable to bind to a given target peptide even in aqueous solvents. This can help us to increase our knowledge of molecular recognition in general and help us to design biosensors for the targeting of cellular processes or for the discovery of new therapeutics. But how should a peptide receptor look? In principle there are two distinct routes one can follow.[15] One can try to rationally design a complete receptor *de novo* with the help of theoretical calculations.[16] The larger the substrate is, the more difficult however this becomes, as theoretical calculations are not yet reliable enough to completely design a tailor-made artificial host for a large substrate. Another possibility is to use a random trial and error approach and to identify suitable receptors with the help of combinatorial chemistry using large and random libraries.[17] Most combinatorial studies in this area involved the synthesis

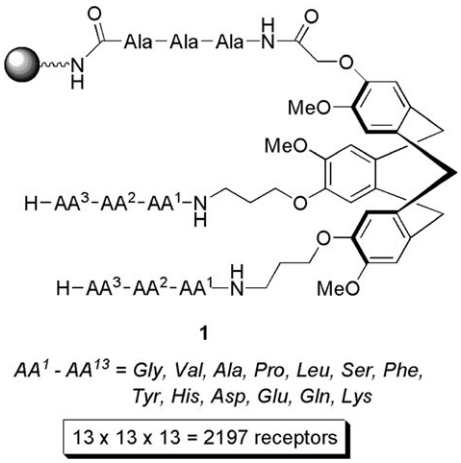
of a large peptide library that is then screened against a given receptor in order to get at a better understanding of host–guest interactions. The alternative strategy, the preparation of combinatorial libraries of host-like molecules, followed by identification of those receptors that bind a given substrate selectively has been much less thoroughly investigated. A few examples using this traditional combinatorial approach will be described in the next section. The best method of course will be to combine both approaches and to use small libraries with tailor-made building blocks specifically designed to bind a peptide target. This much less explored approach will be described later on.

3. Large random libraries

Combinatorial chemistry is a very powerful tool in the effort to find small molecules that alter protein function or lead to new drugs. Proteins or antibodies, as biological macromolecules, show a remarkable capacity for the specific binding of peptides or other organic substrates. However, also rather small natural products such as the antibiotic Vancomycin demonstrate that a huge macromolecular size is not necessary a requirement for efficient supramolecular interactions. Vancomycin recognizes the bacterial dipeptide sequence D-Ala-D-Ala very efficiently ($K_{ass} \sim 10^6$ M^{-1}).[18] The reasons for this high specificity have been the subject of many studies.[19] Unfortunately, Vancomycin-resistant bacterial strains emerged in the last ten years caused by a substitution of the C-terminal D-alanine by D-lactate, resulting in a 1000-fold decrease in affinity and therefore ineffective antibiotic activity of Vancomycin. For the identification of an artificial receptor able to recognize these peptide/depsipeptide sequences various combinatorial libraries as well as otherwise developed receptors have been employed with more or less success.[20]

Chamorro and Liskamp reported the screening of a large combinatorial receptor library derived from a cyclotriveratrylene with three attached peptide arms for the binding of D-Ala-D-Ala and D-Ala-D-Lac.[21] Using a colour-coded substrate, a 2197-member library (**1**) of CTV-based synthetic tripodal receptors was screened in water with phosphate buffer (0.1 N, pH = 7.0) (Fig. 2). Efficient receptors could be identified qualitatively by identification of active beads and subjecting them to Edman degradation. The best receptors could bind the dipeptide more efficiently than the related depsipeptide D-Ala-D-Lac. However, no quantitative information on binding affinities or substrate selectivity was provided, due to the lack of sufficient substrate material, as a result of the one-bead one-compound strategy.

Kilburn and co-workers just recently reported on the screening of a large combinatorial library of tweezer-receptors (>15 000 members) for the binding of the tripeptide *N*-Ac-L-Lys-D-Ala-D-Ala.[22] A guanidinium scaffold was incorporated as a specific recognition site for carboxylate functionality into the receptor structure to identify receptors for peptides with free carboxylate groups. In earlier work, the libraries prepared had only limited diversity, because both arms of the receptor were synthesized simultaneously.[23] Here, Kilburn and co-workers developed a route to unsymmetrical tweezer receptors (**2**) using two different strategies with a careful use of orthogonal protecting groups. The first strategy was the successive preparation of each peptide arm by the split and mix synthesis of Boc- and Fmoc-protected amino acids, respectively (see Fig. 3). The choice of amino acids in each arm was restricted to avoid any ambiguity, as to which arm the respective amino acid derived from, during the identifying process of the final Edman sequencing when two amino acids, one of each arm, are cleaved at the same time. The amino acid used in the first position (AA^1) of the first arm must differ from that of the first position (AA^4) of the second arm. The same holds for the second and third position (see Fig. 3).



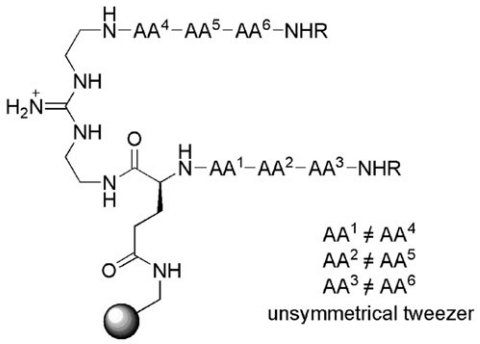
1

AA^1 - AA^{13} = Gly, Val, Ala, Pro, Leu, Ser, Phe, Tyr, His, Asp, Glu, Gln, Lys

13 x 13 x 13 = 2197 receptors

Fig. 2 Combinatorial library of 2197 different artificial tripodal receptors based on a cyclotriveratrylene (CTV) scaffold. Screening and Edman sequencing revealed peptide sequences which are either capable to bind dye labelled D-Ala-D-Ala or D-Ala-D-Lac containing ligands.

An alternative approach was the attachment of a separate coding strand on the beads composed of the same amino acids for both arms (see Fig. 4). In addition phenylalanine was introduced in the coding strand as a useful check that the



2

AA^1 - AA^3 = Gly, L-Leu, L-Phe, L-Glu(OtBu), L-Pro
AA^4 - AA^6 = L-Ala, L-Val, L-Gln, L-Ser(OtBu), L-Lys(Boc)

5 x 5 x 5 x 5 x 5 x 5 = 15625 receptors

Fig. 3 Two-armed tweezer receptor library with a guanidinium head group as a carboxylate binding site for the binding of D-Ala-D-Ala–OH in aqueous solution. The arms are synthesized sequentially to give a structurally more diverse library of unsymmetrical receptors.

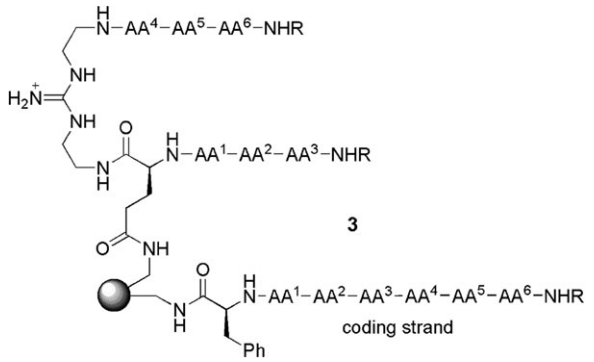


Fig. 4 An additional heptapeptide coding strand allows the identification of the receptor structure *via* Edman degradation.

library synthesis and the sequencing chemistry has worked efficiently. Its only function is the confirmation of the accurate peptide synthesis and the following analysis of the peptide composition. For every bead the last amino acid during the Edman degradation must then be phenylalanine. If not, this indicates a problem in the synthesis of one of the two peptide arms.

These large libraries were screened in aqueous buffer (pH 8.5, borate) with the dye labelled tripeptide *N*-Ac-Lys-D-Ala-D-Ala, using a qualitative binding assay by the observation of stained beads, visualized under a microscope. Some good binding receptors were identified and the most promising one was then resynthesized and studied in more detail. Weak binding with low mM affinities was found for two diastereomeric tripeptides with the resin-bound receptor but no binding data in free solution could be obtained.[22] This shows again the limits of large libraries. Positive hits can only be selected qualitatively and have to be resynthesized on a larger scale for further analysis.

Wennemers and co-workers adapted the method of the one-bead–one-compound library concept for the discovery of new catalysts (Fig. 5). This involves the coimmobilization of one reaction partner **A** with each library member, the potential catalysts, on the same bead.[13c] The second reaction partner **B** in solution is labelled with a dye or a fluorophor. A reaction between **A** and **B** catalysed by the library member on the same bead leads to a covalent connection of the marker with the beads of the active library members. These beads can than be easily selected under a microscope. Again the design of the library allows only discrimination between active and non-active members. There is no information about why and how a

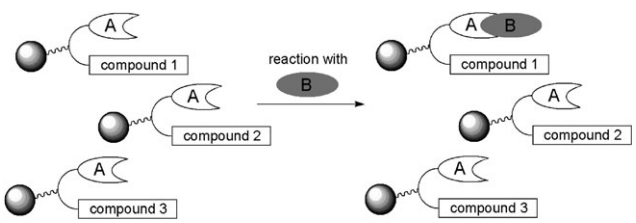


Fig. 5 One-bead–one-compound libraries as a tool for the discovery of active catalysts. In this example only compound **1** is active. This concept allows the qualitative screening of a large number of potential candidates in combinatorial split-and-mix libraries.

compound is acting in the process of catalysis. Such information requires further studies with separately synthesized catalysts and their individual characterization.[24]

These selected examples demonstrate the high potential of combinatorial libraries in the fields of supramolecular chemistry and catalyst development. However, they also show the main weakness of this traditional approach. The library screening provides only a qualitative answer in the form of a "yes" or "no". Hits can be identified and can then be further analyzed, but nothing is known about the non-hits. There is no information of how good or how bad a hit or a non-hit actually is. In other words, a lot of information is lost in the pure qualitative screening of a combinatorial library. However, the low loading of the resins and the large number of library members prevents any further investigations. This is where small but focused libraries come into play.

4. Small focused libraries

Contrary to expectations scientists had some years ago, that large libraries would lead to the discovery of many hits and lead structures, the results of many studies pointed out that biological relevance, design and diversity of the library are more important. In recent years, the concept of target-oriented synthesis (TOS) is of increasing interest.[25] This concept is used in solid-phase syntheses aimed *e.g.* at drug discovery, in particular in the syntheses of focused libraries, where collections of compounds with common structural features that facilitate binding to a preselected target are synthesized. For example, Waldmann and co-workers recently suggested that the use of a biologically validated starting point for combinatorial libraries would significantly improve the hit-rate. Natural products often embody privileged structures that can also evolve into binding to other proteins which are not their initial targets and therefore may result in new lead compounds with enhanced quality.[26] However, it remains questionable whether the regions of chemical space defined by natural products and known drugs are really the best or most fertile regions for the discovery of active compounds? This question led to the concept of diversity-oriented synthesis (DOS) that included the development of pathways leading to the efficient (3–5 step) synthesis of collections of small molecules having rich skeletal and stereochemical diversity with defined coordinates in chemical space also mainly derived from known natural products. The systematic screening of this collection of compounds should advance the fundamental understanding of the roles these diversity elements play in small molecule/protein interactions.

In a *focused* combinatorial library the chances to find a hit are much higher than in a complete random library as the structural diversity is already positively biased for a given problem, *e.g.* binding to a specific target protein. Hence, it is sufficient to use much smaller libraries with only a couple of hundred of different members. This concept of small but focused libraries has already successfully been applied in pharmaceutical and medicinal chemistry.[26] We were interested to use the same approach of small but focused libraries also in the field of supramolecular chemistry for the discovery of artificial peptide receptors.

Spatially separated library members

As already described above, the best way to find an artificial peptide receptor is most likely to combine a rational design approach with the power of combinatorial chemistry. Using such a combined approach, we recently described the screening of a medium-sized combinatorial library of fully flexible one-armed cationic peptide receptors (**4**) for the binding of tetrapeptides in water. The receptors of the library are composed of a carboxylate binding site (CBS) attached to a variable tripeptide unit (see Fig. 6). To ensure strong complexation in polar solvents even for such a short β-sheet, the carboxylate binding site was introduced in the form of a cationic guanidiniocarbonyl pyrrole group. As we can show, this is among the most efficient binding motifs for carboxylates known so far, even in aqueous solutions.[27] A combinatorial variation of the three amino acids in the receptor side chain can then be used to identify receptors in which additional electrostatic and steric interactions between these side chains and the substrates further enhance the binding within the complex and also render the recognition event selective for a specific tetrapeptide. As one of the first target peptides we chose the two tetrapeptides EKAA (**5**) and AAKE (**6**).[28] As mentioned above, the EKAA-tetrapeptide sequence (D-Glu-L-Lys-D-Ala-D-Ala–OH) is interesting in terms of its relevance to bacterial cell wall maturation upon treatment with Vancomycin, leading to bacteria death.

The receptor library was synthesized on Amino-TentaGel® as the solid support according to a standard Fmoc-protocol using the split-mix approach (*vide supra*) in combination with the IRORI™-radio frequency tagging technology.[29] The IRORI microreactors (Fig. 7) are miniaturized devices that contain both a functionalized solid support and a unique tag identifier. The radio frequency tagging technology operates with a radio frequency chip fused in a glass mantle. A unique binary code on every chip allows the read out with a scanning station. Thus, each reactor kan is "tagged" with a unique ID which makes it possible to assign the synthesized product in every step of the preparation.

The kans are designed to be loaded with up to 30 mg of resin beads. This allows the spatially separated synthesis of micromole levels of each library member (compared to only picomole levels using the classical one-bead–one-compound approach). This is enough material of each library member for advanced qualitative and quantitative experiments on

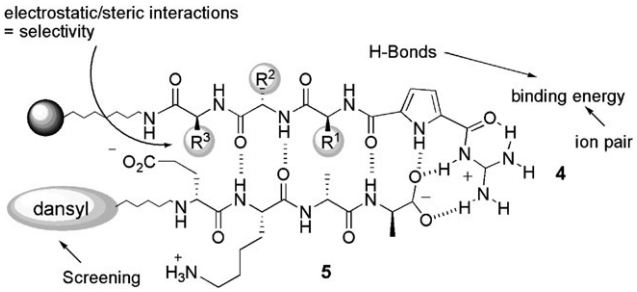


Fig. 6 Schematic representation of complex formation between the receptor library **4** and the dansylated tetrapeptide substrate **5**; dansyl = dimethylaminonaphthalene-1-sulfonyl.

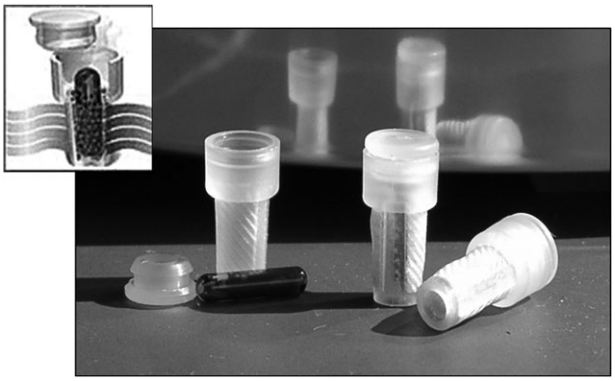


Fig. 7 IRORI MikroKan systems with a loading up to 30 mg resin per reactor. A radio frequency chip inside allows the identification of the resin bound peptide sequence.

binding activity. The synthesis takes place as reagents flow through the outer mesh walls of the microreactors using normal laboratory glassware. A further advantage of this system is that the number of necessary synthesis steps can be kept very low, compared to bigger libraries requiring an additional chemical tagging in every step.[30] Only the actual building blocks of the receptor have to be attached to the resin and no further tagging reaction is required. There is no need of any coding strands or chemical markers. However, the maximal number of library members is limited in a normal chemistry lab to approx. 1000, compared to libraries with several ten thousand members or more using a chemical tagging, due to higher demands on time and costs. During the synthesis of the receptor library **4** the following eight different amino acids were used in each of the three coupling steps: Lys, Tyr, Ser, Glu, Phe, Val, Leu and Trp giving rise to a library with 512 different members. These specific amino acids used were carefully chosen among the proteinogenic amino acids to provide a representative range of varying polar, charged and hydrophobic residues within the final receptor library.

The advantage of such a small and focused solid-phase bound combinatorial receptor library is, besides the fast and time saving synthesis, that the whole library can be tested for a specific feature, in this case its binding properties towards the tetrapeptide substrate, in a single experiment (Fig. 8). For this purpose a fluorescence label in form of a dansyl group was attached *via* a water-soluble spacer to the N-terminus of the tetrapeptide substrate. To probe the entire receptor library qualitatively for its binding properties aliquots of the 512 resin bound deprotected receptors **4** were pooled and the combined mixture incubated with a 5 μm solution of the tetrapeptide substrate in 20 μm bis-tris buffer of pH = 6.0 in water. After the supernatant solution was washed off, the beads were screened under UV light using a fluorescence microscope. A selective binding of the tetrapeptide substrate by some—but not all—of the 512 receptors **4** can be observed as indicated by the strong fluorescence activity of individual beads. Only those beads, on which the attached receptor is capable to bind the peptide, can show the characteristic fluorescence of the dansyl group. All the other receptors which do not bind the peptide under the specific experimental conditions remain dark (Fig. 9).

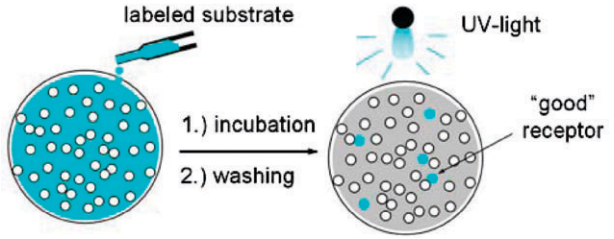


Fig. 8 On-bead assay for the qualitative identification of "active" candidates within the receptor library.

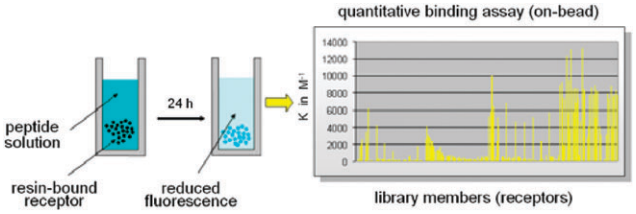


Fig. 10 Determination of binding constants on-bead in a quantitative screening of all library members.

Based on this qualitative screening it is possible to determine binding constants of all library members with the help of the IRORI technique in an on-bead assay using a high throughput microtiter plate reader (Fig. 10).[31] From the fluorescence intensity of the substrate in solution, before and after incubation and the loading of the resin, the association constants for each receptor can be calculated. The binding affinities within the library for tetrapeptide **5** (EKAA) vary from K_{ass} = 17 000 M^{-1} for the best (CBS-KKF) to <20 M^{-1} in buffered water for the worst receptors. This represents a difference in activity of more than two orders of magnitude! For such a structurally closely related library of necessarily limited diversity this represents a remarkably selectivity. The binding of **6** (AAKE) is somewhat less efficient (K_{ass} = 6000 M^{-1} for the most efficient receptors). In all experiments the binding data obtained from the solid-phase screening, could be validated by complexation studies of resynthesized receptors in free solution with NMR and UV titration experiments. Hence, much more information is obtained for the library members compared to the traditional approach. Not only a yes/no answer is provided but also real quantitative data which can be further analyzed and interpreted in detail.

Even though such binding constants determined on a solid support are not the same and in general are less accurate than data obtained in solution, a comparison of relative data within a series of related receptors can at least help rationalize aspects such as complex structure, stability and selectivity on a molecular basis. One can identify structural features that are associated with strong or weak binding. Which parts of our modular receptors are most important for binding or selectivity? What kind of binding sites, electrostatic or hydrophobic, in the various positions of the receptor are needed? In other words a supramolecular structure–binding relationship can be derived from binding data obtained on a solid support. This is not possible from the pure qualitative screening of large libraries as usually performed.

A good example is the formation of a β-sheet like complex between the tetrapeptide Ac-Val-Val-Ile-Ala–OH and the receptor Gua-Lys(Boc)-Ser(OtBu)-Phe, as recently reported by us.[32] This tetrapeptide represents the C-terminal part of the amyloid-β-peptide (Aβ) responsible for plaque formation in Alzheimer's disease. Especially, hydrophobic interactions to this tetrapeptide sequence are thought to promote the self-aggregation of the 42 amino-acid long Aβ. Our quantitative screening of a combinatorial receptor library showed indeed a high preference for a hydrophobic Lys(Boc) in a certain position enabling hydrophobic interactions with the first (Val) and third (Ile) amino acids of the tetrapeptide. The tBoc group is placed between these two side chains of the substrate thereby closing the hydrophobic gap in between. This minimizes the solvent accessible surface (see Fig. 11) and therefore stabilizes the complex. A similar hydrophobic interaction has been predicted, based on molecular mechanics calculations, to play a significant role in the native system.

This work shows the high potential of small and focused libraries in combinatorial chemistry *i.e.* for identifying structural features responsible for high binding activity. Without the quantitative binding data also of the poor receptors this analysis would have not been possible. This kind of structure–activity correlation is not obtainable with the use of large

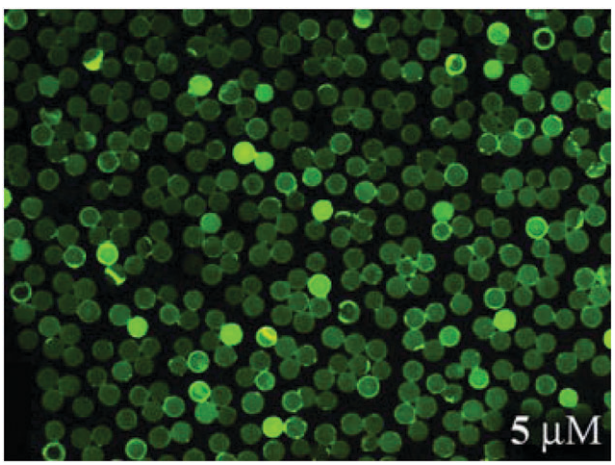


Fig. 9 Incubation of the solid-phase bound receptor library with a fluorescence-labelled substrate reveals a selective interaction with only some substrates within the library.

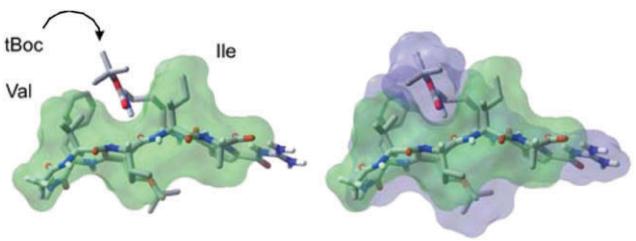


Fig. 11 The tBoc group of the lysine side chain of the receptor fits perfectly into the hydrophobic gap between the first (Val) and third (Ile) side chain of the substrate, causing high binding affinity.

one-bead–one-compound libraries where you get only a simple yes or no answer.

QSAR-analysis

However, when using such small libraries of only a couple of hundred members the question remains, if the size of the initial library is sufficient to provide significant results. In other words to really provide the best possible receptor in this case, respectively. Or are the results suboptimal due to the fact that the library simply did not contain the correct diversity for the given problem? We could answer this question by corroborating the results obtained for the binding affinities of our flexible receptors for the tetrapeptide EKAA by an additional statistical analysis.[33] Quantitative structure–activity relationships (QSARs) correlate properties of the chemical structures under scrutiny to their activity with a mathematical model. The latter can be used to detect patterns and trends in the data. Moreover, it can be used to make predictions for compounds not yet synthesized. We fitted a model based on 49 diverse physical–chemical descriptors for each amino acid in each of the three variable positions in the receptor to the quantitative binding data obtained from the library screening. A plot of the experimental log(K_{ass}) values *vs.* the cross-validated predictions of log(K_{ass}) is shown in Fig. 12 underlining the good quality of the fit. The mathematical model thus obtained was then used to predict the binding affinities of all possible tripeptide sequences in such a receptor which can be obtained by permutation from the 20 proteinogenic amino acids (library size $n = 8000$ members). This virtual library is 15 times larger than the initial experimental library of 512 receptors that we used for our screening. Out of the 7488 virtual receptors not yet synthesized, there were only 16 receptors that were predicted to better bind to the tetrapeptide EKAA than the most active receptors in our small library. However, the binding affinities are all in the same range and are not significantly larger than those already synthesized and analyzed. Therefore, the initial library size was completely sufficient to fully explore the chemical space for this given problem, the binding to this specific peptide target. An actual combinatorial synthesis of this larger 8000 member library would have not paid off but rather would have required an unnecessary use of time and money. Hence, if the library is correctly designed for the question under study, small but focused libraries can provide the same results as much larger but random libraries. In contrast to the large libraries the smaller libraries have the advantage that a full quantitative analysis of all library members is possible. This provides an additional wealth of information otherwise not obtainable.

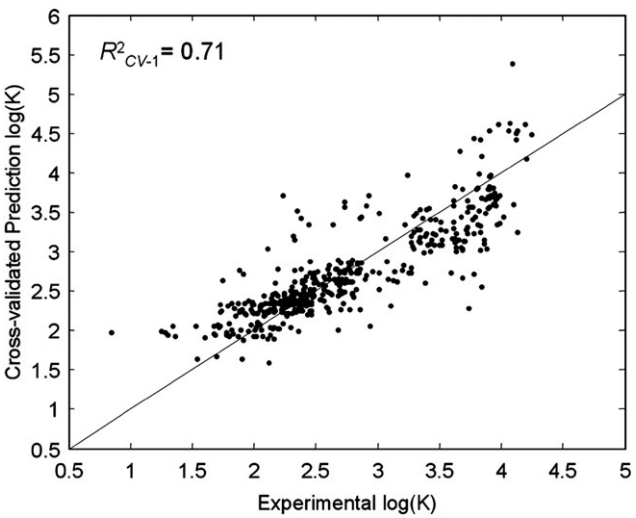


Fig. 12 Experimental *vs.* predicted log(K_{ass}).

Insights into structure and reactivity

A good example for the additional benefit that the information from the quantitative analysis of a whole library can provide is demonstrated in the follow-up work on an artificial peptide receptor showing not only significant substrate selectivity but also a remarkable sequence dependent stereoselectivity. An artificial receptor (CBS-KKF), which in other experiments showed high affinity to the peptidoglycan model peptide D-Glu-L-Lys-D-Ala-D-Ala–OH (*vide supra*) was further examined in a fluorescence screening against a combinatorial substrate library of 320 closely related tetrapeptides to get more information about its binding activity.[34] The substrates presented only three different side chains (Ala, Lys and Glu) and differed only in the absolute configuration of one building block (D/L-Ala) or one chemical linkage (Ala *vs.* Lac), respectively. The screening showed binding constants from <50 M^{-1} up to 27 000 M^{-1} for the best substrate. These are already considerable differences for a library of such a moderate size. The quantitative evaluation resulted in a preference for substrates with anionic amino acids as expected for a tris-cationic receptor. What was not expected and would have been simply overlooked in a traditional combinatorial approach using the simple qualitative screening of a large library was the sequence-dependent stereoselectivity of the receptor. The receptor showed both a distinctive stereoselectivity between D- und L-alanine as well as a remarkable selectivity between D-Ala and D-Lac, *but only at certain positions within the complex*. This could be explained based on the conformational flexibility of the complex. Stereoselectivity requires a rather well defined complex structure and is only possible, when the position where the D-Ala/L-Ala exchange takes place is fixed *at both sides* by strong charge interactions between receptor and substrate. No stereoselectivity is observed, for example when the alanine is in the N-terminal position (see Fig. 13).

5. Perspective

The results described above underline the potential of fully analyzable small libraries compared to large libraries resulting only in qualitative "yes or no" hits. Hence, this alternative approach allows for a fast and economic way to identify *e.g.* potent receptors for a given target by using small but carefully composed combinatorial libraries. The pros and cons of both approaches, as discussed in the text, are once more summarized in Table 1. Small but carefully designed libraries are sufficient to explore the features of even larger ensembles. Hence, the pure size of a combinatorial library is not decisive for the outcome of the screening as long as the library contains the correct range of diversity for the given problem. Of course,

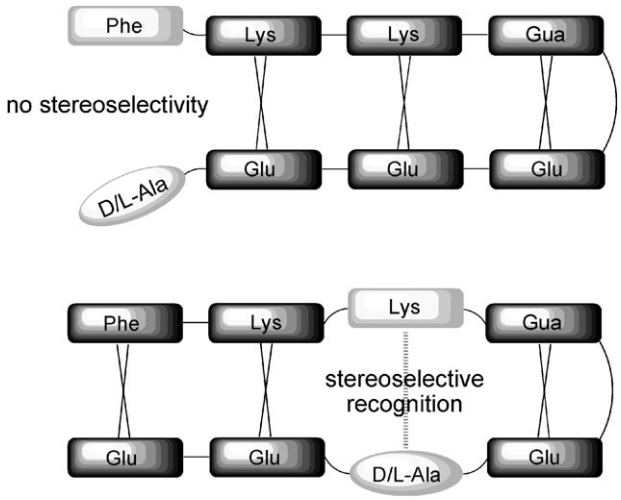


Fig. 13 Screening of a small focused library revealed a remarkable stereoselectivity between D-Ala and L-Ala but only in certain sequences which allow to form overall well defined complexes. This unexpected stereoselectivity is only observed, when the position where the D-Ala/L-Ala exchange takes place is fixed *at both sides* by strong electrostatic interactions between receptor (top) and substrate (bottom).

Table 1 Comparison of small but focused *vs.* large but random libraries

Small but focused libraries	Large but random libraries
Only limited number of library members	Large number of different library members possible
⇒ Only limited diversity	⇒ Much larger diversity
Locally separated library synthesis possible	Only traditional split–mix-synthesis
⇒ No extra tagging required	⇒ Additional tagging necessary (marker, coding strand)
Micromol (10^{-6}) level of each library member available	Only picomol (10^{-12}) level of each library member available
⇒ Qualitative and quantitative screenings possible	⇒ Only qualitative "yes or no" information can be obtained
⇒ Direct comparison of all library members possible	⇒ Selected hits have to be resynthesized for further analysis

the limited structural diversity of a small library requires a careful design to provide this correct diversity needed to answer a certain question. If composed correctly, a small but focused library can be as informative as a large but random library. It is not the pure size but the diversity that is important. However, this requires that libraries are much more thoroughly designed than in a purely random approach. So far this approach might therefore be limited to only certain fields of research which are rather well understood on a molecular basis. However, with an ever increasing and better molecular understanding of recognition phenomena and non-covalent interactions in general, this task to design a suitable library will hopefully become less challenging in the future.

References

1 (*a*) K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski and R. J. Knapp, *Nature*, 1991, **354**, 82–84; (*b*) J. K. Scott and G. P. Smith, *Science*, 1990, **249**, 386–390.

2 (*a*) S. J. Shuttleworth, R. V Connors, J. Fu, J. Liu, M. E. Lizarzaburu, W. Qiu, R. Sharma, M. Wanska and A. J. Zhang, *Curr. Med. Chem.*, 2005, **12**, 1239–1281; (*b*) D. J. Maly, L. Huang and J. A. Ellman, *ChemBioChem*, 2002, **1**, 16–37.

3 C. Y. Cho, E. J. Moran, S. R. Cherry, J. C. Stephans, S. P. A. Fodor, C. L. Adams, A. Sundaram, J. W. Jacobs and P. G. Schultz, *Science*, 1993, **261**, 1303–1305.

4 (*a*) M. Uttamchandani, D. P. Walsh, S. Q. Yao and Y.-T. Chang, *Curr. Opin. Chem. Biol.*, 2005, **9**, 4–13; (*b*) L. A. Thompson and J. A. Ellman, *Chem. Rev.*, 1996, **96**, 555–600.

5 P. M. St. Hilaire and M. Meldal, *Angew. Chem., Int. Ed.*, 2000, **39**, 1162–1179.

6 (*a*) A. Furka, F. Sebestyen, M. Asgedom and G. Dibo, *Int. J. Pept. Protein Res.*, 1991, **37**, 487–493; (*b*) R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo, *Nature*, 1991, **354**, 84–86.

7 K. S. Lam, M. Lebl and V. Krchnak, *Chem. Rev.*, 1997, **97**, 411–448.

8 A. Borchardt and W. C. Still, *Acc. Chem. Res.*, 1996, **29**, 155–163.

9 For reviews on the work of dynamic combinatorial chemistry see: (*a*) J. M. Lehn, *Chem. Eur. J.*, 1999, **5**, 2455–2463; (*b*) S. Otto, R. L. E. Furlan and J. K. M. Sanders, *Drug Discovery Today*, 2002, **7**, 117–125; (*c*) J.-L. Reymond, *Angew. Chem., Int. Ed.*, 2004, **43**, 5577–5579; (*d*) S. Otto, *Curr. Opin. Drug Discovery Dev.*, 2003, **6**, 509–520.

10 (*a*) Z. Grote, R. Scopelliti and K. Severin, *Angew. Chem., Int. Ed.*, 2003, **42**, 3821–3825; (*b*) K. Severin, *Chem. Eur. J.*, 2004, **10**, 2565–2580.

11 (*a*) P. T. Corbet, J. K. M. Sanders and S. Otto, *J. Am. Chem. Soc.*, 2005, **127**, 9390–9392; (*b*) I. Saur and K. Severin, *Chem. Commun.*, 2005, **1**, 1471–1473.

12 (*a*) I. Takeuchi, J. Lauterbach and M. J. Fasolka, *Mater. Today*, 2005, **8**, 18–26; (*b*) H. Koinuma and I. Takeuchi, *Nat. Mater.*, 2004, **3**, 429–438; (*c*) X.-D. Xiang, X. Sun, G. Briceno, Y. Lou, K.-A. Wang, H. Chang, W. G. Wallace-Freedman, S. Chen and P. G. Schultz, *Science*, 1995, **268**, 1738–1740.

13 (*a*) F. L. Tietze, N. Rackelmann and G. Sekar, *Angew. Chem., Int. Ed.*, 2003, **42**, 4254–4257; (*b*) A. H. Hoveyda, *Chem. Biol.*, 1998, **5**, R187–R191; (*c*) P. Krattiger, C. McCarthy, A. Pfaltz and H. Wennemers, *Angew. Chem., Int. Ed.*, 2003, **42**, 1722–1724; (*d*) A. Berkessel, *Curr. Opin. Chem. Biol.*, 2003, **7**, 409–419.

14 (*a*) M. C. Sweeney, A.-S. Wavreille, J. Park, J. P. Butchar, S. Tridandapani and D. Pei, *Biochemistry*, 2005, **44**, 14932–14947; (*b*) K. M. Halkes, P. M. St. Hilaire, P. R. Crocker and M. Meldal, *J. Comb. Chem.*, 2003, **5**, 18–27; (*c*) R. E. Dolle, *J. Comb. Chem.*, 2001, **3**, 477–517.

15 M. H. V. Van Regenmortel, *J. Mol. Recognit.*, 2000, **13**, 1–4.

16 C. Schmuck and L. Geiger, *J. Am. Chem. Soc.*, 2004, **126**, 8898–8899.

17 For reviews on the early work of peptide recognition, see: (*a*) H.-J. Schneider, *Angew. Chem., Int. Ed. Engl.*, 1993, **32**, 848–850; (*b*) M. W. Peczuh and A. D. Hamilton, *Chem. Rev.*, 2000, **100**, 2479–2493; (*c*) For recent work of peptide recognition see: N. Srinivasan and J. D. Kilburn, *Curr. Opin. Chem. Biol.*, 2004, **8**, 305–310; (*d*) C. Chamorro, J.-W. Hofman and R. M. J. Liskamp, *Tetrahedron*, 2004, **60**, 8691–8697.

18 (*a*) C. C. McComas, B. M. Crowley and D. L. Boger, *J. Am. Chem. Soc.*, 2003, **125**, 9314–9315; (*b*) M. Nieto and H. R. Perkins, *Biochem. J.*, 1971, **123**, 780–803.

19 (*a*) C. T. Walsh, *Nature*, 2000, **406**, 775–781; (*b*) D. H. Williams and B. Bardsley, *Angew. Chem., Int. Ed.*, 1999, **38**, 1172–1193.

20 For recent reviews, see: (*a*) J. H. Hartley, T. D. James and C. J. Ward, *J. Chem. Soc., Perkin Trans. 1*, 2000, 3155–3184; (*b*) M. W. Peczuh and A. D. Hamilton, *Chem. Rev.*, 2000, **100**, 2479–2494.

21 C. Chamorro and R. M. J. Liskamp, *Tetrahedron*, 2004, **60**, 11145–11157.

22 J. Shepherd, T. Gale, K. B. Jensen and J. D. Kilburn, *Chem. Eur. J.*, 2006, **12**, 713–720.

23 K. B. Jensen, T. M. Braxmeier, M. Demarcus and J. D. Kilburn, *Chem. Eur. J.*, 2002, **8**, 1300–1309.

24 P. Krattiger, R. Kovasy, J. D. Revell, S. Ivan and H. Wennemers, *Org. Lett.*, 2005, **7**, 1101–1103.

25 For reviews on the work of target- and diversity-oriented synthesis in combinatorial chemistry, see: (*a*) S. L. Schreiber and M. D. Burke, *Angew. Chem., Int. Ed.*, 2004, **43**, 46–58; (*b*) M. D. Burke,

E. M. Berger and S. L. Schreiber, *Science*, 2003, **302**, 613–618; (*c*) M. D. Burke, E. M. Berger and S. L. Schreiber, *Science*, 2000, **287**, 1964–1969.
26 R. Breinbauer, I. R. Vetter and H. Waldmann, *Angew. Chem., Int. Ed.*, 2002, **41**, 2878–2890.
27 For articles on efficient binding motifs for carboxylates based on a guanidiniocarbonyl pyrrole group, see: (*a*) C. Schmuck and W. Wienand, *J. Am. Chem. Soc.*, 2003, **125**, 452–459; (*b*) P. A. Gale, *Coord. Chem. Rev.*, 2003, **240**, 191–221; (*c*) C. Schmuck, *Chem. Eur. J.*, 2000, **6**, 709–718; (*d*) C. Schmuck, *Chem. Commun.*, 1999, 843–844; (*e*) C. Schmuck, *Eur. J. Org. Chem.*, 1999, 2397–2403.
28 C. Schmuck and M. Heil, *Chem. Eur. J.*, 2006, **12**, 1339–1348.
29 For a description of the IRORI© tagging system, see: (*a*) K. C. Nicolaou, J. A. Pfefferkorn, H. J. Mitchell, A. J. Roecker, S. Barluenga, G.-Q. Cao, R. L. Affleck and J. E. Lillig, *J. Am. Chem. Soc.*, 2000, **122**, 9954–9967; (*b*) X.-Y. Xiao, Z. Parandoosh and M. P. Nova, *J. Am. Chem. Soc.*, 1997, **62**, 6029–6033; (*c*) A. W. Czarnik, *Curr. Opin. Chem. Biol.*, 1997, **1**, 60–66.
30 A. Borchardt and W. C. Still, *J. Am. Chem. Soc.*, 1994, **116**, 373–374.
31 (*a*) K. Hirose, *J. Inclusion Phenom. Macrocyclic Chem.*, 2001, **39**, 193–209; (*b*) Y. Cheng, T. Suenaga and W. C. Still, *J. Am. Chem. Soc.*, 1996, **118**, 1813–1814; (*c*) K. A. Connors, *Binding Constants*, John Wiley & Sons, New York, 1987.
32 C. Schmuck and M. Heil, *ChemBioChem*, 2003, **4**, 1232–1238.
33 C. Schmuck, M. Heil, J. Scheiber and K. Baumann, *Angew. Chem., Int. Ed.*, 2005, **44**, 7208–7212.
34 C. Schmuck and P. Wich, *Angew. Chem., Int. Ed.*, 2006, **45**, 4277–4281.