New promise in combinatorial chemistry: synthesis, characterization, and screening of small-molecule libraries in solution

Thomas Carell[1], Edward A Wintner[1], Andrew J Sutherland[1], Julius Rebek, Jr[1]*, Yuriy M Dunayevskiy[2] and Paul Vouros[2]*

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and [2]Department of Chemistry, Barnett Institute, Northeastern University, Boston, MA 02115, USA

Background: The increasing interest in combinatorial chemistry as a tool for the development of therapeutics has led to many new methods of creating molecular libraries of potential lead compounds. Current methods have made it possible to develop libraries of several million compounds. As a result, the limiting factor in the screening of libraries has become the identification and characterization of active species. We have recently described a method for generating libraries of water-soluble compounds containing mixtures of 10^4 to 10^5 different small organic molecules by using generally applicable solution phase chemistry. We set out to develop new methods to characterize and decode these libraries.

Results: Libraries were generated by condensing a multi-acid-chloride core molecule with various amines, producing molecules with functional groups about a rigid backbone. Composition and complexity of the libraries was evaluated using electrospray mass spectrometry to analyze model libraries containing up to 55 different molecules. The number of peaks obtained in mass spectrometry is directly correlated with the complexity of the library, and we were therefore able to deduce which of the expected compounds had in fact been formed in the library, and which of the building blocks in the library were not efficiently used. An iterative selection procedure was developed using this information, which allowed the screening of libraries of up to 50,000 chemical species to produce a competitive inhibitor of the enzyme trypsin.

Conclusions: Our strategy for the identification of active species should be broadly applicable to other methods of generating complex libraries of small molecules. The selection from the library of a compound with desired biological properties augurs well for the potential value of generating and screening complex mixtures of small molecules in solution.

Chemistry & Biology March 1995, **2**:171–183

Key words: combinatorial chemistry, drug design, electrospray mass spectrometry, molecular libraries, trypsin

Introduction

In the past decade, researchers have been intrigued by the promise of combinatorial chemistry to allow rapid access to vast numbers of new chemical entities. Whereas traditional organic chemistry produces single compounds of high purity in a labor-intensive synthetic process, combinatorial chemistry offers the means to generate a large number of different chemicals simultaneously. Modern analytical methods allow the screening of such combinatorial libraries to select with alacrity those species in the generated pool which possess desirable properties [1]. The first attempts at combinatorial synthesis involved peptides, and include the 'pin-method' devised by Geysen *et al.* [2,3] and the 'tea-bag' procedure invented in the Houghten group [4,5]. Both techniques allow the preparation of up to 300 individual peptides at the same time. With the introduction of the 'split-bead technology' of Furka *et al.* [6–9], the parallel synthesis of peptides was increased dramatically to several million compounds in a library of molecules supported on polymer beads. As a result, the limiting factor in the successful screening of libraries has become the characterization of active species; the amount of any individual compound present in a library of millions is too minute for its structure to be determined using conventional spectrometric methods [10].

To overcome the problems associated with structure determination, some researchers have focused on the development of oligonucleotide libraries (in which the active species can be amplified by the polymerase chain reaction prior to structure determination) [11–13] and on the preparation of oligopeptide libraries (in which the structure of an active peptide can be elucidated by direct sequencing) [8,9]. A new trend in combinatorial chemistry is the development of coding schemes. By using either an oligonucleotide code [14–16] or a chemical 'barcode' readable by gas chromatography [17–19], it is likely that it will be possible to efficiently determine the structure of non-peptide active species that are available in only picomolar quantities in the near future; this technology has already been applied successfully to peptides. Despite these recent advances in combinatorial chemistry, the production and screening of large libraries has

*Corresponding authors.

 ISSN 1074-5521

to date not revolutionized the world of drug discovery. One reason may be that the libraries so far investigated have consisted predominantly of oligopeptides and oligonucleotides. The nature of these linear molecules — which are relatively unrestricted in conformation and of low bioavailability — is such that even very active species are often of limited use in the development of new therapeutics. As a result, more drug-related compounds such as benzodiazepines [20], peptoids [21–23], oligocarbamates [24–26] and other synthetic compounds recently termed 'diversomers' [27,28] are now being investigated for their use in combinatorial chemistry. The ideal combinatorial method for the discovery of pharmaceutically valuable chemicals would produce large libraries of small organic molecules with a drug-related structure, yet would still lend itself to rapid synthesis, screening and compound structure determination.

Results

Synthesis of combinatorial libraries

We recently reported a procedure which approaches this goal of the ideal library [29,30]. In a two-step synthetic procedure, mixtures of small organic molecules with a complexity of 10^4–10^5 different compounds were efficiently generated. As our libraries were not made sequentially on a solid support, we were unable to make use of available bead-selection strategies [6–9, 17–19] or coding schemes [14–19]. Instead, an iterative selection procedure was designed through modification and generalization of a screening method first employed by Houghten and colleagues [31,32] for the identification of active peptides in a hexapeptide library solution. This approach trades the speed by which compounds can be identified when tagging procedures are employed for the broadened scope of organic reactions available in solution-phase chemistry. Selection and identification of the structure of an active compound from the new libraries were nevertheless achieved in a period of several weeks.

Ironically, the new approach to library production and selection represents a departure from the teachings of organic synthesis; instead of synthesizing single compounds of high purity, a large mixture of small molecules was synthesized in a single solution. The procedure, described schematically in Fig. 1, involves the coating of a rigid core molecule with a variety of

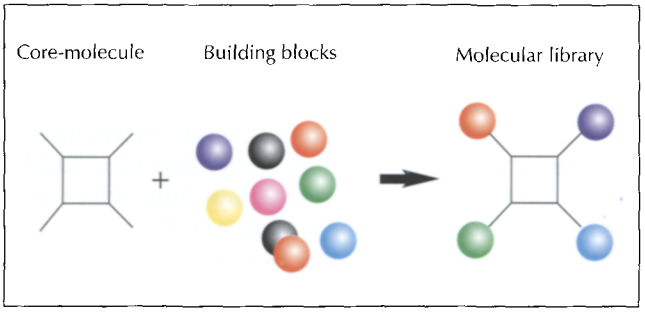


Fig. 1. Schematic representation of the procedure used to generate libraries of small organic molecules.

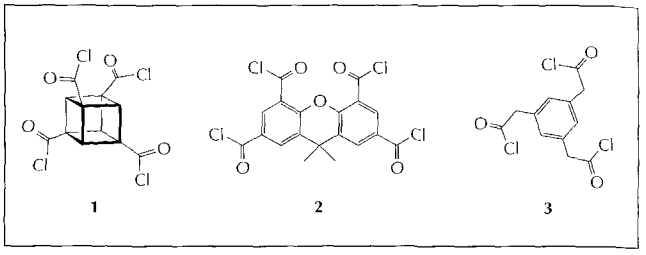


Fig. 2. Core molecules used to generate diverse molecular libraries.

building blocks in a stochastic process. By reacting the rigid core structure in solution with a mixture of small molecules able to form covalent bonds at the attachment sites of the core, libraries were produced in which all molecules possess the same overall shape (as determined by the architecture of the core) yet display a wide array of functionality (as determined by the functional groups of the building blocks).

As core molecules, we investigated the three acid chlorides **1**, **2**, and **3** shown in Fig. 2. We chose relatively rigid core structures from which functional groups would radiate outward, presuming that likely lead compounds would be small organic molecules with convex shapes able to complement the concave surfaces of macromolecular targets such as enzyme active sites, receptor cavities, or nucleic-acid grooves. The highly symmetric and compact cubane derivative (**1**) displays its four acid-chloride groups in a tetrahedral array, while the larger xanthene core molecule (**2**) presents a less symmetric arrangement and is planar in overall shape. The benzene triacid chloride (**3**) sweeps out a different shape-space in orientation of its functional groups from the xanthene or cubane molecules, being less rigid than either.

The cubane compound **1** was prepared from cubane mono-acid by photolysis in oxalyl chloride [33]. The xanthene **2** was prepared from commercial xanthenone via methylation of the 9 position (toluene, trimethylaluminum, 12 h, room temperature (r.t.)), followed by tetrabromination (Br_2, cat. Fe-powder, CH_2Cl_2, 2 h, 20 °C, 4 h reflux). Br/CN exchange (CuCN, N-methyl-2-pyrrolidinone, 2 h, reflux; 20 % nitric acid 10 h, reflux) and subsequent hydrolysis of the tetracyano compound (NaOH, water, 14 h, reflux) provided 9,9-dimethyl xanthene-2,4,5,7-tetracarboxylic acid, which was converted to the tetraacid chloride with oxalyl chloride (CH_2Cl_2, cat. N,N-dimethyl formamide (DMF), reflux, 4 h). The benzene compound **3** was prepared from 1,3,5-tris-carboxymethylbenzene [34] with oxalyl chloride (CH_2Cl_2, cat. DMF, reflux, 4 h).

The theoretical number of different molecules created by reaction of an acid-chloride core with a mixture of amine nucleophiles under ideal reaction conditions (identical reactivity and equimolarity of all building blocks) can be calculated from the combinatorial rule [35] and a set of symmetry factors. These factors depend

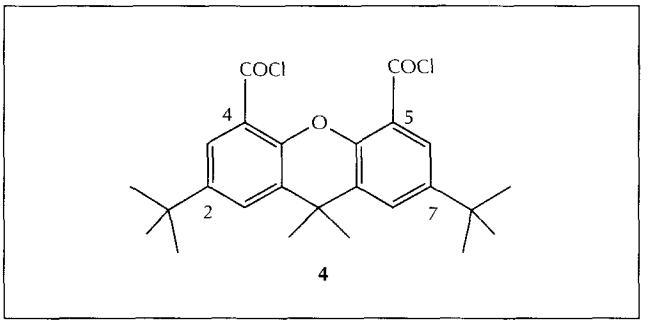
on the symmetry of the core and are determined individually for each core; the higher the symmetry of the core structure, the fewer compounds are generated with a given set of building blocks. For example, 19 different building blocks combined with the highly symmetric cubane core molecule would theoretically generate 11 191 possible compounds, while the same building blocks, when combined with the xanthene core, theoretically generate 65 341 compounds. (Details of the xanthene calculations are given in the Materials and methods section.)

For the preparation of the libraries, a two-step procedure of synthesis and deprotection was employed [29,30]. First, one equivalent of each acid-chloride core molecule was reacted with three or four equivalents (depending on the number of acid chlorides on the core) of an amine mixture. The amine mixtures contained equimolar mixtures of four to twenty building blocks (listed in Materials and methods). These building blocks were L-amino acid derivatives, although clearly a wide variety of nucleophiles could be used in this context. The L-amino acids represent a natural set of biologically relevant functionalities commercially available in their protected forms, and were therefore selected to establish a methodology for our chemistry.

To ensure that all building blocks reacted with the core molecules in high yields, we reacted the xanthene core molecule **2** sequentially with the 20 amines. Each amine listed gave the expected tetra-functionalized product in excellent yield within 30 min reaction time. Additional functional groups in the building blocks other than the desired amines which might react with acid chlorides (for example, the Ser side chain) were blocked with acid-labile protecting groups [36,37]. The hydrophobic nature of the libraries produced with the protected building blocks allowed their separation from unreacted amines and hydrolyzed core material by extraction with solutions of citric acid and sodium hydrogen carbonate. The resulting protected libraries were obtained as white powders even though they contained large mixtures of molecules. In the second step of library production, the protected libraries were treated with reagent K (a trifluoroacetic acid based reagent) [38] that cleaved the various *tert*-butyl, trityl, Boc, and Mtr protecting groups. Precipitation with ether gave white powders with solubility in 10:1 water/dimethylsulfoxide sufficient for screening purposes.

Mass-spectrometric determination of diversity

Electrospray ionization (ESI) mass spectrometry was used to establish that a sizable fraction of the expected compounds are produced in the above synthesis of water-soluble molecular libraries. Experiments were designed to determine if the combination of a xanthene acid chloride core with a set of amines results in the formation of a mixture in which all of the expected compounds are present above a certain concentration threshold. Direct observation of a library of 65 341 compounds was out of the question, and even a library of 136 compounds (only four building blocks) appeared daunting in size. Therefore, smaller yet representative libraries were constructed. Mass spectrometric analysis of these model libraries allowed us to probe the effectiveness of the general synthesis in forming complex mixtures of molecules. For an investigation of oligopeptide libraries with ion-spray mass spectroscopy, see [39].



4

Fig. 3. Truncated core molecule **4** for mass-spectrometric determination of model-library diversity

Based on three criteria, the diacid chloride **4** (Fig. 3) was selected to serve as a core molecule for the model libraries [40]. First, condensation of compound **4** with ten amines leads to only 55 disubstituted compounds, a seemingly manageable number. Second, the two amides formed are at the four and five positions of the xanthene scaffold. As these positions are most susceptible to problems of steric crowding about the xanthene core, compound **4** provided a realistic test for determining any building-block combinations which, because of their bulk, would be disfavored in the synthesis. Third, the two *tert*-butyl groups occupying the two and seven position allowed examination of the precipitation behavior of highly hydrophobic library compounds in the final treatment with ether/*n*-hexane.

For the synthesis of the model libraries, six sets of building blocks were selected from the 20 building blocks: three sets of eight amines, two of nine amines, and one of ten amines. Condensation of these sets of amines with the core molecule **4** was expected to give libraries of 36, 45 and 55 compounds, respectively. With the help of a simple computer program, the building blocks were grouped in sets such that nearly all of the compounds produced from a given set would possess a unique molecular weight. This simplified individual detection of each compound present in the libraries. After their formation, the protected model libraries were deprotected with reagent K [38], then further treated with a solution of trifluoroacetic acid/dichloromethane (4:1) to ensure that all protection groups were cleaved. Finally, the water-soluble model libraries were precipitated with ether and *n*-hexane to give white powders.

To investigate the mass-spectrometric behavior of these disubstituted molecules, several pure xanthene-diamino acids and test mixtures of three compounds (prepared by

condensing the core molecule **4** with one amine or with a mixture of two amines) were analyzed. The ESI mass spectra showed negligible fragmentation under the ion optics settings used, and revealed that ionization efficiency of the xanthene derivatives for a given mode of detection was determined by the nature of the attached amino acids. In positive-ion mode, the derivatives of positively charged amino acids (Lys, Arg, His) gave responses up to ten times higher than negatively- or non-charged derivatives. For mass spectra acquired in negative-ion mode, the opposite picture was observed.

All six model libraries were analyzed in both positive- and negative-ion modes. Taken together, the molecular ion peaks obtained from these measurements are a set of data directly correlated to the diversity of a given molecular library. Since fragmentation was not a factor, we were able to compare the molecular ion peaks in the mass spectra with the molecular weights expected for each model library, and conclude which compounds had been formed and which had not. Because the composition of the model libraries was chosen such that most of the compounds were expected products in more than one model library, the presence or absence of most combinations was checked at least twice. Molecules were considered present if the signal-to-noise ratios of their molecular ion peaks exceeded 10:1; the most intense signals observed had signal-to-noise ratios of 600:1.

Results of the mass spectroscopy (MS) analysis of three model libraries are compiled in Fig. 4. Library L1 was generated by condensation of the truncated core molecule **4** with ten of the twenty amine building blocks; libraries L2 and L3 were generated with eight of the building blocks. On the x- and y-axes are the abbreviations of the building blocks used. Each filled square in the charts represents the presence of one of the expected compounds in that model library. Because the truncated core molecule possesses C_{2v} symmetry, only half of the possible building block combinations give rise to new compounds. The color of the squares indicates whether the corresponding compound was detected as its positive ion (red), its negative ion (blue) or with tandem mass spectrometry (gray). An exemplary mass spectrum of model library L2 is shown in Fig. 5.

Out of 55 expected compounds in model library L1, 43 (78 %) were detected. Missing were predominantly those compounds which contained the arginine-*p*-benzylamide building block (ArgA). Of ten expected signals, only the molecular ion peak of His-ArgA fell within limits of detection. As a result of these findings, the ArgA amine was removed from the set of building blocks used to prepare large libraries for screening purposes. Only three other combinations in model library L1, the Gly-Gly, Gly-Tyr, and Tyr-Tyr compounds, could not be detected. Whereas the Tyr-Tyr combination could be detected in two other libraries (for example, L2), the Gly-Gly combination was missing in all other model libraries we tested. This indicated that in the large libraries used for screening purposes, compounds with two or more Gly building blocks might not be present. The Gly-Tyr combination was tested only in model library L1.

Apart from the compounds containing ArgA, 93 % of the expected compounds are present in model library L1. A

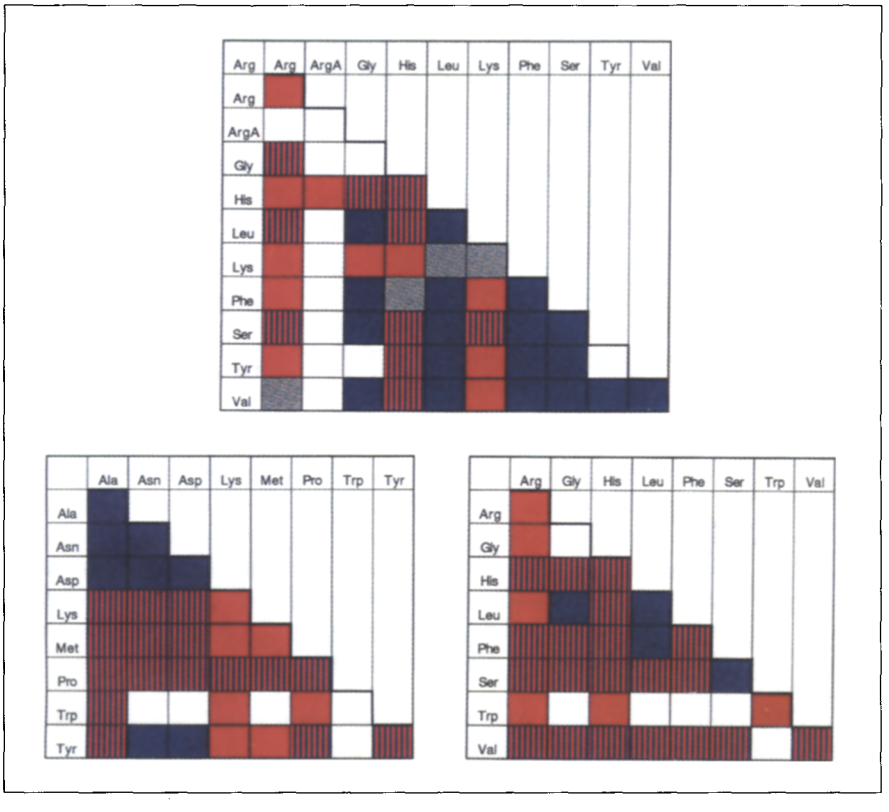


Fig. 4. Mass-spectrometric analysis of libraries L1 (top), L2 (bottom left), and L3 (bottom right). Red: detected as positive ion. Blue: detected as negative ion. Red-blue striped: detected as both positive and negative ion. Gray: detected through MS/MS experiment.

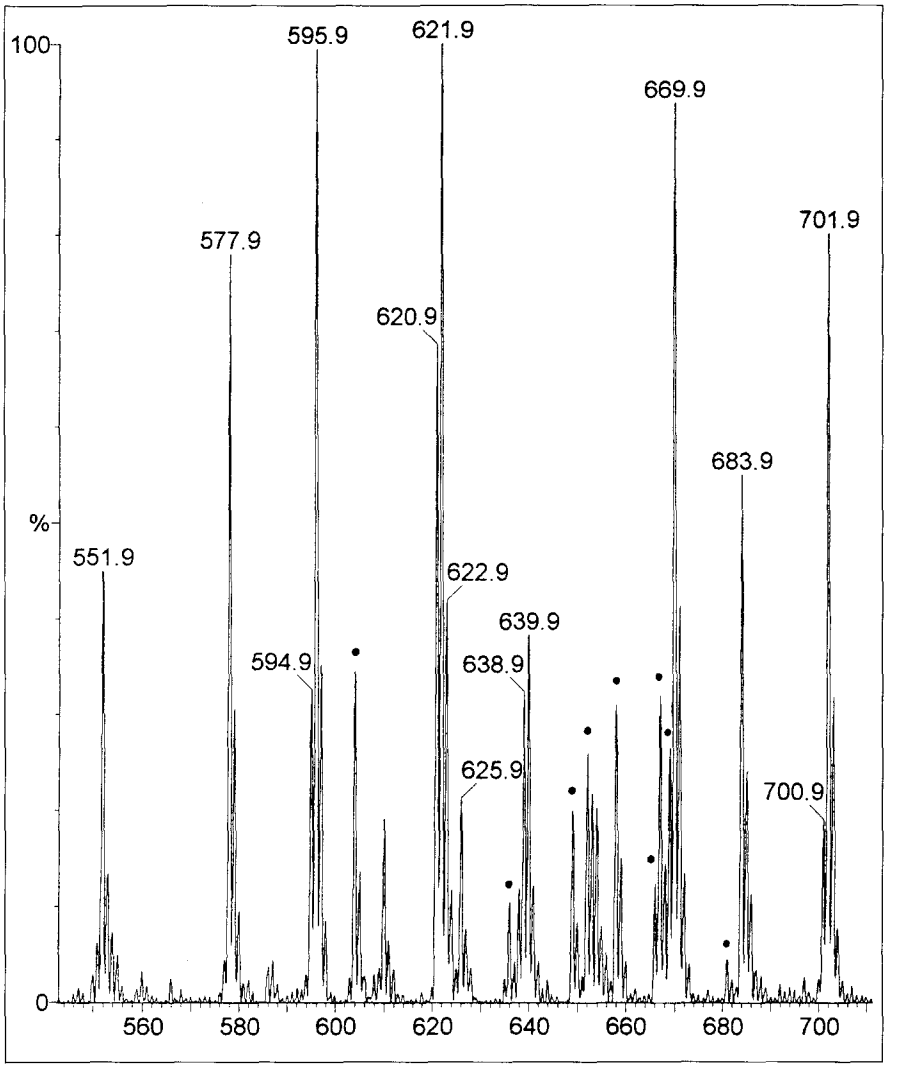


Fig. 5. Negative ion scan (M/Z) of library L2, consisting of the xanthene compound **4** (see Fig. 3) disubstituted with various amino acids. Expected peaks are marked by their molecular weights (M-1) or by a dot (•) for clarity: Ala-Ala (552), Ala-Pro (578), Ala-Asn (595), Ala-Asp (596), Pro-Pro (604), Pro-Asn (621), Pro-Asp (622), Ala-Lys (623), Ala-Met (626), Asn-Asn (638), Asn-Asp (639), Asp-Asp (640), Pro-Lys (649), Pro-Met (652), Ala-Tyr (658), Asn-Lys (666), Asp-Lys (667), Asn-Met (669), Asp-Met (670), Ala-Trp (681), Pro-Tyr (684), Asn-Tyr (701), Asp-Tyr (702). Not shown: Tyr-Tyr (764). The peak at 610 does not match any of the expected products for library L2.

very similar picture was obtained from the analysis of libraries L2 and L3. Here, only one more troublesome building block, tryptophane-methyl ester, was found; less than 50 % of the expected Trp compounds were detected. Nevertheless, in libraries L2 and L3, both of which theoretically contained 36 different compounds, 86 % and 83 % of the expected compounds were present. The absence of Trp compounds may be due to degradation; Trp is an amino acid known to be vulnerable to strongly acidic deprotection conditions. However, as each model library contained some of the expected Trp-substituted xanthenes, this building block was included in the set of building blocks eventually used for large libraries.

One problem which we foresaw in the analysis of the model libraries by positive-ion ESI was the possible overlap, due to C_{13} isotope contribution, of ion peaks differentiated by only one or two atomic mass units (amu). In most of our model libraries, this problem was overcome by the use of negative-ion ESI as a second dimension. As the carboxyl group was the only acidic group in the model libraries, none of the compounds disubstituted with methylester building blocks appeared in negative-ion mass spectra. Thus, many peaks obscured in positive-ion mass spectra were clearly seen in negative-ion mode. Any remaining ambiguities were resolved through tandem mass spectrometry (MS/MS) experiments. The molecular ions of the xanthene-diamino acids underwent collisionally induced fragmentation to form characteristic daughter ions, allowing unambiguous assignment of library compounds via MS/MS even in cases of isobaric molecular ion peaks. For example, MS/MS experiments were applied to model library L1, in which the disubstituted xanthene pairs Arg-Val/Leu-Lys and Lys-Lys/His-Phe had isobaric molecular weights 665 and 694, respectively. In both cases, the resulting MS/MS spectra contained the characteristic fragments of both molecules, confirming the presence of each in the model library. Although unnecessary for our model libraries, chromatographic techniques coupled online to MS can be used to add another dimension of resolving power to the system. Investigation of more complex libraries is under way using capillary electrophoresis coupled to MS (CEMS).

The mass-spectrometric analyses of the model libraries revealed that most of the amines chosen as building

blocks generate the expected condensation products with the truncated core molecule **4**. Exceptions are arginine-*p*-benzylamide, which was removed from future sets of building blocks, and tryptophane-methyl ester. If one makes the reasonable assumption that reaction at the 2 and 7 positions of **2** is no more difficult than reaction at the 4 and 5 positions of **4**, it follows that most of the compounds expected in a large tetra-acid chloride based library will be formed. The compounds which are most probably absent in such libraries contain one or more Trp or multiple Gly building blocks. However, even if 75 % of the expected Trp-containing compounds are absent in a condensation of **2** with 19 building blocks (see screening libraries below), the water-soluble library created still contains over 50 000 different molecules. Mass-spectrometric analysis of the libraries revealed further that very non-polar library molecules are generally present in lower amounts than polar compounds. This effect is probably caused by the final isolation step in which the library material is precipitated with ether and *n*-hexane. However, these non-polar molecules were found to be present in high enough concentrations for the selection assays described below.

In the synthesis of a large mixture of molecules, the generation of some side products was inevitable. MS and MS/MS analyses of the model libraries indicate that relatively few side products were formed in the deprotected, ether/*n*-hexane-precipitated material, however. In the six model libraries, only seven significant peaks were observed that did not match the expected molecular ion peaks for disubstituted xanthenes. The structures of the side products were tentatively assigned from MS/MS data, and include a xanthene-mono-histidine-mono-acid derivative and two mono-amino-acid-mono-alkylamine variants each of His, Lys and Arg. The fragmentation behavior of the xanthene diamino acids and the structure determination of the side products present in the mixtures will be reported elsewhere. Very low concentrations of incompletely deprotected xanthene-diamino acids could also be detected. Overall, however, we estimate a level of side products below 10 % of the total number of compounds in a given library.

Even though the quantitative analysis of mixtures with a diversity of 10^4 to 10^5 is not yet possible, the preparation of specifically designed model libraries and their investigation by ESI mass spectrometry has proven to be a useful technique for the analysis of mixtures of structurally related compounds. With the above mass-spectrometric data in hand, we were confident that our synthetic methodology produced highly diverse libraries of well-defined composition, and we were able to put faith in the results of subsequent screening assays.

Screening of combinatorial libraries against trypsin

We chose to test our system of combinatorial library generation against the readily available enzyme trypsin (for information on screening a peptide library against trypsin, see [41]). Trypsin is a digestive enzyme which is a member of the important class of serine proteases involved in biological functions such as blood clotting (for example, thrombin [42,43]), and blood-pressure regulation (for example, renin). Thrombin inhibitors have been isolated from DNA aptamer libraries [44]. Our goal was to isolate one or more active compounds from our libraries through an iterative selection process. Specifically, we sought to isolate a compound able to inhibit the trypsin-catalyzed cleavage of the amide bond in N^{α}-benzoyl-DL-arginine-*p*-nitroanilide [26, 42–46].

The results of the screening procedure, described briefly in our previous communications [29,30], are reviewed and expanded upon below. For screening purposes, we created three libraries, A1–A3, based on core molecules **1–3**, respectively. Each acid-chloride core was condensed with 19 of the amino-acid derivatives (omitting ArgA). This procedure gave three libraries, theoretically containing 11 191 (cubane core), 65 341 (xanthene core) and 1 330 (benzene core) different compounds. To evaluate the potency of each library, 2.5 mg of a given library was added to the trypsin assay and its inhibitory activity was correlated inversely to the rate of *p*-nitroaniline released from trypsin-catalyzed hydrolysis. The results of this experiment are depicted in round A of Table 1.

Only the xanthene-based library, A2, caused a significant reduction of the enzyme's activity (~30 %). To determine which of the 19 building blocks were most responsible for this interaction, six sublibraries were synthesized. First, cysteine (an amino acid known to cause artifactual results in some enzyme assays) was removed summarily from the set of building blocks, leaving 18 amines. Sublibraries B1–B6 were then prepared with the xanthene core molecule **2** and fifteen of the 18 building blocks, each sublibrary missing three specific amines. Each sublibrary was calculated to consist, in theory, of 25 425 different tetrasubstituted xanthenes.

The screening results of the six mixtures are depicted in round B of Table 1, and show that the sublibrary B3, generated without the building blocks Arg, Lys and His, did not inhibit enzyme activity. This result was not entirely unexpected, given the known preference of trypsin for Lys and Arg at the carbonyl side of the scissile amide bond (often referred to as the P1 position) [47]. Sublibraries B1 and B2 (generated without Gly, Ala, Val and Leu, Ile, Pro respectively) showed the next highest activities. The nine building blocks missing in sublibraries B1–B3 were therefore deemed to be most responsible for the presence of inhibitors in the initial library A2.

As a control for this result, library C1 was constructed with the nine building blocks Gly, Ala, Val, Leu, Ile, Pro, Arg, Lys and His. Activity of the control library was measured relative to other libraries, C2–C7, generated with a random set of 9 to 12 of the 19 amines initially used in library A2. Round C in Table 1 clearly shows that only the control library, C1, possesses significant

Table 1. Screening of combinatorial libraries for inhibition of trypsin.

Round A		Round B		Round C		Round D	
Library	**A (rel)**	**Library**	**A (rel)**	**Library**	**A (rel)**	**Library**	**A (rel)**
Blank	100	Blank	100	Blank	100	Blank	100
A1 Cubane	91	B1 -Gly,Ala,Val	79 X	C1	31 X	D1 -Arg	28
A2 Xanthene	66 X	B2 -Leu,Ile,Pro	85 X	C2	90	D2 -Lys	93 X
A3 Benzene	98	B3 -Arg,Lys,His	95 X	C3	85	D3 -His	21
		B4 -Ser,Thr,Met	68	C4	100	D4 -Leu	46 X
		B5 -Phe,Tyr,Trp	74	C5	82	D5 -Ile	50 X
		B6 -Glu,Asp,Asn	54	C6	90	D6 -Pro	80 X
				C7	82	D7 -Gly	32
						D8 -Ala	27
						D9 -Val	59 X

Round E		Round F				Round G	
Library	**A (rel)**	**Library**	**Positions 4 and 5**	**Positions 2 and 7**	**A (rel)**	**Library**	**A (rel)**
Blank	100	Blank	—	—	100	Blank	100
E1 Lys, Pro, Val, Ile	20 X	F1	Lys, Val	Ile, Pro	37	G1 Tetra COOH	100
E2 Lys, Pro, Val, Leu	26	F2	Lys, Ile	Val, Pro	11 X	G2 Isomer **10**	3
E3 Lys, Pro, Val, Gly	45	F3	Lys, Pro	Ile, Val	100	G3 Isomer **11**	20
E4 Lys, Pro, Val	28	F4	Val, Ile	Lys, Pro	100		
E5 Pro, Val, Ile	100	F5	Val, Pro	Lys, Ile	100		
E6 Lys, Val, Ile	55	F6	Ile, Pro	Lys, Val	100		
E7 Lys, Pro, Ile	100						

Some of the data in this table have previously been published [29,30]. A (rel): percent of trypsin activity in the presence of libraries. Each value represents average trypsin activity for four measurements; measurements are made relative to a blank containing no added library material. Libraries were considered inhibitory if they reduced trypsin activity by > 10 %. The activity of the blank was set to 100 %. X denotes that the library was selected to influence the next round of screening.

Library solutions added in rounds A–D: 2.5 mg in 50 μl DMSO. Round E: 1.5 mg in 50 μl DMSO; round F: 0.5 mg in 50 μl DMSO; round G: 0.25 mg in 50 μl DMSO.

Round A: % of trypsin activity with the initial libraries A1, A2, and A3 constructed from the core molecules **1**, **2**, and **3** respectively plus 19 of the 20 amino acid building blocks (ArgA omitted).

Round B: % of trypsin activity with the six sublibraries B1–B6 constructed from core molecule **2** and 15 of the 18 amino acid building blocks. B1: Gly, Ala, Val omitted. B2: Leu, Ile, Pro omitted. B3: Arg, Lys, His omitted. B4: Ser, Thr, Met omitted. B5: Phe, Tyr, Trp omitted. B6: Glu, Asp, Asn omitted.

Round C: % of trypsin activity with the seven sublibraries C1–C7 constructed from the core molecule **2** and nine to twelve of the amino acid building blocks. C1: Arg, Lys, His, Leu, Ile, Pro, Gly, Ala, Val. C2: His, Leu, Ile, Pro, Ala, Val, Phe, Tyr, Trp, Ser, Thr, Met. C3: Arg, Lys, His, Gly, Phe, Tyr, Trp, Glu, Asp, Asn. C4: His, Leu, Ile, Pro, Ala, Val, Phe, Trp, Glu, Asp, Asn. C5: Arg, Lys, His, Gly, Phe, Tyr, Trp, Ser, Thr, Met. C6: His, Leu, Ile, Pro, Gly, Ala, Val, Phe, Trp, Glu, Asp, Asn. C7: Arg, Lys, His, Pro, Gly, Trp, Glu, Asp, Asn.

Round D: % of trypsin activity with the nine sublibraries D1–D9 constructed from the core molecule **2** and eight of the nine building blocks Arg, Lys, His, Leu, Ile, Pro, Gly, Ala, Val. D1: Arg omitted. D2: Lys omitted. D3: His omitted. D4: Leu omitted. D5: Ile omitted. D6: Pro omitted. D7: Gly omitted. D8: Ala omitted. D9: Val omitted.

Round E: % of trypsin activity with the seven sublibraries E1–E7 constructed from the core molecule **2** and three or four of the building blocks Lys, Leu, Ile, Pro, Gly, Val. E1: Lys, Pro, Val, Ile. E2: Lys, Pro, Val, Leu. E3: Lys, Pro, Val, Gly. E4: Lys, Pro, Val. E5: Pro, Val, Ile. E6: Lys, Val, Ile. E7: Lys, Pro, Ile.

Round F: % of trypsin activity with the six sublibraries F1–F6. Sublibraries were variously substituted with Lys, Ile, Pro and Val at xanthene positions 2, 4, 5 and 7.

Round G: % trypsin activity with the two final inhibitors. G1: control run with 9,9-dimethyl-2,4,5,7-xanthene tetracarboxylic acid. G2: isomer **10**. G3: isomer **11**.

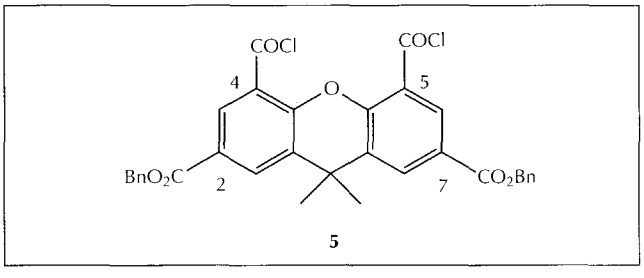
biological activity. This experiment helped to rule out the possibility that the observed differences in inhibitory activity of the sublibraries resulted from differences in library synthesis and handling rather than the absence of certain important building blocks. Because the control library C1 contains, in theory, only 3321 compounds compared to 65 341 present in the initial library A2, and because building-block selection has increased the content of inhibitors (2.5 mg of material now produces 69 % inhibition), the process of building-block selection and sublibrary synthesis can be viewed as an 'amplification' step.

To further narrow the field of possible inhibitors, nine new xanthene sublibraries were prepared using eight of the nine building blocks already selected, with each sublibrary lacking one of these nine amino acids. The screening results obtained from this experiment are presented in round D of Table 1. A sublibrary which showed low inhibitory activity signaled that the building block omitted from that group was crucial for trypsin inhibition. For generation of an active trypsin inhibitor, Lys methyl ester appeared to be the most important building block, followed by Pro, Val, Ile, and Leu.

To establish the most potent combination of the above five building blocks, xanthene libraries were constructed with three or four of these five. (It was by no means clear at this stage that the most potent molecule would contain four different building blocks.) Trypsin assay results are shown in round E of Table 1. Library E1, synthesized with the building blocks Lys, Val, Pro, and Ile, had the highest inhibitory activity, and omission of any one of these four building blocks gave libraries with a lower potency. Thus, the search for a trypsin inhibitor was narrowed to a sublibrary containing 136 compounds which included twelve structural isomers of the Lys-Ile-Pro-Val-xanthene derivative. Since Round E (Table 1) stipulated that the most potent trypsin inhibitor in our libraries was created with four different building blocks, it was clear that one or more of these twelve isomers was the inhibitor in question.

To narrow the remaining possibilities for the structure of the most potent isomer from twelve to two, six new sublibraries (F1–F6, Table 1) were prepared using the dibenzylester xanthene diacid chloride derivative **5** (Fig. 6). The sublibraries were synthesized in a two-step procedure in which compound **5** was treated in a first 'randomization' step with two of the four amines Lys, Ile, Pro, and Val, followed by deprotection of xanthene positions 2 and 7 by hydrogenolysis (ethyl acetate/ethanol, 10 % Pd/C, H_2-atmosphere, 2 h, r.t.). Coupling of the resulting material with the two other building blocks (BOP) and deprotection of the acid-sensitive protection groups with trifluoroacetic acid in CH_2Cl_2 yielded six sublibraries, each with a unique distribution of the four selected building blocks around the xanthene core.

The screening results of these six sublibraries are presented in round F, Table 1. They revealed that only those



5

Fig. 6. Core molecule used for synthesis of sublibraries **F1–F6** (see Table 1).

compounds possessing the combination Lys/Val or Lys/Ile at the 4 and 5 positions and the corresponding Pro/Ile or Pro/Val combination at positions 2 and 7 were active as trypsin inhibitors. Other arrangements of the four selected building blocks on the xanthene core were inactive. Furthermore, of the two active libraries F1 and F2, library F2 (containing the Lys/Ile combination at xanthene positions 4 and 5) was the most active. This result narrowed the structure of a final most potent inhibitor to the two isomers **10** and **11** shown in Fig. 7.

Compounds **10** and **11** were individually prepared via a four-step synthesis from **5**, as outlined in Fig. 7. The dibenzyl-protected diacid chloride **5** was reacted with a mixture of Lys and Ile (CH_2Cl_2, triethylamine, 1.5 h, r.t) and the mixed amide xanthene-monolysine-monoisoleucine compound **6** isolated by flash chromatography (31 % yield). Hydrogenolysis (ethyl acetate / ethanol, 10 % Pd/C, 2 h, r.t.) of the benzylester protecting groups yielded the monolysine-monoisoleucine xanthene diacid **7** (98 %). Coupling of **7** with a Pro/Val building block mixture (BOP, dimethylformamide, triethylamine, 1 h, r.t.) yielded a set of four protected compounds from which two isomeric compounds A and B — both containing all four building blocks (A_1 = Lys, A_2 = Ile, B = Pro and Val) — were isolated by flash chromatography and purified by normal-phase preparative high-pressure liquid chromatography (HPLC).

The assignment of the isomers to the structures of protected **8** and **9** was possible by evaluating two nuclear Overhauser effect (NOE) measurements and a correlated spectroscopy (COSY) spectrum of isomer B. Individual irradiation at the absorption frequencies (δ = 8.00 and δ = 7.93) of two aromatic xanthene protons connected to the same six-membered ring gave strong NOEs with two NH-protons. These protons were assigned to the Val and Ile sub-structures through the COSY spectrum of isomer B. Isomer B therefore corresponds to protected compound **9**, in which Ile and Val are connected to the same benzene ring. Comparison of the chemical shifts of the four aromatic xanthene protons of the isomers **8** and **9** with other tetra-amino acid substituted xanthenes confirmed this assignment.

In the final step of the synthesis, compounds **8** and **9** were deprotected with trifluoroacetic acid in dichloromethane (4 h, r.t., 91 % yield for **10**; 97 % yield for **11**) . The products, compounds **10** and **11**, were purified by reverse-phase preparative HPLC and characterized by fast-atom bombardment mass spectrometry in combination with a detailed study of the fragments by tandem mass spectrometry. The assignment of signals in the NMR spectra was difficult due to several slow conformational interconversions such as a *cis–trans* isomerization of the Pro amide bond. Some of the protons could be assigned to signals in a spectrum measured at 90 °C with the help of a COSY spectrum obtained at 90 °C. NOE studies of compound **9** suggested that, on an NMR time scale, the amide proton of the lysine side

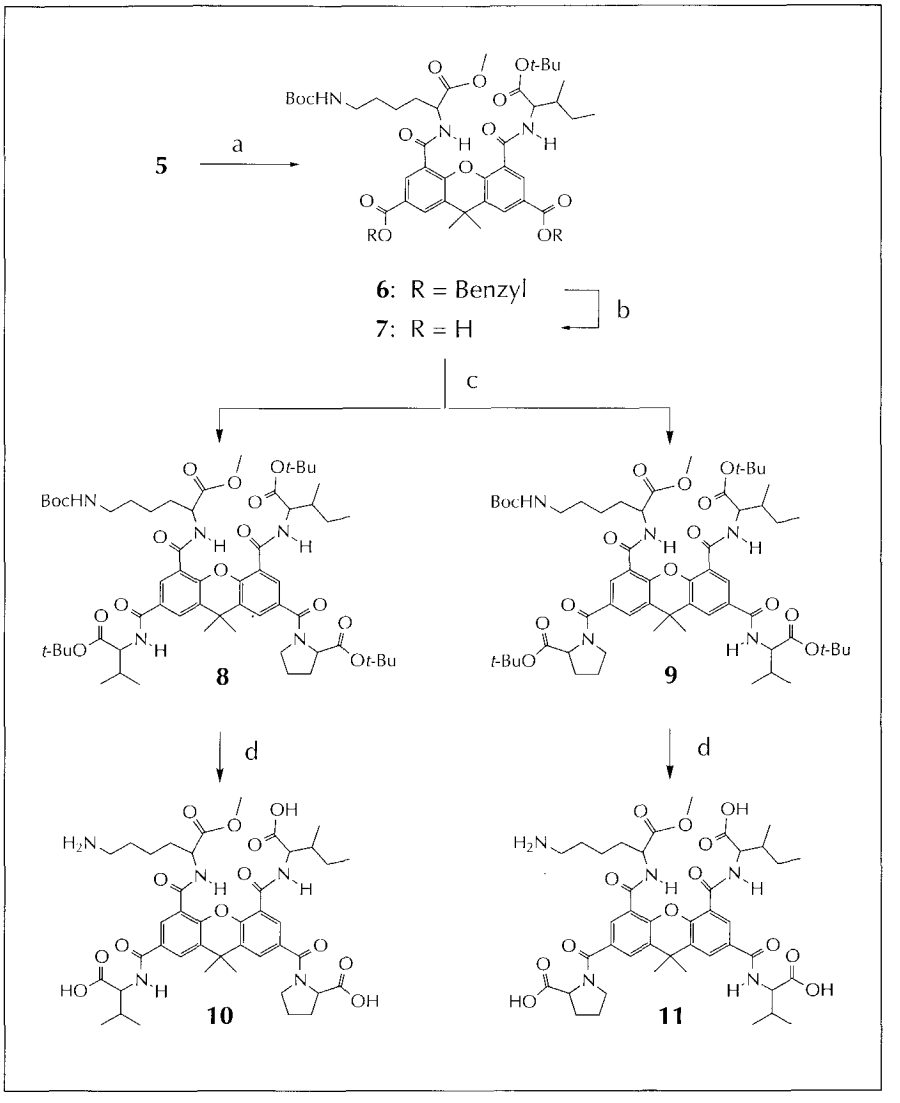


Fig. 7. Synthesis of the final trypsin inhibitor **10** and its isomer **11**. **(a)** H-Lys(Boc)-OMe, H-Ile-O*tert*-butyl, CH_2Cl_2, triethylamine, 1.5 h, r.t., 31 % yield. **(b)** ethyl acetate/ethanol, 10 % Pd/C, H_2-atmosphere, 2 h, r.t., 98 % yield. **(c)** H-Pro-O*tert*-butyl, H-Val-O*tert*-butyl, dimethylformamide, BOP, triethylamine, 1 h, r.t., 30 % yield (**8**), 28 % yield (**9**). **(d)** CH_2Cl_2, TFA, thioanisol, 4 h, r.t., 91 % yield (**10**), 97 % yield (**11**).

chain is locked into a bifurcated hydrogen bond with the xanthene oxygen and the carbonyl oxygen of the adjacent isoleucine. Similar hydrogen bonding was observed in previous work with xanthene structures in our laboratory [24].

Screening of compounds **10** and **11** in the standard assay (round G, Table 1) revealed that both isomers are trypsin inhibitors, with compound **10** being the most potent. The K_i values of both compounds **10** and **11** were obtained by non-linear regression of kinetic data according to the equation for competitive inhibition and additionally by evaluation of Lineweaver–Burke plots. The kinetic data were obtained by measurement of the rate of *p*-nitroaniline released by tryptic cleavage of the substrate benzoyl-L-arginine-*p*-nitroanilide at increasing concentrations with an increasing presence of either compound **10** or compound **11** [48]. These measurements clearly show that both isomers **10** and **11** are competitive trypsin inhibitors, with a K_i of 9.4±0.8 μM for **10** and a K_i of 72±7 μM for **11**. Thus, structure **10** represents the most potent trypsin inhibitor that was selected from our initial xanthene library A2.

The means by which compound **10** inhibits trypsin activity has not yet been determined, but computer modeling suggests that the lysine methyl ester building block is hydrogen bound to Asp189 in the trypsin-specificity pocket (for an introduction to the structure of serine proteases, see [49]). In the model, the xanthene lies flat over the active site of the protein, allowing the other three building blocks to make specific hydrophobic contacts. The model suggests in essence that molecule **10** mimics a trypsin substrate, but is only slowly cleaved by the trypsin 'catalytic triad.' This theory is supported by the fact that after complete inhibition by compound **10**, trypsin activity does slowly recover. Preliminary tests during our K_i determination experiments showed that trypsin activity returned to ~10 % of normal after 24 h.

Discussion

Clearly, the activity initially observed in the xanthene-based library A2 (which should theoretically contain 65 341 compounds) cannot be explained solely by the presence of the inhibitors **10** and **11**. Other inhibitory compounds must have been present in the starting library as well. In the selection strategy, only the most

Table 2. Computer simulation of relative activities of xanthene libraries.

Round b		Round d		Round p		Round q	
Library	A (rel)	Library	A (rel)	Library	A (rel)	Library	A (rel)
b1 -Gly,Ala,Val	81 X	d1 -Arg	39	p1	76 X	q1 -Phe	47
b2 -Leu,Ile,Pro	96 X	d2 -Lys	100 X	p2	92 X	q2 -Met	47
b3 -Arg,Lys,His	100 X	d3 -His	32	p3	46	q3 -Ile	74 X
b4 -Ser,Thr,Met	44	d4 -Leu	57 X	p4	65 X	q4 -Leu	60 X
b5 -Phe,Tyr,Trp	49	d5 -Ile	71 X	p5	39	q5 -Val	85 X
b6 -Glu,Asp,Asn	39	d6 -Pro	76 X	p6	100 X	q6 -Trp	47
		d7 -Gly	32			q7 -Ser	40
		d8 -Ala	42			q8 -Pro	77 X
		d9 -Val	86 X			q9 -Tyr	39
						q10 -Lys	100 X
						q11 -Asp	40
						q12 -Arg	46

A (rel): computer generated relative activity (%) for simulated xanthene libraries. An X denotes selection of a given library to influence the next round of screening.

Round b: (compare to round B in Table 1) relative trypsin activity with the six sublibraries b1–b6. Libraries were simulated as condensations of core molecule **2** and 15 of the 18 amino acid building blocks. b1: Gly, Ala, Val omitted. b2: Leu, Ile, Pro omitted. b3: Arg, Lys, His omitted. b4: Ser, Thr, Met omitted. b5: Phe, Tyr, Trp omitted. b6: Glu, Asp, Asn omitted.

Round d: (compare to round D in Table 1) relative trypsin activity with the nine sublibraries d1–d9. Libraries were simulated as condensations of core molecule **2** and eight of the nine building blocks Arg, Lys, His, Leu, Ile, Pro, Gly, Ala, Val. d1: Arg omitted. d2: Lys omitted. d3: His omitted. d4: Leu omitted. d5: Ile omitted. d6: Pro omitted. d7: Gly omitted. d8: Ala omitted. d9: Val omitted.

Round p: relative trypsin activity with the six sublibraries p1–p6. Libraries were simulated as condensations of core molecule **2** and 15 of the 18 amino acid building blocks, but using an alternate grouping from that of Table 1. p1: Phe, Met, Ile omitted. p2: Leu, Val, Trp omitted. p3: Ala, Thr, Gly omitted. p4: Ser, Pro, Tyr omitted. p5: His, Asn, Glu omitted. p6: Lys, Asp, Arg omitted.

Round q: relative trypsin activity with the twelve sublibraries q1–q12. Libraries were simulated as condensations of core molecule **2** and eleven of the twelve building blocks Phe, Met, Ile, Leu, Val, Trp, Ser, Pro, Tyr, Lys, Asp, Arg. q1: Phe omitted. q2: Met omitted. q3: Ile omitted. q4: Leu omitted. q5: Val omitted. q6: Trp omitted. q7: Ser omitted. q8: Pro omitted. q9: Tyr omitted. q10: Lys omitted. q11: Asp omitted. q12: Arg omitted.

active building blocks or groups of building blocks were selected for the generation of further sublibraries. These choices guided us directly to the inhibitors **10** and **11**, with compound **10** being eight times more active than its structural isomer **11**. The question then arises whether the described selection strategy automatically results in the isolation of the most active compound in a given library. Would a different initial grouping of the building blocks have resulted in the isolation of structurally different inhibitors? Could the strategy fail to select other compounds present in library A2 that had a higher inhibitory activity than compound **10**?

To address these questions, a computer program was written to simulate the activities observed for all the discussed sublibraries. The simplifying assumptions were that 1) inhibition was due to molecules binding at the active site of trypsin, and 2) each building block at a certain xanthene position added an incremental value to the 'energy of binding' for that molecule. Based on the empirical evidence of molecules **10** and **11**, the values of 'binding energy' for each building block at each xanthene position were adjusted until the program was able roughly to reproduce sublibrary activities. As shown in Table 2 (lower case letters for libraries denote computer generated results), the program reproduced the results of Table 1 such that libraries b1–b3 would be selected over libraries b4–b6, and Lys, Pro, Val, Ile and Leu would be selected from the sublibraries d1–d9. The program reproduced rounds B and D qualitatively, but not quantitatively.

Application of the computer program gave several insights into our selection strategy. First, given the assumptions made, it seems likely that molecule **10** was merely the tip of a broad family of inhibitors; beneath the tip lay other less-potent inhibitors which nevertheless contributed greatly to the activity of the computer-generated libraries. This explains the relatively high activity of the initial xanthene library A2. Assuming that the computer program mirrors reality, if one postulates a 'molecular landscape' [50] in which valleys mark compounds that bind weakly and mountains mark families of inhibitors, our selection procedure is less a screening for individual inhibitors than it is a group selection procedure which provides a means of ascending a feature in a molecular landscape to its peak of activity.

The program further gave evidence that when given a single, well defined mountain in a molecular landscape, our screening procedure is not dependent on the grouping of building blocks to find the highest point of activity. In the initial simulated screening procedure, the combination Lys, Ile, Pro, Val was selected independently of how the building blocks were grouped. For example,

when grouped as in libraries p1–p6 (Table 2), four sublibraries p1, p2, p4, and p6 would be selected, leading to sublibraries q1–q12 (Table 2). The sublibraries q3, q4, q5, q8 and q10 would be selected from this round, once again converging to building blocks Ile, Leu, Val, Pro and Lys (compare to actual screening round D).

When the computer program was changed to introduce a second peak in the molecular landscape (that is, a second set of inhibitors with different building blocks was defined with an energy of binding equal to or greater than that of **10**), the landscape feature (and therefore the ultimate inhibitory molecule) which the computer selected did depend on initial groupings of building blocks. Simulations demonstrated that by employing our selection strategy, it was impossible to miss an active compound orders of magnitude more potent than compound **10**, but it was possible to overlook a compound with comparable or even slightly higher activity. The computer program thus confirmed that the described selection strategy of ever-narrowing sublibraries will result in the isolation, if they exist, of one or more of the most active compounds present in a combinatorial library such as A2.

Significance

In a few weeks of work, the procedures used in this experiment in combinatorial chemistry yielded a single compound giving micromolar inhibition of trypsin from a large library of small organic molecules free in solution. Whether or not this xanthene derivative is the best possible inhibitor in the library, the iterative screening method is effective. This success suggests that the synthesis and screening of molecular libraries from a core molecule and a variety of building blocks could be a valuable tool in the search for potent therapeutic lead compounds. Our results show that the procedure is a useful means of exploring a structural landscape of small molecules, complementing the existing repertoire of combinatorial procedures. It is emphasized that while this study used predominantly the readily available L-amino acid derivatives and only three core molecules, any nucleophile that would give products of reasonable hydrolytic stability could be used with these or other core molecules. Furthermore, the procedure could easily be adapted to optimize lead structures. Specifically, the four building blocks selected in this study, Lys, Ile, Pro, and Val, could be re-combined with the xanthene core and a new set of similar amines (for example, homolysine and ornithine). The resulting library could be re-screened to select still more potent trypsin inhibitors. Given the large number of possible amines and multi-acid-chloride cores, the potential of this method for the generation of new libraries is unbounded.

Materials and methods

General

All commercially available reagents were used without further purification. The diacid of compound **4** is commercially available from Aldrich Chemical Company. All solvents were purchased from Malinckrodt, and were analytical-reagent quality. Bovine pancreas trypsin was purchased from Sigma. All protected amino acids were obtained from Advanced ChemTech and Novabiochem. ^{1}H NMR spectra were obtained on Bruker AC-250, Varian XL-300, Varian UN-300 and Varian VXR-500 spectrometers. For preparative HPLC a Waters 600E system was used with a Waters 490I multiwavelength detector (Det. 275 nm). All UV measurements were performed on a Perkin-Elmer Lambda-2 spectrometer. Mass spectra for product characterization were obtained on a Finnigan MAT 8200 system. For the analysis of the diversity of the generated libraries by electrospray mass spectrometry, a VG Quatro instrument was used.

Standard procedure for the preparation of acid chlorides

To synthesize an acid chloride [29,30], 0.6 mmol of the corresponding acid was dissolved in 20 ml of dichloromethane. 1.5 mmol oxalyl chloride were added in the case of the diacids 9,9-dimethyl-2,7-di-*tert*-butyl xanthene-4,5-dicarboxylic acid and 9,9-dimethyl xanthene-2,7-dicarboxylic acid benzylester-4,5-dicarboxylic acid, 2.5 mmol oxalyl chloride for 1,3,5-tris-carboxymethylbenzene-tricarboxylic acid and 3 mmol oxalyl chloride for 9,9-dimethyl xanthene-2,4,5,7-tetracarboxylic acid. After addition of two drops of dimethyl-formamide, the reaction mixture was heated to reflux for 4 h and concentrated *in vacuo* to yield compounds **4**, **5**, and **2** as tan solids and compound **3** as a yellow oil. The products were dried *in vacuo* over night. The acid chlorides were used without further purification.

Standard procedure for the preparation of libraries

List of the amine building blocks used to prepare libraries for mass-spectrometric investigations and screening purposes: L-alanine-*tert*-butyl ester (Ala), N^g-4-methoxy-2,3,6-trimethylbenzene-sulfonyl-L-arginine (Arg), N^g-4-methoxy-2,3,6-trimethylbenzene-sulfonyl-L-arginine-*p*-methoxy-benzyl amide (ArgA), L-asparagine-*tert*-butyl ester (Asn), O^4-*tert*-butyl-L-aspartic acid *tert*-butyl ester (Asp), S-trityl-L-cysteine (Cys), O^5-*tert*-butyl-L-glutamic acid *tert*-butyl ester (Glu), glycine-methyl ester (Gly), N^{im}-trityl-L-histidine (His), L-isoleucine-*tert*-butyl ester (Ile), L-leucine-*tert*-butyl ester (Leu), N^ε-Boc-L-lysine-methyl ester (Lys), L-methionine-methyl ester (Met), L-phenylalanine-*tert*-butyl ester (Phe), L-proline-*tert*-butyl ester (Pro), O-*tert*-butyl-L-serine-*tert*-butyl ester (Ser), O-*tert*-butyl-L-threonine-methyl ester (Thr), L-tryptophane-methyl ester (Trp), O-*tert*-butyl-L-tyrosine-methyl ester (Tyr), L-valine-*tert*-butyl ester (Val).

To make a library [29,30], 0.22 mmol of the acid-chloride core molecule was dissolved in 10 ml of dichloromethane. The amine mixture composed of an equimolar mixture of the desired amines (total molarity of amines = total molarity of acid-chloride groups) was added and the reaction mixture was vigorously stirred. Triethylamine (0.5 ml) was added and the reaction mixture stirred for 2 h at r.t. The reaction mixture was diluted with 250 ml of dichloromethane and washed twice with 1 M citric acid solution, twice with saturated sodium hydrogen carbonate solution, and once with water. The organic phase was separated, dried with $MgSO_4$ and concentrated *in vacuo* to afford a tan foam.

Table 3. Diversity calculations.

Combination type	C_{2v} symmetry multiplier	Combinatorial rule [35]	Combinations for *m*= 4	8	15	19
AAAA	1	$m!/1!(m-1)!$	4	8	15	19
AAAB/AABB	8	$m!/2!(m-2)!$	48	224	840	1368
AABC	18	$m!/3!(m-3)!$	72	1008	8190	17442
ABCD	12	$m!/4!(m-4)!$	12	840	16380	46512
		Total combinations	136	2080	25425	65341

Standard procedure for the deprotection of libraries

To deprotect libraries [29,30] 100 mg of the library material was stirred with reagent K [38] (6 ml) at r.t. for 4 h. The solution was concentrated *in vacuo* to yield a brown oil. The libraries were precipitated by addition of 15 ml of a mixture of diethylether/*n*-hexane (1:1). The tan precipitate was filtered off and washed three times with diethylether/*n*-hexane (1:1). The obtained library material was redissolved in dichloromethane/trifluoroacetic acid (1:4, 20 ml) and stirred at r.t. for another 3 h. The reaction mixture was concentrated *in vacuo* and the oily residue solidified by addition of 15 ml of diethylether/*n*-hexane (1:1). The precipitated final library material was filtered off, washed four times with diethylether/*n*-hexane (1:1), and dried *in vacuo* overnight.

Screening procedure

Bovine pancreatic trypsin (2.5 mg) was dissolved in 10 ml of an 0.005 M solution of HCl in water (trypsin stock solution). Then, 50 mg of N-benzoyl-D,L-arginine-*p*-nitroanilide were dissolved in 5 ml of dimethylsulfoxide (substrate stock solution). Library material was dissolved in 200 µl of dimethylsulfoxide (for amount of library material, see Table 1). Of this library stock solution, 50 µl was diluted with 500 µl of buffer (pH 8.2, 0.5 M Tris HCl, 0.04 M $CaCl_2$) and 10 µl of the trypsin stock solution. High buffer concentration was necessary to stabilize the pH in those experiments where 2.5 mg of library material was added; buffer concentration was constant for all screening and K_i determination experiments. The slightly cloudy mixtures were centrifuged for 5 min, and 400 µl of the clear supernatant was transformed into a 1.0 ml disposable UV cuvette filled with 500 µl of buffer. (Only the supernatant was used to avoid UV absorption by cloudy mixtures.) The solutions were mixed and the UV absorption at 405 nm was measured and noted (0-value). Then, 50 µl of the substrate solution was added and the solutions mixed thoroughly. UV absorption at 405 nm was measured at 10 min, 20 min, and 40 min, and the obtained absorption minus the 0-value was noted (activity value) and used for activity comparisons. Final activity percentage was obtained by comparison of activity value to the blank activity value measured without added library material (blank activity = 100 %).

Diversity calculations

Calculation of the theoretical number of tetra-functionalized compounds produced with xanthene core **2** and *m* building blocks is shown in Table 3. The total number of compounds equals the sum of the number of compounds with one (AAAA), two (AABB, AAAB), three (AABC), and four (ABCD) different building blocks.

Computer simulation

The computer program, designed to simulate the relative inhibition of various xanthene libraries, used the following scheme:

1. Based on empirical evidence, generate a 'binding score' for all building blocks at a given xanthene position. For example, at xanthene position 5, give Ile +3, Leu or Val +2.5, any other hydrophobic building block +1, etc.
2. Based on the number of building blocks chosen, create a library of one each of all possible xanthene molecules containing four amino acids.
3. For each molecule in the library, sum the binding scores of the building blocks at positions 2, 4, 5 and 7.
4. Based on the total binding score, generate a score for the amount of inhibition which each molecule contributes. For example, inhibition score = $e^{(\text{binding score})}$.
5. Sum the inhibition scores of all molecules in the library to get the total reduction in trypsin activity.
6. Compare the reduction in trypsin activity to other libraries generated with the same number of building blocks.
7. If the comparison of library inhibition does not qualitatively match real-life comparison of the same libraries, adjust the binding scores in step 1 and repeat.

Supplementary material available

Individual synthetic procedures and product characterizations, preparation of sublibraries F1–F6, synthesis of compounds **10** and **11**, kinetic studies, Lineweaver–Burke plots.

Acknowledgements: We thank Dr B Tsao for the measurement of COSY-spectra and NOE-experiments, JS Nowick and Ken Shimizu for the original syntheses of 2 and 5, Dr Bashir-Hashemi for providing cubane-1,3,5,7-tetracarboxylic acid chloride 1, and Professor J Stubbe for advice regarding the enzyme assay. This research was supported by the National Institutes of Health, the Alexander-von-Humboldt Foundation (Feodor-Lynen postdoctoral fellowship to Dr T Carell), the National Science Foundation, and the American Chemical Society, Organic Division (predoctoral fellowship and graduate fellowship to E. A. W.).

References

1. Jung, G. & Beck-Sickinger, A.G. (1992). Multiple peptide synthetic methods and their applications. *Angew. Chem. Int. Ed. Engl.* **31**, 367–383.
2. Geysen, H.M., Meloen, R.M. & Barteling, S.J. (1984). Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. Natl. Acad. Sci. USA* **88**, 3998–4002.
3. Geysen, H.M., Rodda, S.J., Mason, T.J., Tribbick, G. & Schoofs, P.G. (1987). Strategies for epitope analysis using peptide synthesis. *J. Immunol. Methods* **102**, 259–274.
4. Houghten, R.A. (1985). A general method for the rapid solid phase synthesis of large numbers of peptides. *Proc. Natl. Acad. Sci. USA* **82**, 5131–5135.
5. Houghten, R.A., DeGraw, S.T., Bray, M.K., Hoffmann, S.R. & Frizzell, N.D. (1986). Simultaneous multiple peptide synthesis. *Biotechniques* **4**, 522–528.

6. Furka, Á., Sebestyén, F., Asgedom, M. & Dibó, G. (1991). A general method for the rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res.* **37**, 487–493.
7. Furka, Á., Sebestyén, F., Asgedom, M. & Dibó, G. (1988). Abstr. 14th Int. Congr. Biochem., Prague, Czechoslovakia. Vol. 5, p. 47.
8. Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M. & Knapp, R.J. (1991). A new type of synthetic peptide library for the identification of ligand binding activity. *Nature* **354**, 82–84.
9. Salmon, S.E., *et al.*, & Felder, S. (1993). Discovery of biologically active peptides in random libraries. *Proc. Natl. Acad. Sci. USA* **90**, 11708–11712.
10. Brummel, C.L., Lee, I.N.W., Zhous, Y., Benkovic, S.J. & Winograd, N. (1994). A mass spectrometric solution to the address problem of combinatorial chemistry. *Science* **264**, 399–402.
11. Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment. *Science* **24**, 505–510.
12. Bartel, D.P. & Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418.
13. Beaudry, A.A. & Joyce, G.F. (1992). Directed evolution of an RNA enzyme. *Science* **257**, 635–641.
14. Brenner, S. & Lerner, R.A. (1992). Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381–5383.
15. Nielsen, J., Brenner, S. & Janda, K.D. (1993). Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc.* **115**, 9812–9813.
16. Kerr, J.M., Banville, S.C. & Zuckermann, R.N. (1993). Encoded combinatorial peptide libraries containing non-natural amino acids. *J. Am. Chem. Soc.* **115**, 2529–2531.
17. Ohlmeyer, M.H.F., *et al.*, & Still, W.C. (1993). Complex synthetic chemical libraries indexed with molecular tags. *Proc. Natl. Acad. Sci. USA* **90**, 10922–10926.
18. Borchardt, A. & Still, W.C. (1994). Synthetic receptor binding elucidated with an encoded combinatorial library. *J. Am. Chem. Soc.* **116**, 373–374.
19. Nestler, H.P., Bartlett, P.A. & Still, W.C. (1994). A general method for molecular tagging of encoded combinatorial chemical libraries. *J. Org. Chem.* **59**, 4723–4724.
20. Bunin, B.A. & Ellman, J.A. (1992). A general and expedient method for the solid phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.* **114**, 10997–10998.
21. Zuckermann, R.N., *et al.*, & Moos, W.H. (1994). Discovery of nanomolar ligands for seven transmembrane G-protein coupled receptors from a diverse peptoid library. *J. Med Chem.* **37**, 2678– 2685.
22. Simon, R.S., Martin, E.S., Miller, S.M., Zuckermann, R.N., Blaney, J. M. & Moos, W.H. (1994). *Techniques in Protein Chemistry. Part V.* J. Wiley & Sons., San Diego.
23. Kessler, H. (1993). Peptoids — a new approach to the development of pharmaceuticals. *Angew. Chem. Int. Ed. Engl.* **32**, 543–544.
24. Cho, C.Y., *et. al.*, & Schultz, P.G. (1993). An unnatural biopolymer. *Science* **261**, 1303–1305.
25. Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T. & Solas, D. (1991). Light directed, spacially addressable, parallel chemical synthesis. *Science* **251**, 767–773.
26. Rozsnyai, L.F., Benson, D.R., Fodor, S.P.A. & Schultz, P.G. (1992). Photolithographic immobilization of biopolymers on a solid support. *Angew. Chem. Int. Ed. Engl.* **31**, 759–761.
27. DeWitt, S.H., Kieley, J.S., Stankovic, C.J., Schroeder, M.C., Reynolds-Cody, D.M. & Pavia, M.R. (1993). Diversomers — an approach to nonpeptide, nonoligomeric chemical diversity. *Proc. Natl. Acad. Sci. USA* **90**, 6909–6913.
28. Liskamp, R.M.J. (1994). Opportunities for new chemical libraries — unnatural biopolymers and diversomers. *Angew. Chem. Int. Ed. Engl.* **33**, 633–636.
29. Carell, T., Wintner, E.A., Bashir-Hashemi, A. & Rebek, J., Jr. (1994). A novel procedure for the synthesis of libraries containing small orgainic molecules. *Angew. Chem. Int. Ed. Engl.* **33**, 2059–2061.
30. Carell, T., Wintner, E.A. & Rebek, J., Jr. (1994). A solution phase screening proceedure for the isolation of active compounds from a library of molecules. *Angew. Chem. Int. Ed. Engl.* **33**, 2061–2064.
31. Houghten, R.A., Pinilla, C., Blondelle, S.E., Appel, J.R., Dooley, C.T. & Cuervo, J.H. (1991). Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **354**, 84–86.
32. Pinilla, C., Appel, J.R., Blanc, P. & Houghten, R.A. (1992). Rapid identification of high affinity peptide ligands using positional scanning synthetic peptide combinatorial libraries. *Biotechniques* **13**, 901–905.
33. Bashir-Hashemi, A. (1993). Photochemical carboxylation of cubanes. *Angew. Chem. Int. Ed. Engl.* **32**, 612–613.
34. Newman, M.S. & Lowrie, H.S. (1954). Synthesis of 1,3,5-benzene triacetic acid. *J. Am. Chem. Soc.* **76**, 6196–6197.
35. Weiss, N.A. & Hassett, M.J. (1991). *Introductory Statistics.* (3rd edn), pp. 218–219, Addison-Wesley Publishing Co., New York.
36. Bodanszky, M. (1984). *Principles of Peptide Synthesis.* Springer Verlag, Heidelberg.
37. Bodanszky, M. & Bodanszky, A. (1984). *The Practice of Peptide Synthesis.* Springer Verlag, Heidelberg.
38. King, D., Fields, C. & Fields, G. (1990). A cleavage method which minimizes side reactions following FMOC solid phase peptide synthesis. *Int. J. Pept. Protein Res.* **36**, 255–266.
39. Metzger, J.W., Wiesmüller, K-H., Gnau, V., Grünges, J. & Jung, G. (1993). Mass spectrometry of synthetic peptide libraries. *Angew. Chem. Int. Ed. Engl.* **32**, 894–896.
40. Nowick, J.S., Ballester, P., Ebmeyer, F. & Rebek, J., Jr. (1990). Convergent functional groups 9: complexation in new molecular clefts. *J. Am. Chem. Soc.* **112**, 8902–8906.
41. Eichler, J. & Houghten, R.A. (1993). Identification of substrate-analog trypsin inhibitors through the screening of synthetic peptide combinatorial libraries. *Biochemistry* **32**, 11035–11041.
42. Zwaal, R.F.A. & Henker, H.C. (1986). *Blood Coagulation.* Elsevier, New York.
43. Salzman, E.W. (1992). Low molecular weight heparin and other new antithrombotic drugs. *New Eng. J. Med.* **326**, 1017–1019.
44. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E. H. & Toole, J.J. (1992). Selection of single stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566.
45. Erlanger, B.F., Kokowsky, N. & Cohen, W. (1961). The preparation and properties of two new chromogenic substrates of trypsin. *Arch. Biochem. Biophys.* **95**, 271–278.
46. Gaertner, H.F. & Puigserver, A.J. (1992). Increased activity and stability of poly(ethylene glycol) modified trypsin. *Enzyme Microb. Technol.* **14**, 150–155.
47. Laskowski, M. Jr & Kato, I. (1990). Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593–626.
48. Segel, I.H. (1993). *Enzyme Kinetics.* John Wiley & Sons, Inc., New York.
49. Branden, C. and Tooze, J. (1991). *Introduction to Protein Structure.* Garland Publishing, Inc., NY.
50. Holland, J.H. (1992). Genetic algorithms. *Sc. Am.* 66–72.

Received: **23 Jan 1995**; revisions requested: **30 Jan 1995**; revisions received: **27 Feb 1995**. Accepted **28 Feb 1995**.