From Split-Pool Libraries to Spatially Addressable Microarrays and Its Application to Functional Proteomic Profiling**

Nicolas Winssinger, Jennifer L. Harris, Bradley J. Backes, and Peter G. Schultz*

Numerous technologies have been developed to investigate cellular events on a genome-wide scale. Oligonucleotide arrays provide information on changes in mRNA expression levels in response to a variety of physiological stimuli.[1, 2] Two-dimensional gel electrophoresis,[3] or other chromatographic separation methods, in conjunction with mass spectroscopy[4] offer a more direct analysis of proteome function.[5] Technologies have also been developed for genome-wide analysis of protein structure.[6] In a more targeted analysis of protein function, maps of protein–protein[7] and protein–DNA[8] interactions have been reported as well as preliminary work towards a protein chip.[9] Methods to monitor the catalytic activity of proteins on a genome-wide scale also provide critical insights into cellular activity.[10, 11]

Small molecules have long been used to analyze and control the catalytic activity of enzymes as well as to modulate biological networks by acting as agonists or antagonists of receptors.[12] As such, microarrays of small molecule inhibitors or substrates provide a tool for profiling cellular activity. If one hopes to discriminate between the more than 30000 potential gene products in humans, it is clear that microarrays containing large collections of compounds will be a necessity. High density microarrays of peptides and synthetic oligomers have been reported (40000 compounds cm^{-2}),[13] however the photolithographic techniques used limit the range of accessible molecular diversity. More recently, several small molecules have been printed on a glass slide in an effort to merge robotic printing and split-pool libraries.[14] While split-pool library synthesis[15] is far more efficient for the generation of molecular diversity than parallel synthesis,[16] the identity of each library member is unknown and must be individually decoded for each active library member.[17, 18] If one wishes to screen such libraries against more than 30000 gene products, the decoding of library members becomes problematic.

Herein we report a method for the preparation of small molecule microarrays using positionally encoded libraries and its application to functional proteomic profiling in a model system. Libraries of small molecules tethered to peptidonucleic acid (PNA)[19] tags are constructed as shown in Figure 1. The PNA tag serves two purposes, first to encode the synthetic history of the small molecule and second to positionally encode the identity of the small molecule by its location upon

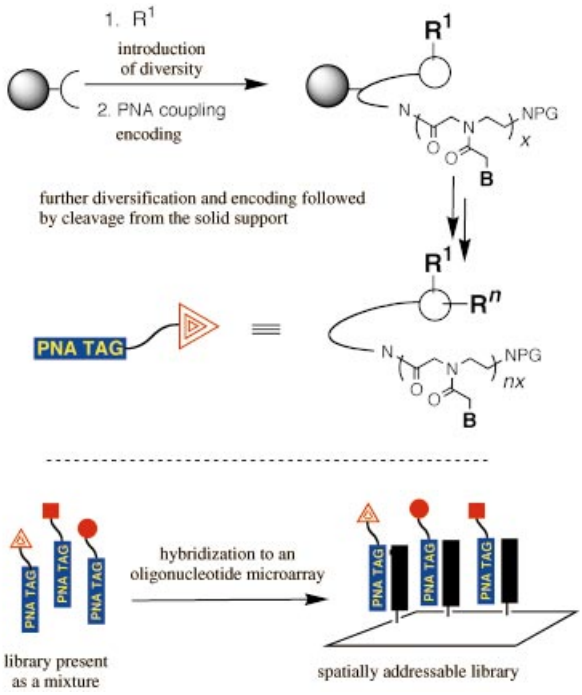


Figure 1. Split-pool synthesis of PNA-encoded combinatorial libraries. R^n = element of diversity present in library, B = base of the peptidonucleic acid, x = number of bases encoding a single element of diversity, n = number of chemical diversity-introducing steps, PG = protecting group.

hybridization to an oligonucleotide microarray. The choice of PNA as the encoding oligonucleotide was based on its desirable hybridization properties, the flexibility of its synthesis, and its chemical robustness.[20] For library synthesis, the oligomerization of PNAs relies on an amide bond formation, one of the mildest reaction in organic chemistry. Furthermore, the wide array of possible protecting groups for the nitrogen of the PNA's N-terminus should accommodate a wide range of diversity-introducing reactions. Finally, in terms of hybridization properties, the lack of negative charges on the PNA backbone increases its affinity for DNA and reduces the influence of salt concentration on hybridization strength.

PNA-encoded libraries of protein ligands can be screened against several targets simultaneously by incubating the library with the various targets containing different fluorophores. Upon hybridization of the mixture to an oligonucleotide array, fluorescent detection reveals the identity and selectivity of library members for each target. Although attractive for drug discovery, this strategy does not lend itself to profiling biological samples since it is difficult to uniformly label all proteins in a sample of interest. Conversely, the PNA–small molecule conjugate can be synthesized with a fluorophore such that, upon incubation with a sample of interest, the PNA–small molecule conjugate bound to a macromolecule can be separated from the unbound PNA–

[*] P. G. Schultz, N. Winssinger, J. L. Harris, B. J. Backes
Department of Chemistry and the Skaggs Institute for Chemical Biology
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla CA 92037 (USA)
Fax: (+1) 858-784-9440
E-mail: schultz@scripps.edu
and
The Genomics Institute of the Novartis Research Foundation
3115 Merryfield Row, Suite 200, San Diego, CA 92121 (USA)

[**] Funding was provided by the Skaggs Institute for Chemical Biology (P.G.S.) the Genomics Institute of the Novartis Research Foundation (J.L.H. and B.J.B.) and a fellowship from the National Institute of Health (N.W.). We thank Affymetrix for their generous donation of the GenFlex tag arrays.

  1433-7851/01/4017-3152 $ 17.50+.50/0

small molecule conjugate by size-exclusion chromatography.[21] As shown in Figure 2, hybridization of the high molecular weight fraction to a chip will reveal the identity of the small molecules bound to macromolecules, thereby generating a profile of protein function. The correlation between profiles and phenotypes may be rapidly assessed in the biological system using the small molecules identified in the profile while their molecular target(s) can be determined by affinity chromatography.

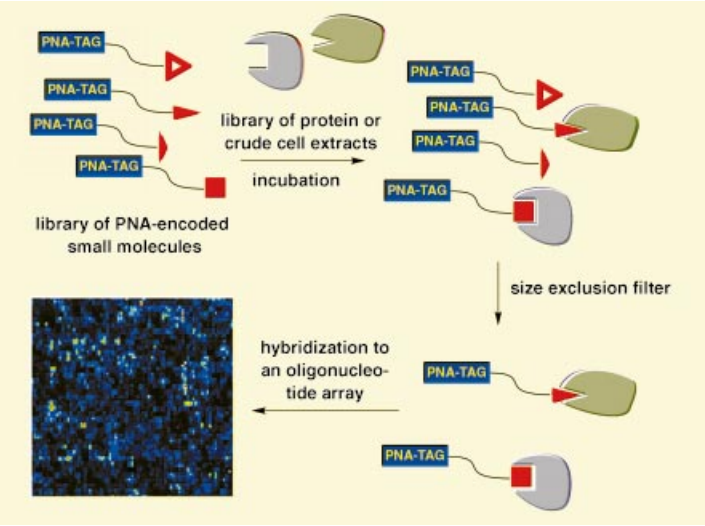
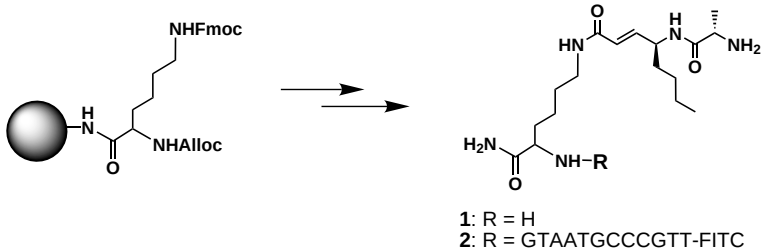


Figure 2. Screening of PNA-encoded libraries.

In order to demonstrate the feasibility of this approach, we focused on mechanism-based cysteine protease inhibitors containing an acrylamide functionality.[22] Preliminary studies to determine the optimal length of PNA indicated that 12mers have good hybridization properties and allow ample sequence variation to encode very large libraries. The synthesis was carried out on acid-labile Rink resin with mutually compatible Fmoc and Alloc protecting groups for the inhibitor and PNA synthesis, respectively. The side chains and bases were protected with acid-labile groups (Scheme 1).[23] The design of inhibitors was based on the information gathered from a previously developed method to rapidly assess the substrate specificity of proteases.[24] A comparison of the activity (Table 1) of compounds **1** and **2** (Scheme 1) against cathepsin C and cathepsin L reveals that the PNA tag does not significantly affect the activity or selectivity of the inhibitor



Scheme 1. Synthesis of the designed cathepsin L inhibitor with (**2**) and without (**1**) PNA tag on a Rink amide resin. Fmoc = 9-fluorenylmethoxycarbonyl, Alloc = allyloxycarbonyl.

Table 1. Kinetic inhibition of cathepsins by acrylamide inhibitors **1** and **2**.

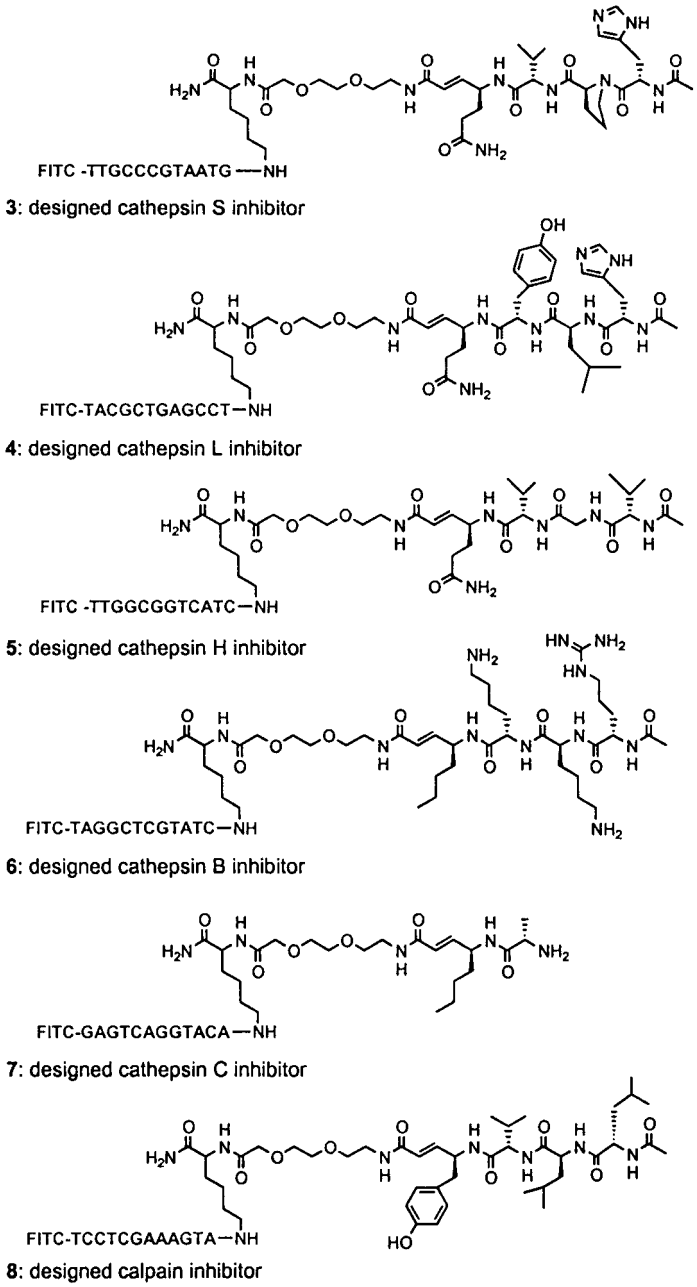
	Cathepsin C		Cathepsin L	
	IC_{50} [μM]	$k_{inact.}/K_i$ [$M^{-1}s^{-1}$]	IC_{50} [μM]	$k_{inact.}/K_i$ [$M^{-1}s^{-1}$]
1	17.6	40	> 2000	–
2	14.1	70	> 2000	–

(Table 1). An additional PEG spacer was incorporated in the library synthesis to insure good water solubility.

A series of compounds designed to inhibit cathepsin S, L, H, B, C, and calpain (Scheme 2) were synthesized. The PNA sequences were selected to hybridize to the terminal twelve residues of the 20mer probes of a GenFlex tag array[25] (arrays of this type are currently available at a density of 400000 features cm^{-2}). The PNA tags only hybridize to a portion of the array probe, and it was expected that each probe would have different hybridization properties. Hybridization of a mixture of the six probes (0.45 pmol of each in 150 μL) afforded the results shown in Figure 3, panel A. The difference in intensity of each array feature reflects the differences in melting temperature of the individual probes. Importantly, despite such differences in melting temperature, 30% changes in probe concentration were reliably detected. An equimolar mixture of the six compounds (**3**–**8**, 28 pmol) was incubated with commercially available purified cathepsin C (100 μM, 20 μL[26]) for 2 hours at 23 °C, passed through a size exclusion column (BioRad, Bio-Sil, SEC 125-5) to remove material below 10 kDa, and hybridized to a GenFlex tag array. As shown in Figure 3, panel C, hybridization afforded the expected signal for the probe corresponding to cathepsin C, whereas a control experiment lacking cathepsin C gave no signal (panel B). The same experiment was performed with cathepsin L (10 μM, 20 μL[26]) using ten-fold less protein. Direct detection of the fluorescein gave a weak signal, but this signal could be amplified using an anti-fluorescein goat antibody followed by a biotininalted anti-goat antibody and phycoerythrin-labeled streptavidin (Figure 3, panel D). These results show that the proposed size exclusion separation is effective to separate the bound PNA–ligand conjugates from the unbound ones, that PNA is efficient for positional encoding, and that small molecule–PNA conjugates can be used to probe protein function in a microarray format.

Cathepsins have been postulated to play important roles in many physiological and pathological functions including antigen presentation,[27] tumor progression,[28] osteoporosis,[29] cataract formation,[30] and parasitic virulence.[31] A library based on the described tools enables the identification of specific cathepsin inhibitors, the rapid diagnostic identification of specific cathepsin activity in biological samples, and the identification of novel proteases. The ability to array small molecule libraries prepared by split-pool combinatorial synthesis in a spatially addressable format allows for multiplexed screening in a highly miniaturized format to generate profiles of cellular activity. This approach should also be applicable to other classes of proteases, kinases, and phosphatases, as well as receptors. Screening large libraries as a single mixture in solution prior to hybridization may

  1433-7851/01/4017-3153 $ 17.50+.50/0



Scheme 2. Chemical structure of designed PNA-tagged cysteine protease inhibitors **3**–**8**. FITC = fluorescein.

prove useful as it minimizes nonspecific interactions of surfaces with the targets. Finally, this technology may prove to be a general and practical solution to prepare and screen microarrays of other biomolecules including antibodies and other proteins.

Received: May 10, 2001 [Z17078]

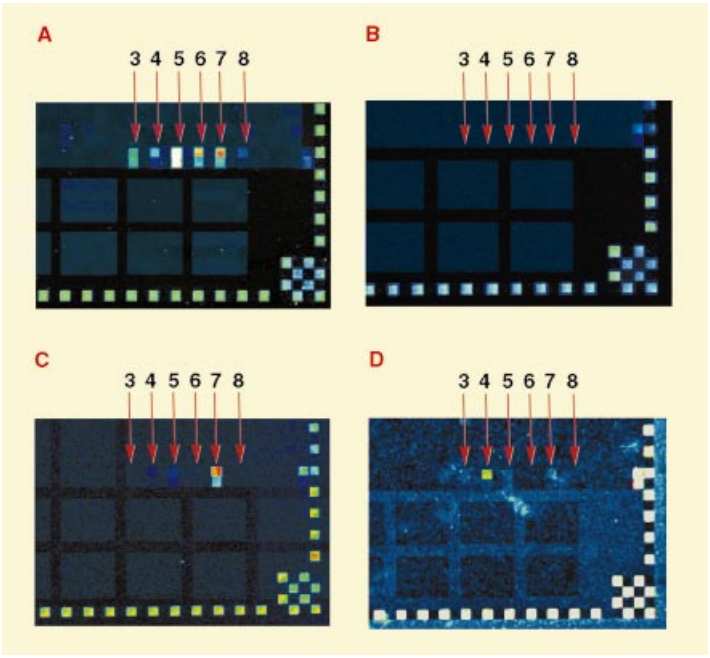


Figure 3. Hybridization of probes **3**–**8**. A) Hybridization of 0.45 pmol (3.0 nM) of each probe; B) control: incubation of **3**–**8** (1.4 μM, 20 μL) for 2 h at pH 5.5,[26] size-exclusion chromatography, hybridization; C) incubation of **3**–**8** (1.4 μM) with cathepsin C (100 μM, 20 μL) for 2 h at pH 5.5,[26] size-exclusion chromatography, hybridization; D) incubation of **3**–**8** (1.4 μM) with cathepsin L (10 μM) for 2 h at pH 5.5,[26] size-exclusion chromatography, hybridization.

[1] a) D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton, E. L. Brown, *Nat. Biotechnol.* **1996**, *14*, 1675; b) J. L. DeRisi, V. R. Lyer, P. O. Brown, *Science* **1997**, *278*, 680.
[2] D. J. Lockhart, E. A. Winzeler, *Nature* **2000**, *405*, 827.
[3] N. L. Anderson, N. G. Anderson, *Electrophoresis* **1998**, *19*, 1853.
[4] D. Figeys, A. Ducret, J. R. Yates III, R. Aebersold, *Nat. Biotechnol.* **1996**, *11*, 1544.
[5] For reviews, see: a) G. L. Corthals, V. C. Wasinger, D. F. Hochstrasser, J. C. Sanchez, *Electrophoresis* **2000**, *21*, 1104; b) S. P. Gygi, G. L. Corthals, Y. Zhang, Y. Rochon, R. Abersold, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9390.
[6] a) E. Abola, P. Kuhn, T. Earnest, R. C. Stevens, *Nat. Struct. Biol.* **2000**, *7*, 973; b) R. C. Stevens, *Curr. Opin. Struct. Biol.* **2000**, *10*, 558.
[7] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg, *Nature* **2000**, *403*, 623.
[8] V. R. Iyer, C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, P. O. Brown, *Nature* **2001**, *409*, 533.
[9] a) H. Zhu, J. F. Klemic, S. Chang, P. Bertone, A. Casamayor, K. G. Klemic, D. Smith, M. Gerstein, M. A. Reed, M. Snyder, *Nat. Genet.* **2000**, *26*, 283; b) P. Arenkov, A. Kukhtin, A. Gemmell, S. Voloshchuk, V. Chupeeva, A. Mirzabekov, *Anal. Biochem.* **2000**, *278*, 123; c) G. MacBeath, S. L. Schreiber, *Science* **2000**, *289*, 2760.
[10] a) P. Goullet, *J. Gen. Microbiol.* **1975**, *87*, 97; b) C. M. Kam, A. S. Abuelyaman, Z. Li, D. Hudig, J. C. Powers, *Bioconjugate Chem.* **1993**, *4*, 560; c) A. S. Abuelyaman, D. Hudig, S. L. Woodward, J. C. Powers, *Bioconjugate Chem.* **1994**, *5*, 400; d) Y. Liu, M. P. Patricelli, B. F. Cravatt, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14694; e) D. Greenbaum, K. F. Medzihradszky, A. Burlingame, M. Bogyo, *Chem. Biol.* **2000**, *8*, 569.
[11] a) D. W. Nicholson, A. Ali, N. A. Thornberry, J. P. Vaillancourt, C. K. Ding, M. Gallant, Y. Gareau, P. R. Griffin, M. Labelle, Y. A. Lazebnik, N. A. Munday, S. M. Raju, M. E. Smulson, T.-T. Yamin, V. L. Yu, D. K. Miller, *Nature* **1995**, *376*, 37; b) B. J. Backes, J. L. Harris, F. Leonetti, C. S. Craik, J. A. Ellman, *Nat. Biotechnol.* **2000**, *18*, 187; c) T. Nazif, M. Bogyo, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2967.
[12] a) N. S. Gray, L. Wodicka, A.-M. W. H. Thunnissen, T. C. Norman, S. Kwon, F. H. Espinoza, D. O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S.-H. Kim, D. J. Lockhart, P. G. Schultz, *Science* **1998**, *281*, 533; b) D. T. Hung, T. F. Jamison, S. L. Schreiber, *Chem. Biol.* **1996**, *3*, 623.
[13] a) S. P. A. Fodor, R. J. Leighton, M. C. Pirrung, L. Stryer, A. T. Lu, D. Solas, *Science* **1991**, *251*, 767; b) C. Y. Cho, E. J. Moran, S. R. Cherry,

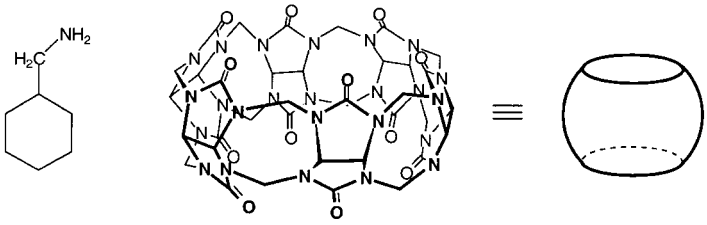
  1433-7851/01/4017-3154 $ 17.50+.50/0

J. C. Stephans, S. P. A. Fodor, C. L. Adams, A. Sundaram, J. W. Jacobs, P. G. Schultz, *Science* **1993**, *261*, 1303.
[14] a) G. MacBeath, S. L. Schreiber, *J. Am. Chem. Soc.* **1999**, *121*, 7967; b) P. J. Hergenrother, K. M. Deprew, S. L. Schreiber, *J. Am. Chem. Soc.* **2000**, *122*, 7849.
[15] a) Á. Furka, F. Sebestyén, M. Asgedom, G. Dibó, *Highlights of Modern Biochemistry*, *Proceedings of the 14th International Congress of Biochemistry* (Prague, Czechoslovakia, 1988), VSP, Ultrecht, **1988**, *13*, pp. 47–47; b) Á. Furka, F. Sebestyén, M. Asgedom, G. Dibó, *Int. J. Pept. Protein Res.* **1991**, *37*, 487; c) S. E. Lam, E. Salmon, M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82; d) R. A. Houghton, P. Clemencia, S. E. Blondelle, J. R. Appel, C. T. Dooley, J. H. Cuervo, *Nature* **1991**, *354*, 84.
[16] The number of final products in a split-pool library is exponentially related to the number of diversity-introducing reactions whereas it is linearly proportional in parallel synthesis.
[17] a) S. Brenner, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381; b) M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett, M. A. Gallop, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10700.
[18] a) M. H. J. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler, W. C. Still, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10922; b) K. C. Nicolaou, X.-Y. Xiao, Z. Parandoosh, A. Senyei, M. P. Nova, *Angew. Chem.* **1995**, *107*, 2476; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2289.
[19] P. E. Neilsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497.
[20] For example, PNA is compatible with 95% TFA used in numerous cleavages.
[21] It is important that the rate of dissociation between the small molecule and its target is slow relative to the time taken for the size-exclusion separation.
[22] a) J. S. Kong, S. Venkatraman, K. Furness, S. Nimkar, T. A. Shepherd, Q. M. Wang, J. Aube, R. P. Hanzlik, *J. Med. Chem.* **1998**, *41*, 2579; b) T. J. Caulfield, S. Patel, J. M. Salvino, L. Liester, R. Labaudiniere, *J. Comb. Chem.* **2000**, *2*, 600; c) C. Walsh, *Tetrahedron* **1982**, *38*, 871.
[23] All the compounds synthesized exhibited satisfactory analytical and functional characteristics.
[24] J. L. Harris, B. J. Backes, F. Leonetti, S. Mahrus, J. A. Ellman, C. S. Craik, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7754.
[25] http://www.affymetrix.com/products/gc_genflex.html. The sequences of the chip's probes are available from Affymetrix.
[26] Buffer: 100 mM NaOAc, pH 5.5; 100 mM NaCl; 1.0 mM EDTA, 0.01% Brij-35; 2.0 mM DTT.
[27] R. J. Riese, P. R. Wolf, D. Bromme, L. R. Natkin, J. A. Villadangos, H. L. Ploegh, H. A. Chapman, *Immunity* **1996**, *4*, 357.
[28] I. M. Berquin, B. F. Sloane, *Adv. Exp. Med. Biol.* **1996**, *389*, 281.
[29] M. J. Bossard, T. A. Tomaszek, S. K. Thompson, B. Y. Amegadzie, C. R. Hanning, C. Jones, J. T. Kurdyla, D. E. McNulty, F. H. Drake, M. Gowen, M. A. Levy, *J. Biol. Chem.* **1996**, *271*, 12517.
[30] T. R. Shearer, M. Azuma, L. L. David, T. Murachi, *Invest. Ophthalmol. Visual Sci.* **1991**, *32*, 533.
[31] J. H. McKerrow, E. Sun, P. J. Rosenthal, J. Bouvier, *Annu. Rev. Microbiol.* **1993**, *47*, 821.

Two Mechanisms of Slow Host–Guest Complexation between Cucurbit[6]uril and Cyclohexylmethylamine: pH-Responsive Supramolecular Kinetics**

Cesar Marquez and Werner M. Nau*

Supramolecular materials capable of performing specific functions are of great current interest.[1–3] Ultimately, their suitability will greatly depend, in particular for the case of prospective supramolecular machines, on the speed by which their operations and functions are performed. A prediction and optimization of this speed requires knowledge of the sequence and the rates by which the elementary supramolecular processes occur, which defines the emerging field of supramolecular kinetics.[4] We have recently introduced a fluorescent probe for measurement of very fast host–guest association processes with cyclodextrins[5] and now report a study of the factors which govern the other extreme of very slow host–guest complexation kinetics between cyclohexylmethylamine (**1**) and the container compound cucurbit[6]uril (CB6). The pH proved crucial for the kinetics and the rate



1 CB6

constants turned out to be unrelated to the thermodynamics of complexation; rather, they are related to the degree of protonation of the guest. The distinct pH dependence of the kinetics of this host–guest complexation process points to two different mechanisms, in which the ingression of the protonated guest is retarded by the formation of an association complex, while the unprotonated guest can enter the cavity directly with a rate constant 20-fold larger. Such a regulation of the supramolecular kinetics between an organic guest and a container-type host molecule through pH is unprecedented.

CB6 is a glycoluril macrocycle with two "crowns" of six ureido-carbonyl groups on both rims suitable of complexing organic ammonium salts through ion–dipole and hydrogen bonding interactions.[6, 7] The small cavity openings of CB6 have been considered as portals, the size of which (about 4 Å)[8] determines whether and how fast complexation occurs. The inner, wider cavity (5.5 Å in width and 6 Å in height)[8] is suitable to accommodate nonpolar organic residues through

[*] Prof. Dr. W. M. Nau, Dipl.-Chem. C. Marquez
Departement Chemie der Universität
Klingelbergstrasse 80, 4056 Basel (Switzerland)
Fax: (+41) 61-267-3855
E-mail: Werner.Nau@unibas.ch

[**] This work was supported by the Swiss National Science Foundation (projects 620-58000 and 4047-057552) within the Swiss National Research Program "Supramolecular Functional Materials".

  1433-7851/01/4017-3155 $ 17.50+.50/0