Pergamon

Bioorganic & Medicinal Chemistry, Vol. 4, No. 5, pp 717–725, 1996

0968-0896/96 $15.00+0.00

S0968-0896(96)00068-5

'Mutational SURF': A Strategy for Improving Lead Compounds Identified from Combinatorial Libraries

Susan M. Freier,[a,*] Danielle A. M. Konings,[b,†] Jacqueline R. Wyatt[a] and David J. Ecker[a]

[a]*ISIS Pharmaceuticals, 2292 Faraday Avenue, Carlsbad, CA 92008, U.S.A.*

[b]*Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, U.S.A.*

Abstract—Synthesis and testing of mixtures of compounds in a combinatorial library offers the potential of much greater throughput than the 'one compound, one well' approach. When mixtures of compounds are screened, however, pooling and deconvolution strategies must be employed to identify the most active compound in the library. The possibility exists that the most active compound will not be identified. We have developed a theoretical model of library deconvolution using the well characterized properties of nucleic acid hybridization to calculate activities of individual molecules in libraries of more than 250,000 compounds. Calculations using this model have been employed to evaluate strategies for pooling and deconvolution. In the presence of errors in synthesis and testing, iterative deconvolution or position scanning sometimes identified a compound with sub-optimal activity. We describe a procedure called 'mutational SURF' in which 'mutants' of the selected compound are individually synthesized and tested. Simulations of mutational SURF using our model libraries suggest that mutational SURF provides an efficient method for improving the activity of lead compounds identified from combinatorial libraries. Copyright © 1996 Elsevier Science Ltd

Introduction

A critical step in drug discovery is identification of a lead compound with the desired activity. Before development as a therapeutic agent, a lead compound will likely be modified to optimize its pharmacological, pharmacokinetic, toxicological, and other properties consistent with the requirements of a successful drug product. Traditionally, lead compounds have been identified either by random screening of natural products or synthetic chemicals or by rational modification of chemicals' with known activity.

Development of automated synthesis techniques has enabled preparation of chemical libraries with extraordinary diversity and unprecedented numbers of novel compounds. These 'combinatorial libraries' provide a new source of compounds for drug discovery.[1–8] Synthesis and testing of mixtures of compounds in a combinatorial library offers the potential of much greater throughput than the 'one compound, one well' approach. When mixtures of compounds are screened, however, pooling and deconvolution strategies must be employed to identify a single, and hopefully most active, compound in the library. An iterative deconvolution strategy which we refer to as SURF (Synthetic Unrandomization of Randomized Fragments)[9,10] can be used to identify a single compound from a mixture (Table 1). Iterative deconvolution strategies have been used to identify peptides which bind tightly to antibodies[11–16] or other protein targets,[15,17,18] oligonucleotides with antiviral activity,[9,10,19,20] and non-oligonucleotide compounds which inhibit LTB_4 and PLA_2.[21] A second technique that can be used to identify a single compound from a mixture is position scanning (Table 1). It differs from SURF in that it is non-iterative and the subsets are overlapping. Position scanning has been used to identify active compounds from peptide[15,22,23] and non-peptide[24,25] libraries.

When mixtures of compounds are tested and a deconvolution procedure is employed, the possibility arises that the most active molecule in the library may not be identified. For example, among the molecules in complex combinatorial libraries, there may be many that have moderate activity and some with activity nearly equal to that of the most active compound. These molecules with sub-optimal activity could impede selection of the most active compound. Alternatively, errors in either synthesis or activity assay may lead to identification of a sub-optimal compound.

To evaluate the importance of molecules with sub-optimal activity on the outcome of SURF deconvolution, we have previously used a model system based on oligonucleotide hybridization to perform computer simulations of deconvolution of combinatorial libraries.[26] We examined two library-target pairs. A library of all RNA 9-mers binding to the target sequence GUGUGGGCA had very few suboptimal binders and the same library binding to UGGGCA had many suboptimal binders. Characteristics of these two

Key words: combinatorial libraries, deconvolution.
†Present address: Department of Microbiology, Southern Illinois University, Carbondale, IL 62901, U.S.A.

Table 1. Deconvolution strategies for chemical libraries of all tetramers composed of the five monomers A–E

Molecules in subset	SURF[a] No. molecules per subset	Most active subset	Molecules in subset	Position scanning[b] No. molecules per subset	Most active subset
XNNN	125	**A**NNN	**X**NNN	125	**A**NNN
AXNN	25	**AC**NN	N**X**NN	125	NC**N**N
ACXN	5	**ACE**N	NN**X**N	125	NN**E**N
ACEX	1	**ACED**	NNN**X**	125	NNN**D**
	Selected molecule:	**ACED**		Selected molecule:	**ACED**

[a]SURF deconvolution begins with synthesis of a non-overlapping set of mixtures by incorporating a unique monomer at a common position of each subset. The subsets are tested separately and the one with greatest activity is identified. A second set of compound mixtures is prepared with each subset containing the fixed monomer showing greatest activity from the previous round. In addition, another position is fixed with each of the unique monomers to give another set of subsets. The complexity of the mixture is reduced and the process is repeated until a unique molecule is identified.

[b]Position scanning is a non-iterative technique which has been used with peptide[15,22,23] and non-peptide[24,25] libraries. At each position in the oligomer, a series of mixtures is synthesized with a different monomer in the fixed position. Each of the mixtures is tested separately and the selected molecule is deduced by selecting the monomer from the most active mixture from each position set. In principal, only a single round of screening is required to define the most active molecule.

systems span those of many other combinatorial libraries.[26] We found that within achievable experimental and library synthesis error, iterative deconvolution methods usually found either the best molecule or one with affinity within five-fold that of the best. Position scanning was less successful with these libraries. Although SURF was reasonably successful, a compound with sub-optimal activity was, none-the-less, sometimes selected. Below we extend the simulations described previously[26] to include a procedure we call 'mutational SURF'. In mutational SURF, each position of a selected molecule is systematically 'mutated' to see if substitution at that position improves activity. If a molecule with improved activity is identified, the process is repeated until no improvement in activity is observed. We demonstrate below that when iterative SURF is followed by mutational SURF the likelihood of selecting a molecule with optimal activity is increased.

Results

Characterization of two molecular landscapes

Oligonucleotide hybridization is the only known molecular binding interaction where calculations based upon experimentally determined parameters can accurately predict the association constants of very large numbers of molecules.[27] We defined the activity of a compound as its association constant for the target (see Experimental). Using hybridization energies it is possible to calculate the activity for several hundred thousand different molecules for a specific target to create a 'molecular landscape'.[28] We created two different landscapes by selecting two RNA targets, a 9-mer and a 6-mer, and a library of all possible 262,144 RNA 9-mers. The energy profiles for these two library target pairs have been previously described in detail.[26] It is important to realize that, once generated, the binding energy profile for each landscape simply represents a series of affinities of molecules for a target which could be a model for any macromolecular interaction. The two landscapes described here had distinctly different profiles which simulate a fairly broad range of binding interactions.

For simplicity, we refer to the activity of the highest affinity molecule in the library as the 'best binder' and any molecule within five-fold of that is called a 'good binder'. Against the 9-mer target there were two best binders (with equal affinity) and only 12 (0.005% of the library) good binders. Almost all of the molecules (98.7%) were less than 10^{-5}-times as active as the two best compounds. Relatively few molecules contributed to activity of the library in this landscape; the activity of the whole library was only eight-times greater than the activity of the single best compound.

In contrast, against the 6-mer target there were many compounds with significant activity. There were 16 molecules with the best affinity and 2414 molecules (0.9% of the library) with good binding activity. Only 37% of the molecules were less than 10^{-5}-times as active as the most active compounds. In this landscape the library was 2000-fold more active than the single most active compound. These two distinct molecular landscapes provided us with test systems to investigate pooling and deconvolution strategies that hopefully will mimic real situations in combinatorial library screening.

Simulations of iterative deconvolution

To simulate iterative deconvolution, the library of 262,144 molecules was divided into four non-overlapping subsets with a single position fixed and an equimolar mix of all four monomers at the other positions. Activities of each subset were determined as described in the experimental. The subset with greatest activity was further divided into four non-overlapping subsets and the procedure was repeated until a unique molecule was selected. As illustrated in Table 1, the division into four subsets during each round is deter-

mined by the choice of fixed position. For libraries synthesized using split bead strategies,[29-31] the last position synthesized is typically fixed in the first round of deconvolution because fewer synthetic steps are required if the fixed position is synthesized last. In principle, however, any order of fixed positions is possible. Our earlier studies demonstrated that order of unrandomization had little effect on the outcome of a SURF experiment.[26] For the present studies we used randomly chosen orders of unrandomization.

In reality, measured activities have experimental errors associated with them. Thus, if two subsets have similar activities, it may not always be possible to pick out the most active subset. To simulate experimental error in the activity assay, a Monte Carlo strategy was used to add a two-fold error to each activity measurement. Figure 1 (squares) plots the results of 500 Monte Carlo simulations of iterative deconvolution for each of the two targets. For the 9-mer target, a best binder was identified 55% of the time and 94% of the time a good binder was selected. For the 6-mer target, a best binder was rarely selected (1% of the time), but 97% of the simulations resulted in identification of a good binder.

Simulations of position scanning

Position scanning[23] (see Table 1) is a non-iterative deconvolution technique where a set of mixtures is synthesized for each position of the oligomer and a single position is fixed in each subset. The most active compound is deduced by selecting the monomer from the most active subset at each position. Simulations of position scanning were also performed with two-fold experimental error in the activity measurement. Figure 1 (circles) plots the results of 500 Monte Carlo simulations of position scanning for each of the two targets. Position scanning was less successful than iterative SURF. For the 9-mer target a good binder was identified only 71% of the time and for the 6-mer target a good binder was identified only 19% of the time.

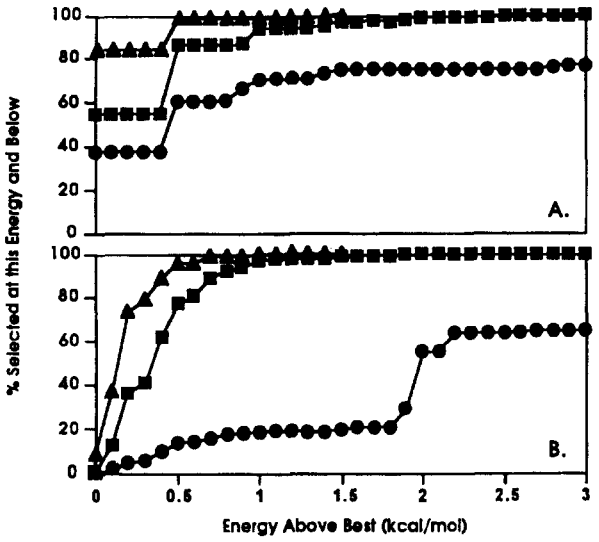

Figure 1. The effect of mutational SURF on the success rate of iterative deconvolution and position scanning. For each target and each deconvolution strategy, 500 simulations were performed with two-fold Monte-Carlo error in the activity of each sample. Due to the two-fold error, different simulations sometimes resulted in selection of a different compound. The percent of simulations resulting in selection of a compound at this energy or better was plotted vs. energy for the library of RNA 9-mers binding to 5′-GUGUGGGCA-3′ (A) or 5′UGGGCA-3 (B). Data are plotted for position scanning without mutational SURF (●), iterative deconvolution without mutational SURF (■), and iterative deconvolution followed by mutational SURF (▲).

Simulations of mutational SURF without experimental error

The results in Figure 1 demonstrate that in the presence of two-fold error in the activity assay the most active molecule (best binder) was frequently not selected. Frequency of selection of a good binder was greater but in the worst case only 19% of the selected molecules were good binders. A strategy which we call 'mutational SURF' was tested to see if it could improve activity of the selected molecule. Mutational SURF begins after identification of a unique molecule or 'starting compound' with detectable activity. In the examples in Table 2 and Figure 1, the starting compound was selected using iterative deconvolution or position scanning but any procedure which identifies an active compound could be used. In mutational SURF, a set of 'mutant' molecules related to the 'starting compound' is synthesized and tested. These mutant molecules include all compounds in which the monomer at a single position is substituted with each different monomer. In our library with four monomers and nine positions, there were 27 (9 positions × 3 mutants at each position) substitution mutants. We also included two cyclically permuted mutants where all of the residues of the linear molecule were shifted one position left or right. Cyclic permutations were included to allow for molecules that bound in different registers. Examples of both substitution and permutation mutations for our library are listed in Table 2. The first round of mutational SURF consists of synthesis and testing of the mutant molecules as individual compounds. If one or more mutants are more active than the starting compound, the most active one becomes the new starting compound and a new set of mutants is synthesized and tested. The process is repeated until no improvement in activity is observed.

Table 2 illustrates mutational SURF using a compound selected during simulations of iterative SURF with two-fold experimental error. This compound, AGCCCGCAC was about 20-fold less active than the most active compound in the library. Although selection of a compound with such poor activity occurred infrequently (<1% of the simulations), it provided a good test of mutational SURF. In the first round of mutational SURF, activities of 29 mutant molecules were compared to the activity of the initial compound and a mutant with 10-fold improved activity was identified. In the second round of mutational SURF a mutant with two-fold improved activity was identified

and in the third round of mutational SURF, no improvement in activity was observed. Thus, synthesis and testing of 87 (3 x 29) individual compounds (0.03% of the library) resulted in identification of the most active compound in the library.

This procedure was repeated on 218 different sub-optimal binders selected from deconvolution of the library against the 9-mer target. These 218 starting compounds were identified during simulations of position scanning with experimental error. In all cases, the most active molecule in the library was identified. On average, 3.6 rounds of mutational SURF were required before no improvement in activity was observed. In the worst case, seven rounds of mutational SURF were required to identify the most active molecule in the library. For 204 of the 218 compounds tested (94%), activity improved at least five-fold during the first round of mutational SURF. The 14 molecules that improved less than five-fold during the first round of mutational SURF were among the most active studied. Activity of these 14 compounds was on average only 4.8-fold less than that of the most active compound. Thus, the first round of mutational SURF resulted in improvement of less than five-fold because there was less room for improvement.

Mutational SURF was also applied to sub-optimal binders selected during simulations of iterative SURF with experimental error performed with the 6-mer target. The most active molecule with an IC_{50} of 40 nM was identified in two of ten examples and in the other examples the molecule identified had an IC_{50} of 47 nM. In these examples, one to three rounds of mutational SURF were required.

Mutational SURF on randomly selected compounds

The success of mutational SURF on sub-optimal compounds selected using standard deconvolution procedures motivated us to simulate mutational SURF on compounds selected randomly from the library. For each library-target pair, 500 molecules were selected at random. Each was subjected to mutational SURF without experimental error. For the 9-mer target, all 500 simulations resulted in selection of a best binder.

Table 2. Example of mutational SURF without experimental error[a]

Round 1		Round 2		Round 3	
Molecule	IC_{50}(pM)	Molecule	IC_{50}(pM)	Molecule	IC_{50}(pM)
Initial molecule		Initial molecule			
AGCCCGCAC	26.9	GCCCGCACA	2.36	**GCCCACACA**	**1.05**
Mutant molecules		Mutant molecules		Mutant molecules	
CGCCCGCAC	22.9	ACCCGCACA	1560	ACCCACACA	691
GGCCCGCAC	26.9	CCCCGCACA	3500	CCCCACACA	1560
UGCCCGCAC	26.9	UCCCGCACA	4120	UCCCACACA	1830
AACCCGCAC	24,500	GACCGCACA	76,400	GACCACACA	33,900
ACCCCGCAC	55,200	GGCCGCACA	106,000	GGCCACACA	46,900
AUCCCGCAC	64,900	GUCCGCACA	361	GUCCACACA	160
AGACCGCAC	1,200,000	GCACGCACA	76,400	GCACACACA	33,900
AGGCCGCAC	1,670,000	GCGCGCACA	106,000	GCGCACACA	46,900
AGUCCGCAC	4,120	GCUCGCACA	307	GCUCACACA	136
AGCACGCAC	871,000	GCCAGCACA	9270	GCCAACACA	12,800
AGCGCGCAC	1,200,000	GCCGGCACA	7880	GCCGACACA	10,900
AGCUCGCAC	3500	GCCUGCACA	99	GCCUACACA	60.6
AGCCAGCAC	106,000	**GCCCACACA**	**1.05**	GCCCCCACA	33,900
AGCCGGCAC	89,800	GCCCCCACA	33,900	GCCCGCACA	2.36
AGCCUGCAC	1120	GCCCUCACA	20,900	GCCCUCACA	20,900
AGCCCACAC	12.0	GCCCGAACA	24,500	GCCCAAACA	15,100
AGCCCCCAC	387,000	GCCCGGACA	20,900	GCCCAGACA	12,800
AGCCCUCAC	238,000	GCCCGUACA	136	GCCCAUACA	60.6
AGCCCGAAC	64,900	GCCCGCCCA	5700	GCCCACCCA	2530
AGCCCGGAC	64,900	GCCCGCGCA	5.31	GCCCACGCA	2.36
AGCCCGUAC	1560	GCCCGCUCA	2980	GCCCACUCA	1320
AGCCCGCCC	4120	GCCCGCAAA	222	GCCCACAAA	98.6
AGCCCGCGC	60.6	GCCCGCAGA	222	GCCCACAGA	98.6
AGCCCGCUC	2150	GCCCGCAUA	51.5	GCCCACAUA	22.9
AGCCCGCAA	160	GCCCGCACC	10.2	GCCCACACC	4.52
AGCCCGCAG	160	GCCCGCACG	2.36	**GCCCACACG**	**1.05**
AGCCCGCAU	222	GCCCGCACU	5.31	GCCCACACU	2.36
GCCCGCACA	**2.36**	CCCGCACAG	4840	CCCACACAG	2150
CAGCCCGCA	587	AGCCCGCAC	26.9	AGCCCACAC	12.0
winseq:	2.36	winseq:	1.05	winseq:	1.05
GCCCGCACA	2.36	GCCCACACA	1.05	GCCCACACA	1.05

[a]This library of 262,144 compounds contained two most active molecules with IC_{50} = 1.05 pM.

On average, 7.3 rounds of mutational SURF were required. For the 6-mer target, 51% of the simulations resulted identified the most active compound, 49% of the simulations identified a compound nearly as active as the best binder. On average, 4.7 rounds of mutational SURF were required.

Simulations of mutational SURF with experimental error

Experimental error in activity measurement could lead to a failure of mutational SURF. Therefore we performed 500 simulations of iterative SURF in the presence of two-fold error. The molecule identified by iterative SURF was then subjected to mutational SURF with two-fold error in the activity measurements. Figure 1 (triangles) plots the results of these simulations. For both targets, mutational SURF significantly improved the success rate. For the 9-mer target, an average of 3.4 rounds of mutational SURF were required before no improvement was observed. For the 6-mer target 14.2 rounds were necessary (Table 3).

We also performed simulations of position scanning in the presence of two-fold error followed by mutational SURF with two-fold error. The results for position scanning followed by mutational SURF were virtually identical to those for iterative deconvolution followed by mutational SURF (Table 3). When random selection preceded mutational SURF with experimental error, the success rate was only slightly lower than that for mutational SURF following iterative deconvolution or position scanning (Table 3).

The relatively large number of rounds of mutational SURF required for the 6-mer target resulted from the high number of compounds with good activity for this target. Typically only three rounds of mutational SURF were required to identify a molecule with activity within two-fold that of the best. The difficulty was that there were 920 compounds with activity within two-fold that of the most active and many of these differed from one another by a single 'mutation'. Thus, even if a best binder was identified, there was reasonable likelihood that the two-fold error would cause a mutant molecule to appear more active than the starting compound and an additional round of mutational SURF would be required. This process continued for an average of 14.2 rounds before no mutant molecule appeared more active than the starting compound. In contrast, for the 9-mer target, there were two best compounds and no others with activity within two-fold that of the best. Consequently, with the 9-mer target, the two-fold error did not substantially increase the number of rounds of mutational SURF required.

To test if the efficiency of mutational SURF could be improved, we performed simulations in which mutational SURF was terminated as soon as no mutant showed an improvement in activity greater than a cutoff based on experimental error. Results for two different cutoff values are listed in Table 3. Cutoff at two-fold resulted in only a slightly lower success rate and a substantial decrease in the number of rounds of mutational SURF required when mutational SURF followed iterative deconvolution or position scanning. Cutoff at five-fold resulted in a significant decrease in the activity of the selected compounds. Incorporation of either cutoff greatly reduced the activity of selected compounds when mutational SURF followed random selection.

Varients of mutational SURF

We also explored a variant of mutational SURF in which not all 29 mutants were prepared and tested

Table 3. Average relative activity of selected molecule and average number of rounds of required for mutational SURF using different strategies for identification of a starting compound[a]

Target: Initial selection procedure	9-mer						6-mer					
	Iterative		Position scanning		Random		Iterative		Position scanning		Random	
	Relative activity[b]	Average rnds[c]	Relative activity[b]	Average rnds[c]	Relative activity[b]	Average rnds[c]	Relative activity[b]	Average rnds[c]	Relative activity[b]	Average rnds[c]	Relative activity[b]	Average rnds[c]
No mutational SURF	0.6		0.05		$<7\times10^{-10}$		0.5		0.007		$<9\times10^{-6}$	
No cutoff[d]	0.9	3.4	0.9	3.7	0.4	10.2	0.7	14.2	0.7	13.9	0.6	16.4
Cutoff at two-fold[d]	0.8	1.8	0.8	2.2	0.05	7.3	0.6	3.9	0.6	4.8	0.4	6.2
Cutoff at five-fold[d]	0.6	1.01	0.5	1.9	0.002	2.3	0.5	1.03	0.5	1.3	0.2	3.5

[a]Mutational SURF followed iterative deconvolution, position scanning or random selection of a compound from the library. Averages are over 500 simulations. Two-fold Monte Carlo error was included in all activity determinations.
[b]Average activity of selected molecule relative to the most active compound in the library.
[c]Average number of rounds of mutational SURF required before the cutoff criterion was met.
[d]Simulations of mutational SURF with no cutoff were continued until no mutant compound was more active than the starting compound. Simulations of mutational SURF with cutoff were stopped when no mutant was more than two-fold (or more than five-fold) more active than the starting compound.

simultaneously. Instead, each mutant was synthesized and tested, one at a time, and as soon as a compound with improved activity was identified, that compound was assigned to be the initial molecule for a new round of mutational SURF. This strategy was as successful as the strategy in which all 29 mutant compounds were tested and the most active was moved on to the next round. This varied strategy required more rounds of mutational SURF, 10.5 for the 9-mer target and 6.5 for the 6-mer target. The average number of compounds prepared each round was 7. This variant of mutational SURF offers the advantage that fewer syntheses are required. For example, for the 9-mer target, the variant procedure required on average 10.5 rounds × 7 molecules per round (73.5) compared to 7.3 rounds × 29 molecules per round (211.7) for the standard procedure. The disadvantage is, however, that each compound would have to be synthesized and tested before it can be determined which compound to prepare next. No efficiency from simultaneous synthesis or testing could be gained.

A second variant that was examined was exclusion of the two cyclical permutation mutants. We performed mutational SURF on 500 compounds selected at random from each of the two landscapes with testing of only the 27 substitution mutants each round. When the cyclical permutations were omitted, mutational SURF was much less successful. For the 9-mer target, activity of the selected molecule was, on average, 335-fold less active than the best binder. For the 6-mer target, this factor was only three-fold, but was still much worse activity than was obtained when all 29 mutants were included. Thus, access to molecules of different register was required for the success of mutational SURF for these library-target pairs.

Discussion

Is mutational SURF worthwhile?

Our results suggest mutational SURF is a very effective way of correcting mistakes that may have been made during deconvolution. In the absence of experimental error, the most active molecule, or one with nearly equal activity, was always selected. Even in the presence of two-fold error in the activity assay, mutational SURF improved the success rate of iterative SURF (Fig. 1).

Mutational SURF does, however, require synthesis and testing of several individual molecules. The question therefore arises; "is mutational SURF worth the effort?" The answer, of course, depends on how much improvement in activity is gained by mutational SURF and how many rounds of synthesis and testing mutational SURF requires. This point is illustrated by two examples taken from the Monte Carlo simulations of iterative SURF with two-fold error followed by mutational SURF with two-fold error. In the first example, from the 9-mer target, two rounds of mutational SURF resulted in an 11-fold improvement in activity. In this example, mutational SURF was probably worth the effort; substantial improvement in activity was achieved with relatively little synthetic effort. In the second example, from the 6-mer target, 57 rounds of mutational SURF actually resulted in a 1.6-fold *reduction* in activity. In this example, a large synthetic effort was not profitable; the molecule identified after mutational SURF was *less* active than the molecule identified by iterative SURF without mutational SURF.

Although it is impossible to know how much improvement mutational SURF will yield and how much effort will be required until it has been completed, a single round of mutational SURF can provide a strong hint as to how efficient further rounds of mutational SURF will be. In the simulations of mutational SURF performed on 218 sub-optimal binders from the library binding to the 9-mer target, the greatest improvement in activity was usually during round 1. Each subsequent round generally gave less improvement than the previous round. Thus, it may be most efficient to perform mutational SURF until the round to round improvement is less than some cutoff. A cutoff equal to the experimental error would take advantage of the large improvement in activity sometimes observed in the early rounds, but would avoid the situation described above for the 6-mer target where the large number of molecules with activity within two-fold that of the best resulted in 14.2 rounds of mutational SURF, many of which were ineffective. Simulations in Table 3 suggest that termination of mutational SURF when the improvement in activity is less than the assay error can substantially reduce the number of rounds of mutational SURF required without substantially reducing activity of the selected compound.

Is deconvolution necessary?

Mutational SURF was originally designed to correct mistakes that could occur with iterative deconvolution or position scanning. Simulations on molecules chosen at random from the library demonstrated that mutational SURF was reasonably successful when the initial deconvolution process was omitted and the starting compound was chosen at random (Table 3).

The success of mutational SURF on randomly selected molecules forces us to pose the question, "is the preliminary round of iterative deconvolution or position scanning necessary?" Beginning with a randomly selected molecule typically required more rounds of mutational SURF than beginning with a molecule selected by iterative SURF or position scanning. In addition, the success rate was reduced greatly if a cutoff was employed so the increased number of rounds associated with no cutoff would be necessary. Mutational SURF requires synthesis of many individual molecules and iterative deconvolution

or position scanning frequently requires synthesis using the split bead technique. Therefore, synthetic requirements associated with additional rounds of mutational SURF may be less than those required for iterative deconvolution or position scanning. Thus, if only synthetic requirements are considered, it appears beginning with a randomly selected compound could be more efficient than beginning with a molecule selected by iterative deconvolution or position scanning.

The difficulty of beginning with a randomly selected compound is that often the selected compound possesses no detectable activity so identification of a mutant with improved activity may be difficult. The likelihood of a randomly selected compound possessing detectable activity depends, naturally, on the detection limit of the assay. It is interesting, however, to compare the activities of randomly selected molecules to activities for the first round of iterative deconvolution or position scanning. If iterative deconvolution or position scanning is to be successful, it would require a detection limit for IC_{50} of 25 nM for the 9-mer target and 7 μM for the 6-mer target. For the 9-mer target, 99.4% of the molecules in the library were less active than this limit. For the 6-mer target, this figure was 96.3%. Thus, it is possible to have sufficient sensitivity for iterative deconvolution or position scanning to be successful while more than 95% of randomly selected molecules have undetectable activity. The difficulty of finding a molecule randomly with which to start mutational SURF suggests that iterative deconvolution and position scanning are still valuable 'front ends' for mutational SURF.

Can these results be applied to chemical combinatorial libraries?

Our simulations have demonstrated that mutational SURF is an effective strategy for finding the best compound in a library following iterative SURF or position scanning. However, what is the relevance of this technique to 'real' combinatorial libraries of small molecules binding to targets such as enzymes or membrane receptors? Any set of interactions between molecules in a library and a target can be reduced to a table of binding energies of each individual molecule for the target, which creates a molecular landscape. The two landscapes used in our calculations had binding energy ranges similar to those of many known drug-receptor complexes. The best binder against the 9-mer target had an IC_{50} of 1 pM and the 6-mer target's best binder had an IC_{50} of 40 nM.

However, in a real test of mixtures the absolute values of the best and good binders are only important in that they must generate a detectable signal. What is important in deconvolution is the number of tight binders and their energies relative to each other and the rest of the molecules in the library. For the 9-mer target, there were two best binders and only 12 (0.005%) good binders, while the 6-mer target had 16 best binders and 2414 (0.9%) good binders. The landscape for any 'real' library of small molecules will vary for each library and target, but our simulated landscapes are likely to bracket many real situations.

One factor that can be compared between experiments and simulations is the activities of the mixtures compared to the activities of the individual molecules deconvoluted from them. If only one compound contributes to the activity of a mixture, then the 'expected' improvement in activity is equal to the complexity of the mixture. We calculate the ratio of 'expected' improvement to actual improvement observed upon deconvolution to a single compound. We call the ratio of these activities the suboptimal binding factor (SBF), which is a measure of how much more active the mixture is than it would be if only the final selected compound contributed to its activity.[26] The results that we obtained in deconvolutions of our model library compared favorably with experimental deconvolutions published by many groups with respect to the suboptimal binding factor[26,28] and are also similar to more recently published deconvolutions,[32,33] including deconvolutions of small molecule chemical libraries.[24,34] This suggests that our simulated landscapes are similar to some real ones.

The aspect of our simulated landscape that may be less like small molecule chemical libraries is the relative weight of a single substitution on the overall binding energy of the molecule. In our examples, mutations at some sites affected activity as little as two-fold and at other sites the effect was as much as 4.5×10^4-fold. Although there may be small molecules where the improvement gained from substitution at a single site is greater than 4.5×10^4 or where the range of effects of substitutions at different sites is greater than that for our model, the wide range of effects observed in our model suggest it will be relevant to many real systems.

A potential limitation of mutational SURF is the fact that, although all possible single mutations are made and tested at once, improvements that require two simultaneous mutations to improve activity will not be detected. Simultaneous double mutations that improve the activity of molecules have been observed in experimental directed molecular evolution experiments.[35] Thus mutational SURF in its present form will improve the molecule with iterative parallel sampling of all single mutant space, but not identify coupled double mutants.

In summary, mutational SURF provides a technique that can correct mistakes that may be made when deconvolution techniques are used to identify an active lead from a combinatorial library. This procedure used compounds synthesized and tested one at a time avoiding potential errors associated with synthesis and testing of mixtures. Our simulations suggested that the success rate of deconvolution techniques are signifi-

cantly improved when they are followed by mutational SURF. In the presence of experimental error, the efficiency of mutational SURF was reduced when there were many compounds with activity near that of the most active. Stopping mutational SURF when the improvement drops below a cutoff equal to the experimental error should prevent these inefficiencies. Finally, the results presented above provide further demonstration of the utility of our model system for rapid, inexpensive, evaluation of strategies for identification of an active compound from a combinatorial library.

Experimental

The targets for the 9-mer and 6-mer landscapes were, respectively, 5'GUGUGGGCA-3' and 5'UGGGCA-3'. Methods for calculation of free-energies for library molecules binding to the RNA targets have been described previously.[26] We define the activity of a molecule as the reciprocal of the concentration needed to bind 50% of the target molecules,

$$\text{activity} = 1/IC_{50} = K_A = \exp(-\Delta G^\circ_{37}/(RT)),$$

where K_A is the association constant for the molecule, $-\Delta G^\circ_{37}$ is the binding free energy, R is the gas constant (0.001987 kcal/mol/K) and T is temperature (310.15 K).

Activities of each subset were calculated as the average activity of the compounds in the subset.[24,26] This calculation assumes no synergism or antagonism between compounds within a subset. Effects of experimental error were simulated by assuming the observed activities had a log normal distribution about the true activity. Observed activities for each subset were generated using standard Monte Carlo techniques.[36] Mutational SURF was simulated by calculating energies of the appropriate mutant molecules and adding a Monte Carlo error if appropriate.

Acknowledgments

The authors thank Dr. Stuart Kauffman for helpful discussions regarding characterization of molecular landscapes. Computing resources were generously provided by the University of Colorado.

References

1. Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1233.

2. Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. *J. Med. Chem.* **1994**, *37*, 1385.

3. Ecker, D. J.; Crooke, S. T. *BioTechnology* **1995**, *13*, 351.

4. Terrett, N. K.; Gardner, M.; Gordon, D. W.; Kobylecki, R. J.; Steele, J. *Tetrahedron* **1995**, *51*, 8135.

5. Eichler, J.; Houghten, R. A. *Molecular Med. Today* **1995**, 174.

6. Janda, K. D. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10779.

7. Houghten, R. A. *Methods: A Companion to Methods in Enzymology* **1994**, *6*, 354.

8. Pinilla, C.; Appel, J.; Blondelle, S.; Dooley, C.; Dorner, B.; Eichler, J.; Ostresh, J.; Houghten, R. A. *Biopolymers* 1995, *37*, 221.

9. Ecker, D. J.; Vickers, T. A.; Hanecak, R.; Driver, V.; Anderson, K. *Nucl. Acids Res.* **1993**, *21*, 1853.

10. Wyatt, J. R.; Vickers, T. A.; Roberson, J. L.; Buckheit, R. W., Jr.; Klimkait, T.; DeBaets, E.; Davis, P. W.; Rayner, B.; Imbach, J. L.; Ecker, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1356.

11. Geysen, H. M.; Rodda, S. J.; Mason, T. J. *Mol. Immunol.* **1986**, *23*, 709.

12. Blake, J.; Litzi-Davis, L. *Bioconjugate Chem.* **1992**, *3*, 510.

13. Geysen, H. M.; Rodda, S. J.; Mason, T. J.; Tribbick, G.; Schoofs, P. G. *J. Immunol. Methods* **1987**, *102*, 259.

14. Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84.

15. Houghten, R. A.; Appel, J. R.; Blondelle, S. E.; Cuervo, J. H.; Dooley, C. T.; Pinilla, C. *BioTechniques* **1992**, *13*, 412.

16. Edmundson, A. B.; Harris, D. L.; Fan, Z.-C.; Guddat, L. W.; Schley, B. T.; Hanson, B. L.; Tribbick, G.; Geysen, H. M. *Proteins* **1993**, *16*, 246.

17. Owens, R. A.; Gesellchen, P. D.; Houchins, B. J.; DiMarchi, R. D. *Biochem. Biophys. Res. Commun.* **1991**, *181*, 402.

18. Eichler, J.; Houghten, R. A. *Biochemistry* **1993**, *32*, 11035.

19. Buckheit, R. W., Jr.; Roberson, J. L.; Lackman-Smith, C.; Wyatt, J. R.; Vickers, T. A.; Ecker, D. J. *AIDS Res. Hum. Retroviruses* **1994**, *10*, 1497.

20. Ecker, D. J.; Wyatt, J. R.; Vickers, T. *Nucleosides Nucleotides* **1995**, *14*, 1117.

21. Davis, P. W.; Vickers, T. A.; Wilson-Lingardo, L.; Wyatt, J. R.; Guinosso, C. J.; Sanghvi, Y. S.; DeBaets, E. A.; Acevedo, O. L.; Cook, P. D.; Ecker, D. J. *J. Med. Chem.* **1995**, *38*, 4363.

22. Dooley, C. T.; Houghten, R. A. *Life Sci.* **1993**, *52*, 1509.

23. Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *BioTechniques* **1992**, *13*, 901.

24. Pirrung, M. C.; Chen, J. *J. Am. Chem. Soc.* **1995**, *117*, 1240.

25. Smith, P. W.; Lai, J. Y. Q.; Whittington, A. R.; Cox, B.; Houston, J. G.; Stylli, C. H.; Banks, M. N.; Tiller, P. R. *Bioorg. Med. Chem. Lett.* 1994, *4*, 2821.

26. Freier, S. M.; Konings, D. A. M.; Wyatt, J. R.; Ecker, D. J. *J. Med. Chem.* **1995**, *38*, 344.

27. Turner, D. H.; Sugimoto, N.; Jaeger, J. A.; Longfellow, C. E.; Freier, S. M.; Kierzek, R. *Cold. Spring. Harb. Symp. Quant. Biol.* **1987**, *52*, 123.

28. Kauffman, S. A. *The Origins of Order, Self-Organization and Selection in Evolution*; Oxford University: Oxford, 1993.

29. Furka, A.; Sebestyn, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487.

30. Lebl, M.; Krchnák, V.; Sepetov, N. F.; Seligmann, B.; Strop, P.; Felder, S.; Lam, K. S. *Biopolymers* **1995**, *37*, 177.

31. Stanková, M.; Issakova, O.; Sepetov, N. F.; Krchnák, V.; Lam, K. S.; Lebl, M. *Drug Dev. Res.* **1994**, *33*, 146.

32. Campbell, D. A.; Bermak, J. C.; Burkoth, T. S.; Patel, D. V. *J. Am. Chem. Soc.* **1995**, *117*, 5381.

33. Terrett, N. K.; Bojanic, D.; Brown, D.; Bungay, P. J.; Gardner, M.; Gordon, D. W.; Mayers, C. J.; Steele, J. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 917.

34. Carell, T.; Wintner, E. A.; Rebek, J., Jr. *Angew. Chem. (Engl)* **1994**, *33*, 2061.

35. Beaudry, A. A.; Joyce, G. F. *Science* **1992**, *257*, 635.

36. Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: San Francisco, 1992; pp 75–95.

(*Received in U.S.A.* 7 *September* 1995; *accepted* 13 *December* 1995)