Biotechnology et alia **7:**1-6

Directed Evolution and Solid Phase Enzyme Screening

Edward J. Bylina*, William J. Coleman, Christina L. Grek, Mary M. Yang and Douglas C. Youvan
KAIROS Scientific Inc., Bldg. 62, 3350 Scott Blvd., Santa Clara, CA 95054 USA

ABSTRACT

A new digital imaging spectrophotometer and a series of colorimetric solid phase assays have been developed to screen bacterial libraries expressing mutagenized enzymes undergoing directed evolution. This high-throughput solid-phase assay system (known as 'Kcat Technology') can detect less than a 20% difference in enzyme rates within microcolonies grown at a nearly confluent density of 500 colonies per cm^2 on an assay disk. Each microcolony is analyzed simultaneously at single-pixel resolution (1.5 megapixels; 75 micron/pixel), requiring less than 100 nanoliters of substrate per measurement, a 1000-fold reduction over conventional liquid phase assays. Here we report the successful identification of variants of Agrobacterium β-glucosidase (Abg) – a glycosidase with broad substrate specificity that favors cleavage of glucosides over galactosides – by simultaneously assaying two different substrates tagged with spectrally distinct chromogenic reporters. In this study, we have used combinatorial cassette mutagenesis based on the phylogeny of homologous glycosidase genes and known point mutants in Abg to generate a recombinant library of over 10 million enzyme variants involving partial randomization of 13 amino acid residues near the active site. Different enzyme variants that show epimeric specificity – either for the hydrolysis of glucoside or galactoside – were identified by solid-phase screening. This solid phase screening technology is compatible with a variety of mutagenesis methods including error-prone mutagenesis, sequential random mutagenesis, and gene shuffling. However, using combinatorial cassette mutagenesis, this greater screening capacity enables one to survey large sequence spaces in which many residues are changed simultaneously. Isolating optimized enzymes may be accomplished through recursive cycles of cassette mutagenesis using a GA-based technique, Recursive Ensemble Mutagenesis (REM) – emulating the natural evolution of genes by point mutation and recombination. This technology should lead to the isolation of new enzyme activities that are useful in the synthesis of various substances, including specialty chemicals and pharmaceuticals.

Keywords: Imaging, spectroscopy, kinetics, microcolonies, REM, mutagenesis, DNA, software

1. INTRODUCTION

The enzymes that are currently available for use in synthetic chemistry have evolved over millions of years to be efficient and selective catalysts for the chemical reactions taking place in living systems. However, many potential industrial applications for these catalysts involve substrates, organic solvents and other reaction conditions that were never encountered in nature. Protein engineering can be used to change the properties of these natural catalysts to suit the needs of synthetic chemistry and manufacturing. Much progress has been made by both industrial and academic laboratories in demonstrating how enzymes can be tailored for optimal performance in industrial applications by directing their evolution *in vitro*. By carefully controlling *in vitro* mutation efficiencies and screening for enhanced catalytic properties over multiple generations, researchers have developed new enzymes that are hundreds of times more active than the natural enzymes in chemical process environments.

Despite these successes, further exploitation of this enzyme engineering is severely limited by the current need to employ time-consuming liquid-phase assays for screening mutant libraries. Many companies are attempting to develop new technologies that will overcome such screening bottlenecks. Recent advances in protein engineering, genomics and combinatorial chemistry have prompted major pharmaceutical and biotechnology firms to accelerate their efforts by developing automated high-throughput liquid-phase screening systems. Conventional 96-well plates (100-200 μl per well) are typically used in liquid-phase screening. Improvements in assay miniaturization would allow acceleration in the rate of screening, reduction in the cost per assay, and conservation of compounds that are either expensive or difficult to synthesize and purify. The high-throughput screening system presented in this paper utilizes a solid-phase assay disk which may approach the limit to which these assays may be miniaturized on a 2-dimensional surface.

---

*Correspondence: Email: ebylina@kairos-scientific.com; http://www.kairos-scientific.com; Telephone: 408-567-0400;
Fax: 408-567-0440

Feasibility of the Kcat Technology was demonstrated using the model enzyme system ('Abg') from *Agrobacterium faecalis* that hydrolyzes β-glucosides. The catalytic mechanism of the wild-type protein and several active site variants has been studied in detail (Trimbur et al., 1992; Wang et al., 1995). The enzyme is quite nonspecific, catalyzing the hydrolysis of substituted glucosides, galactosides, xylosides, fucosides, and arabinosides (Kempton & Withers, 1992). Given the versatility of this enzyme, it appeared to be an ideal candidate for directed evolution experiments. In the Abg experiments described below, we demonstrate the ability to differentiate Abg mutants differing in substrate specificity for indolyl derivatives of galactoside and glucoside. We further demonstrate the capability of characterizing these enzyme variants in solid-phase kinetics assays.

Abg activity was screened using the type of reaction shown in Figures 1 and 2. Hydrolytic cleavage of the glycosidic bond at the 3-position of the indolyl derivative generates the protonated (hydroxyl) form of indoxyl, which deprotonates and tautomerizes. In the presence of oxygen, two indoxyl molecules are then spontaneously oxidized and the C-2 carbon is deprotonated, which causes them to rapidly dimerize. The final product is an intensely colored indigo dye, which precipitates out of solution, and therefore does not diffuse away from the lysed bacterial microcolony. This is important, since absorbance is monitored as a function of time in the solid phase assay.

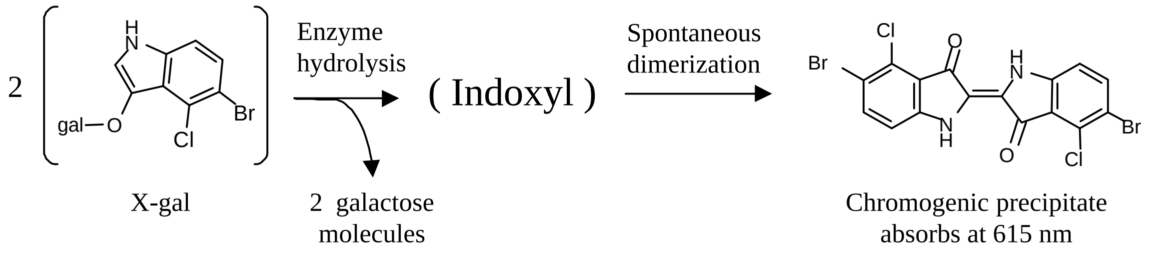


Figure 1. Enzymatic hydrolysis of an indolyl derivative used in an exemplary colorimetric solid phase assay. The classical X-gal reaction commonly used in molecular biology is shown here. The hydrolysis of other indolyl derivatives (see the next figure) follows the same pathway to form a variety of colored precipitates.

Indolyl derivatives	R$_1$	R$_2$	R$_3$	R$_4$	Indigo product λmax (nm)
Red	H	H	H	Cl	540
Magenta	H	H	Br	Cl	565
Iodo	H	H	I	H	575
X	H	Cl	Br	H	615
Green	CH$_3$	H	H	H	665
Y	H	H	H	H	680

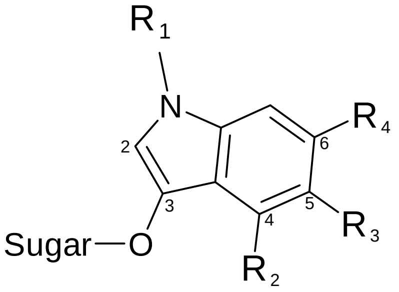


Figure 2. Indolyl derivatives and the colors of their indigo products. A variety of different colors can be produced depending on the identity of the R groups. The sugar moiety shown here can be replaced by other compounds to make the molecule a substrate for a variety of enzymes. For example, replacing the sugar with a carboxylic acid generates ester substrates for lipases and esterases.

2. METHODS AND RESULTS

The active site nucleophile of Abg is the glutamic acid residue at position 358 (Withers et al., 1990). An Abg "active site phylogeny" was constructed in this region of the protein as follows. A BLAST search of the non-redundant GenBank CDS (Altschul et al., 1997) identified 20 related sequences using a search sequence comprising amino acids 350-385 of Abg. Protein sequence differences between Abg and these sequences were tabulated. Additional related protein sequences and sequence changes of active Abg mutants (Trimbur et al., 1992) were also incorporated . Figure 3 shows this phylogeny for Abg residues 361 to 373. This phylogenetic information was used to design a combinatorial cassette library. Experiments can also be based on random mutagenesis using NNN, NNK, or NNS triplets for the zeroth iteration of REM, as previously described (Arkin and Youvan, 1992a,b; Delagrave et al., 1993; Delagrave and Youvan, 1993). In this paper, phylogeny has been used to 'bootstrap' the method – as described by Goldman et al., 1992.

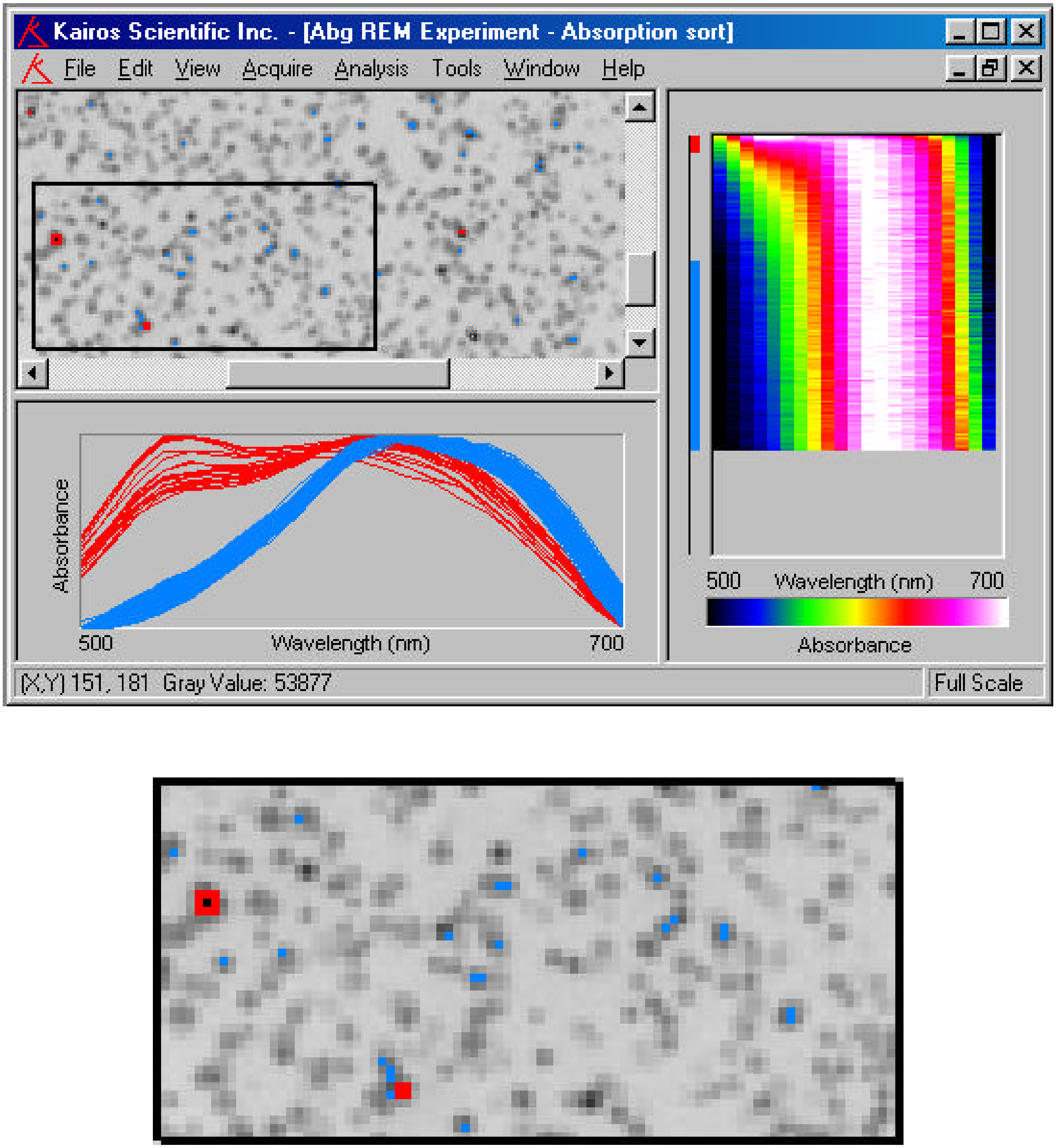
A brief outline of the REM library construction follows: An Abg expression plasmid was constructed by cloning the Abg structural gene into the pQE70 expression vector (QIAgen). The Ala389 GCC codon was changed to GCA, resulting in the introduction of a BglII restriction site at this position. A doped oligonucleotide corresponding to the 90-bp NarI-BglII fragment (encoding Gly360 to Ala389 in Abg) was synthesized based on phylogenetic data entered into the 'CyberDope' computer program (KAIROS Scientific). A REM cassette was synthesized by PCR amplification of this doped oligonucleotide:

5’CACCGAAAACGGCGYCKSCWHCRAKRWKRDGNTTSWGRAWDGCRRGVTCMAWGACCAGCCGCGTCTCGAT
TATTACGCCGAACACCTCGGCATCGTCGCAGATCTCATCC 3’

The BglII-BsaHI digested REM cassette was ligated into the BglII-NarI digested expression vector and transformed into M15[pREP4]. The resulting library was screened using the Kcat instrument (KAIROS Scientific). Information on the Kcat instrument and the CyberDope program can be found at: www.kairos-scientific.com. More detailed methodology for performing Kcat Technology can be found in US patent 5,914,245 (Bylina et al., 1999) at: www.USPTO.gov. The software and algorithms used to sort pixels (or features) in the data shown below are described in another manuscript in these proceedings (Yang et al., 2000).

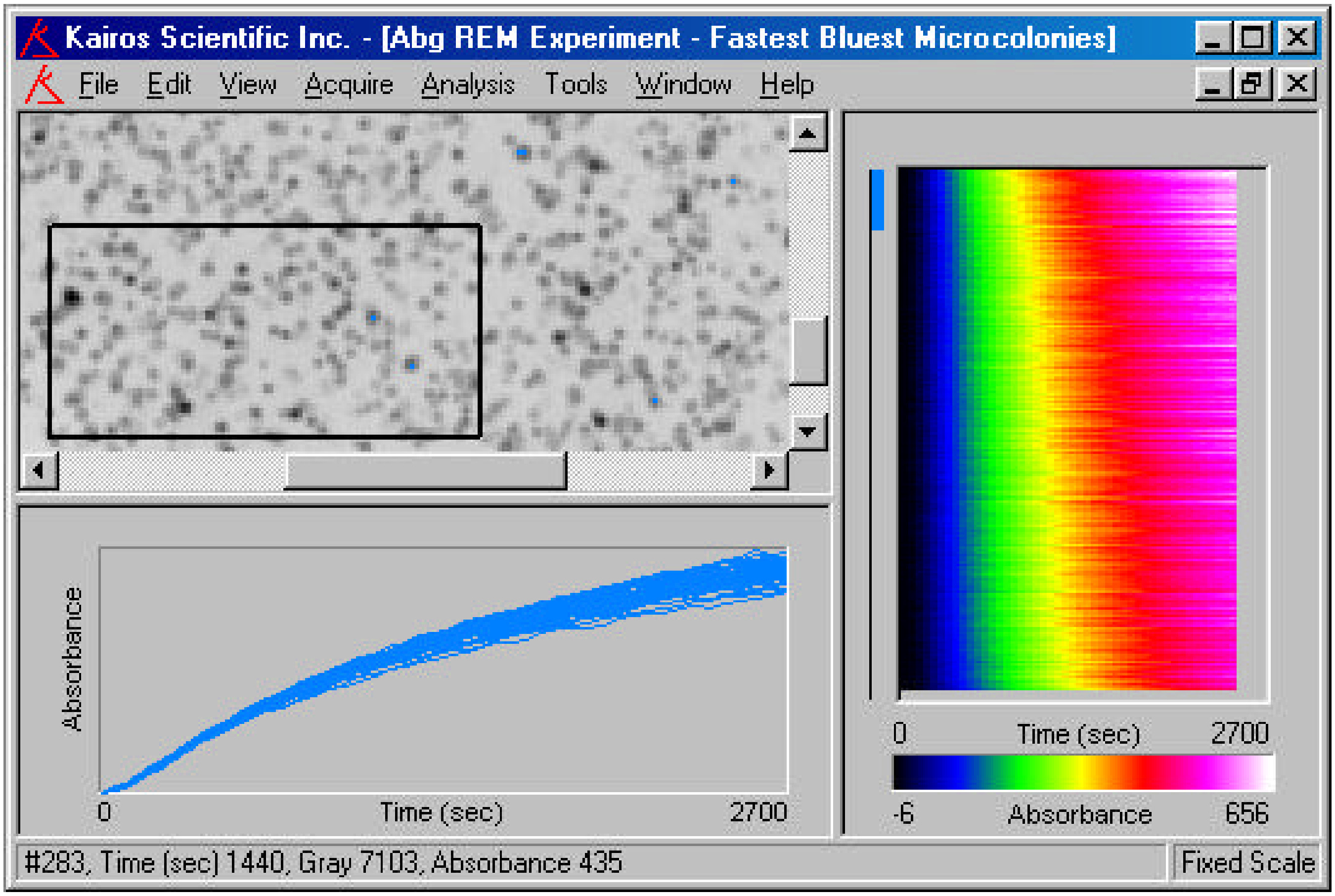
														N			
								Y	A					F			
							H		S	E	P		I	S		A	
				R		M	R	T	T	S	D		K	Q		Y	
				F	D	E	F	F	Q	T	N	G	N	V		E	
				V	**S**	**S**	**K**	**I**	**V**	**F**	I	**K**	H	T		**K**	
				I	**G**	**I**	**E**	**K**	**E**	**L**	**Q**	**E**	**C**	**K**	**L**	**Q**	
				L	**A**	**F**	**D**	**D**	**K**	**I**	**V**	**D**	**S**	**R**	**I**	**H**	
	358																374
	E	N	G	**A**	**C**	**Y**	**N**	**M**	**G**	**V**	**E**	**N**	**G**	**E**	**V**	**N**	D
wt:	GAA	AAC	GGC	GCC	TGC	TAC	AAT	ATG	GGC	GTC	GAA	AAC	GGC	GAG	GTC	AAT	GAC
REM:	GAA	AAC	GGC	GYC	KSC	WHC	RAK	RWK	RDG	NTT	SWG	RAW	DGC	RRG	VTC	MAW	GAC
				A	C	Y	N	M	G	V	E	N	G	E	V	N	
				V	A	F	D	D	K	I	V	D	S	R	I	H	
					G	I	E	K	E	L	Q	E	C	K	L	Q	
					S	S	K	I	V	F	L	K		G		K	
						T		V	R								
						N		E	M								
								N									

Figure 3. The combinatorial region of the cassette used to mutagenize the Abg gene. The wild-type (wt) sequence from residue 358 to 374 is given above the mixed nucleotide sequence for the combinatorial cassette. The entire sequence of this doped oligonucleotide can be found in the text. This oligonucleotide was used for the first cycle of Recursive Ensemble Mutagenesis (REM). Phylogenetic data for glucosidases is given above the DNA sequences. Amino acid residues (in bold type face) were input into the CyberDope program (including wt), and the residues encoded by the REM cassette are given below the DNA sequences. In order to reduce combinatorial complexity to ~ 10^8, some of the residues in the phylogeny were omitted (not in bold type face).



(X,Y) 151, 181 Gray Value: 53877

Figure 4. Graphical User Interface of the Kcat instrument (above) and an enlarged area of the assay disk (below) showing the spectral analysis of Abg mutagenesis on Red-gal and X-glu. A high resolution 24-bit RGB version of this image can be found on-line in *Biotechnology et alia* at http://www.et-al.com/searchable/abstracts/SolidPhase1.html. Backcoloring is used to highlight pixels in the image (upper left) and absorption spectra (lower left) for every pixel on the assay disk. Sorting was based on the 540 nm : 610 nm ratio. Pixels within microcolonies that favor the substrate Red-gal are sorted to the top of the contour plot and backcolored in red. Pixels within microcolonies that favor the substrate X-glu are sorted to the bottom of the contour plot and backcolored in blue. Pixels within microcolonies that favor the substrate X-glu are sorted further by catalytic velocity in Figure 5, below.



#283, Time (sec) 1440, Gray 7103, Absorbance 435

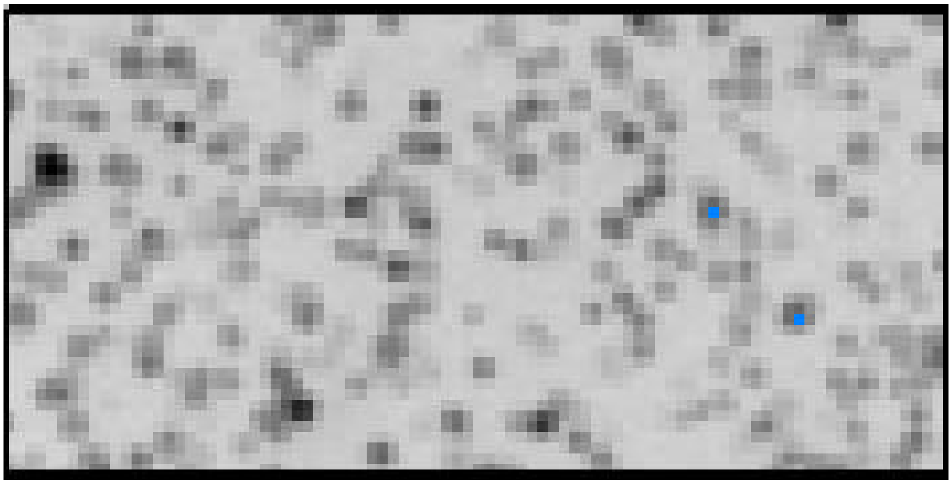


Figure 5. Graphical User Interface of the Kcat instrument (above) and an enlarged area of the assay disk (below) showing the kinetic analysis of Abg mutants that favor the substrate X-glu. A high resolution 24-bit RGB version of this image can be found on-line in *Biotechnology et alia* at http://www.et-al.com/searchable/abstracts/SolidPhase1.html. Highlighted pixels (blue) within microcolonies indicate enzyme variants with the highest velocity after sorting for X-glu specificity.

3. SUMMARY

We have successfully developed a high-throughput solid-phase assay system that can be used to identify interesting enzyme variants generated by mutagenesis. This Kcat Technology can be used to distinguish small differences in both the kinetic and spectral properties of enzyme variants expressed in microcolonies. The Abg variants identified during the screening described in this paper have been isolated and are currently being characterized.

ACKNOWLEDGEMENTS

This work was supported by NIH R44GM55470, NIH R43GM60073, and by KAIROS IR&D Funds.

REFERENCES

1. S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” *Nucleic Acids Res.* **25**, pp.3389-3402, 1997.
2. A.P. Arkin, and , D.C. Youvan, "Optimizing Nucleotide Mixtures to Encode Specific Subsets of Amino Acids for Semi-Random Mutagenesis," *Bio/Technology*, **10**, pp. 297-300, 1992.
3. A.P. Arkin, and , D.C. Youvan, "An Algorithm for Protein Engineering: Simulations of Recursive Ensemble Mutagenesis," *PNAS*, **89**, pp. 7811-7815, 1992.
4. E.J. Bylina, W.J. Coleman, M.R. Dilworth, C.M. Silva, M.M. Yang and D. C. Youvan, “Solid phase enzyme kinetics screening in microcolonies,” U.S. Patent 5914245, 1999.
5. S. Delagrave, E.R. Goldman, and D.C. Youvan, "Recursive Ensemble Mutagenesis," *Protein Engineering*, **6**, pp. 327-331, 1993.
6. S. Delagrave and D.C. Youvan, "Searching Sequence Space to Engineer Proteins: Exponential Ensemble Mutagenesis," *Bio/Technology*, **11**, pp. 1548-1552, 1993.
7. E.R. Goldman, and D.C. Youvan, D.C, "An Algorithmically Optimized Combinatorial Library Screened by Digital Imaging Spectroscopy," *Bio/Technology*, **10**, pp. 1557-1561, 1992.
8. J.B. Kempton and S.G. Withers,” Mechanism of Agrobacterium beta-glucosidase: kinetic studies,” *Biochemistry* **31**, 9961-9969, 1992.
9. D.E.Trimbur, R.A.J. Warren, and S.G. Withers, “Region-direcetd mutagenesis of residues surrounding the active site nucleophile in β-glucosidase from Agrobacterium faecalis,”*J. Biol. Chem.* **267**, pp. 10248-10251, 1992.
10. Q. Wang, D. Trimbur, R.Graham R.A. Warren, and S.G. Withers, “Identification of the acid/base catalyst in Agrobacterium faecalis beta-glucosidase by kinetic analysis of mutants,” Biochemistry **34**, 14554-14562, 1995.
11. S.G. Withers, R.A.J. Warren, I.P. Street, K. Rupitz, J.B. Kempton, and R. Aebersold, J. Am. Chem. Soc. **112**, pp. 5887-5889, 1990.
12. M. M.Yang, Edward J. Bylina, W. J. Coleman, Michael R. Dilworth, S. J. Robles, M. A. Tanner, and D. C. Youvan, "Graphical User Interfaces for Single-Pixel Spectroscopy," *SPIE* (this volume).
13. D.C. Youvan, "Searching Sequence Space. *Bio/Technology*," **13**, pp. 722-723, 1995.