

## ***In vitro* Evolution of Beta-glucuronidase into a Beta-galactosidase Proceeds Through Non-specific Intermediates**

**Ichiro Matsumura and Andrew D. Ellington\***

*Institute of Cellular and Molecular Biology, ICMB A4800/MBB 3.424, University of Texas, 26th and Speedway Austin, TX 78712, USA*

The *Escherichia coli* beta-glucuronidase (GUS) was evolved *in vitro* to catalyze the hydrolysis of a beta-galactoside substrate 500 times more efficiently ( $k_{\text{cat}}/K_m$ ) than the wild-type, with a 52 million-fold inversion in specificity. The amino acid substitutions that recurred among 32 clones isolated in three rounds of DNA shuffling and screening were mapped to the active site. The functional consequences of these mutations were investigated by introducing them individually or in combination into otherwise wild-type *gusA* genes. The kinetic behavior of the purified mutant proteins in reactions with a series of substrate analogues show that four mutations account for the changes in substrate specificity, and that they are synergistic. An evolutionary intermediate, unlike the wild-type and evolved forms, exhibits broadened specificity for substrates dissimilar to either glucuronides or galactosides. These results are consistent with the "patchwork" hypothesis, which postulates that modern enzymes diverged from ancestors with broad specificity.

© 2001 Academic Press

**Keywords:** beta-glucuronidase; beta-galactosidase; substrate specificity; *in vitro* evolution; DNA shuffling

\*Corresponding author

### **Introduction**

The rules that govern the evolution of enzyme substrate specificity remain incompletely understood. The preponderance of apparently neutral mutations in molecular evolution (Kimura, 1983) makes it difficult to identify changes that affect protein function through sequence comparisons alone. Site-directed (Chen *et al.*, 1996; Harris & Craik, 1998; Hedstrom *et al.*, 1992; Onuffer & Kirsch, 1995) and cassette mutagenesis studies (Altamirano *et al.*, 2000; el Hawrani *et al.*, 1996; Graham *et al.*, 1993; Palzkill & Botstein, 1992) can sometimes generate novel catalysts as efficient and specific as natural enzymes. These "rational" and "semi-rational" approaches to protein re-design are predicated upon a correct understanding of

the relationships between protein structure and function.

Proteins can be "improved" without structural information by randomly mutating whole genes; the resulting library can be expressed and screened for variants exhibiting altered specificity (Forney *et al.*, 1989; Joo *et al.*, 1999; May *et al.*, 2000; Stemmer, 1994b; Yano *et al.*, 1998; Zacco & Gherardi, 1999; Zhang *et al.*, 1997). So far, enzyme variants generated by whole gene random mutagenesis have generally proven less specific (May *et al.*, 2000; Yano *et al.*, 1998) or less catalytically active (Forney *et al.*, 1989; Zhang *et al.*, 1997) than natural enzymes. Random mutagenesis is also useful because it produces structure-function information that can be utilized to further improve a protein (Miyazaki & Arnold, 1999; Oue *et al.*, 1999). Furthermore, it allows us to approximate and study the early stages of adaptation at the molecular level. In particular, we are interested in whether enzymes directly evolve new substrate specificities, or whether intermediates with more general recognition abilities first arise.

We chose a model system that would allow us to study the structural and functional consequences of adaptive molecular evolution. The *Escherichia coli* beta-galactosidase (*lacZ*) and beta-glucuronidase

Present address: I. Matsumura, Department of Biochemistry, Emory University School of Medicine, Rollins Research Center, Room 4119, Atlanta, GA 30322, USA.

Abbreviations used: GUS, beta-glucuronidase; X-gal, 5-bromo-4-chloro-3-indolyl-beta-D-galactoside; pNP-*para*-nitrophenyl-beta-D-; oNP-, *ortho*-nitrophenyl-beta-D-.

E-mail address of the corresponding author: andy.ellington@mail.utexas.edu

dase (*gusA*) genes diverged from an ancient common ancestor (Henrissat, 1991), and their gene products catalyze the hydrolysis of similar glycoside substrates. Beta-galactosides and beta-glucuronides differ only at the C4 (hydroxyl group up or down) and C5 (hydroxymethyl or carboxyl) positions. Beta-galactosidase, however, has no detectable activity in reactions with beta-glucuronides. GUS exhibits weak beta-galactosidase activity that is six orders of magnitude lower ( $k_{cat}/K_M$ ) than its beta-glucuronidase activity (see below). The putative catalytic domains of the two enzymes are only 25% identical in amino acid sequence, and the nucleotide sequences could not be aligned by the GAP algorithm (GCG, Madison, WI). The structural determinants of substrate specificity are therefore difficult to predict through comparative approaches. We sought to identify them by evolving the specificity of GUS *in vitro* and dissected the structural and functional changes that occurred.

## Results

### *In vitro* evolution

We were able to detect the beta-galactosidase activity of wild-type GUS in a *lacZ*-deficient *E. coli* colony by using a strong over-expression system (Matsumura *et al.*, 1999) and saturating concentrations of the histochemical substrate, X-gal. Detection of the desired activity in the ancestral form is generally a prerequisite for *in vitro* evolution by whole gene random mutagenesis because mutations in a library of limited diversity are unlikely to enhance the activity by more than a factor of 2 (Moore & Arnold, 1996). Random mutations were introduced into the wild-type *gusA* gene *via* mutagenic PCR (Cadwell & Joyce, 1992; Zacco *et al.*, 1996), and the resultant *gusA* libraries were subcloned into the expression vector pET28a(+), and transformed into *E. coli* strain, DH5ΔLAC(DE3).

In the first round of screening, 7000 transformants were propagated on LB agar plates containing 25 μg/ml kanamycin, adsorbed onto nitrocellulose filters and transferred to similar plates containing the inducer, 0.5 mM IPTG, and 4.5 mM X-gal. After 24 hours of induction, we isolated 36 colonies that exhibited more beta-galactosidase activity than control colonies expressing the wild-type GUS. After each round of screening, the selected clones were individually restreaked onto LB-kanamycin plates, propagated, adsorbed to nitrocellulose filters and induced on LB-kanamycin/IPTG/X-gal plates. This enabled us to confirm the phenotypes, and to unambiguously identify the most active beta-galactosidases from each round. After the first round, 17 of the clones were reproducibly more active than the wild-type GUS.

The *gusA* variants isolated in the first round were amplified by PCR, pooled and randomly

mutated, and recombined by DNA shuffling (Stemmer, 1994a); this causes the mutations to re-segregate, potentially uniting beneficial mutations and excluding neutral and deleterious ones (Zhang *et al.*, 1997). The library of randomly recombined alleles was subcloned back into the expression vector, and screened for clones exhibiting further enhancement of beta-galactosidase activity. In the second round, 20,000 transformed colonies were screened, and ten clones that reproducibly turned blue within 2.5 hours were isolated. These were randomly recombined and subcloned, and 7500 colonies were screened. Five colonies that reproducibly turned blue within an hour of induction were isolated. Additional rounds of shuffling and screening did not yield variants with any further increases in fitness.

### Sequencing

The 32 *gusA* variants isolated in the three rounds of evolution were sequenced (Table 1). Sixteen of the 17 round 1 clones were mutated at one of four sites: D508, T509, S557 or N566. We observed a pronounced transition bias (>80%), which suggested that further screening would not have yielded additional beneficial mutations. The T509A, S557P and N566S substitutions were unified in a single round of DNA shuffling and screening (Table 1, clones 2.1 and 2.8), and were almost fixed by the third round (Table 1, clones 3.1-3.4). In contrast, the most common round 1 mutation, D508G, was extinct by the third round. The K568Q substitution did not occur in the first round but was fixed by the third round. These patterns suggest that epistatic interactions occurred during *in vitro* evolution, which we confirmed for K568Q through site-directed mutagenesis.

The tertiary structure of *E. coli* GUS has not been determined, so the substitutions were mapped onto the crystal structure of the human GUS (Jain *et al.*, 1996). This homologue is ~50% identical in amino acid sequence and has the same substrate specificity, so it is reasonable to assume that the alpha carbon trace is conserved. The recurring substitutions at positions 508, 509, 557, 566 and 568 mapped to two loops near the putative catalytic residues (Islam *et al.*, 1999) (Figure 1). These loops cannot be superimposed on the analogous loops of the *E. coli* beta-galactosidase structure (Jacobson *et al.*, 1994), so we cannot say whether the substitutions we identified are structurally convergent. Any of these changes could potentially enhance the blue colony phenotype by a mechanism unrelated to substrate specificity, such as increased protein expression, so we dissected their effects by site-directed mutagenesis.

### Characterization

We introduced single substitutions that occurred (T509A, S557P, N566S) in the first round into otherwise wild-type *gusA* genes by recombinant

**Table 1.** Recurring sequence substitutions

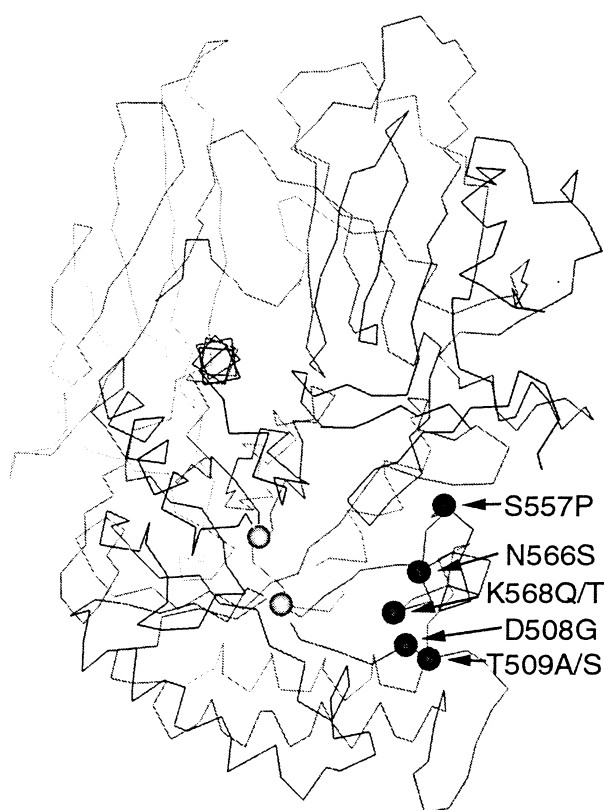
Wild-type	No. aa sub 0	No. sil	77 K	508 D	509 T	557 S	566 N	568 K	604 Stop
1.1	1	1	-	G	-	-	-	-	-
1.2	1	0	-	G	-	-	-	-	-
1.3	2	1	-	G	-	-	-	-	-
1.4	1	0	-	G	-	-	-	-	-
1.5	2	0	-	G	-	-	-	-	-
1.6	3	1	-	-	S	-	-	-	-
1.7	4	0	-	-	A	-	-	-	-
1.8	3	0	E	-	-	P	-	-	-
1.9	3	0	-	-	-	P	-	-	-
1.10	1	2	-	-	-	P	-	-	-
1.11	3	2	-	-	-	P	-	-	-
1.12	1	0	-	-	-	-	S	-	-
1.13	2	1	-	-	-	-	S	-	-
1.14	3	0	-	-	-	-	S	-	-
1.15	2	0	-	-	-	-	S	-	-
1.16	2	3	-	-	-	-	S	-	-
1.17	5	0	-	-	-	-	-	-	-
2.1	9	1	-	-	A	P	S	-	W
2.2	5	3	-	-	S	P	-	Q	-
2.3	5	4	-	-	A	P	-	Q	-
2.4	6	6	-	G	-	P	S	-	-
2.5	3	4	-	-	-	P	-	T	-
2.6	5	1	-	-	A	P	-	Q	W
2.7	4	1	T	-	-	P	-	T	-
2.8	4	2	-	-	A	P	S	-	-
2.9	5	1	T	-	-	P	-	T	-
2.10	7	4	T	-	A	P	-	Q	W
3.1	9	4	-	-	A	P	S	Q	W
3.2	9	4	-	-	A	P	S	Q	W
3.3	7	6	T	-	A	P	S	Q	-
3.4	7	1	-	-	A	P	-	Q	-
3.5	9	4	-	-	-	P	-	Q	-

The *gusA* gene was randomly mutated, and the resulting library was expressed and screened for beta-galactosidase activity as described in the text. The 32 *gusA* variants isolated in three rounds of screening were sequenced (one strand, using four primers) using the Perkin Elmer-Applied Biosystems protocol (Foster City, CA). The inferred amino acid changes at the seven positions that recurred five or more times during the course of *in vitro* evolution are listed. Other mutations are classified as amino acid changing or silent. All mutations are listed in the JMB webpage. In the round 2 and 3 variants, the stop codon was replaced with a tryptophan residue, which also added the plasmid-encoded LQNSSVVDKLAALAEHHHHHHH to their carboxyl termini.

DNA methods. The K578Q mutant, which did not occur until the second round, was also made. In a similar way, since clones 2.1 and 3.1 (Table 1) were the most functional representatives isolated in the second and third rounds of screening, respectively, we introduced the T509A/S557P/N566S (round 2), T509A/D531E/S557P/N566S (round 2) and T509A/S557P/N566S/K568Q (from round 3) substitutions into a wild-type *gusA* background. Each of these site-directed single and multiple mutants, as well as the wild-type and clones 2.1 and 3.1 (Table 1), were expressed and purified. Each GUS variant was reacted with varying concentrations of the chromogenic beta-galactoside substrate, pNP-galactoside, and five other analogues. The results of all kinetics experiments are reported in terms of  $k_{cat}/K_m$ , or as ratios of  $k_{cat}/K_m$  values. The variants that contained only substitutions in the active-site loops (508-568) always behaved identically with the original variants containing additional sequence substitutions (data not shown); mutations outside the active-site loops apparently do not significantly affect the kinetic properties of the enzymes. Furthermore, the T509A/S557P/N566S

variant behaved identically with T509A/D531E/S557P/N566S, showing that D531E has no functional effect.

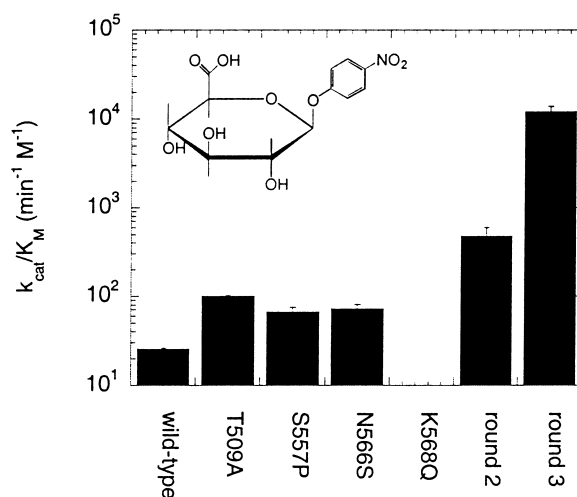
As expected, the specific beta-galactosidase activities of the GUS variants progressively improved throughout the course of the selection. The T509A, S557P, and N566S substitutions in isolation each enhance  $k_{cat}/K_m$  two- to fourfold (Figure 2). The K568Q substitution is deleterious in isolation, but increases catalytic efficiency by a factor of 25 in the context of T509A/S557P/N566S. This shows that T509A, S557P, N566S and K568Q are synergistic with respect to catalytic efficiency. Synergism is rare in artificially evolved systems (Zaccolo & Gherardi, 1999) because step-wise selection schemes, including our own, systematically select against mutations that do not enhance fitness in isolation. It is commonplace in wild-type enzymes, such as in the D32/H64/S221 "catalytic triad" of *Bacillus amyloliquefaciens* subtilisin (Carter & Wells, 1988). The second-order rate constant of the T509A/S557P/N566S/K568Q GUS variant (round 3 loop) is  $1.2 \times 10^4 \text{ min}^{-1} \text{ M}^{-1}$  in reactions with pNP-galactoside, 500-fold greater than that of



**Figure 1.** Homology mapping of amino acid substitutions. The alpha carbon trace of a monomer of human GUS is shown (Jain *et al.*, 1996), but the residues are numbered according to the *E. coli* protein. Putative catalytic residues, E413 and E504 (Islam *et al.*, 1999), are represented as white balls. The sites that were mutated six or more times among the 32 clones isolated in three rounds of *in vitro* evolution (Table 1) are indicated as black balls.

the wild-type (Figure 2). The improvement is similar to that of the evolved *ebgA* beta-galactosidase in reactions with lactose (Hall, 1999), but the  $k_{cat}/K_m$  value remains 10,000-fold worse than the comparable value of the wild-type *lacZ* enzyme (Zhang *et al.*, 1997).

We studied the effects of the beneficial mutations upon interactions with the C4 and C5 substituents by reacting the T509A/S557P/N566S/K568Q GUS (round 3) variant with a series of substrate analogues (Figure 3). The wild-type enzyme recognizes beta-glucuronides (C5 carboxylate, C4 hydroxyl "down") and the enzyme variants were screened for their ability to catalyze hydrolysis of a beta-galactoside analogue (C5 hydroxymethyl, C4 hydroxyl "up"). The wild-type and evolved variants were reacted with the *p*-nitrophenyl derivatives of these as well as those of beta-galacturonide (C5 carboxylate, C4 hydroxyl up) and beta-glucoside (C5 hydroxymethyl, C4 hydroxyl down). The 2,900,000-fold preference of the wild-type for beta-glucuronides was inverted to a 18-fold preference

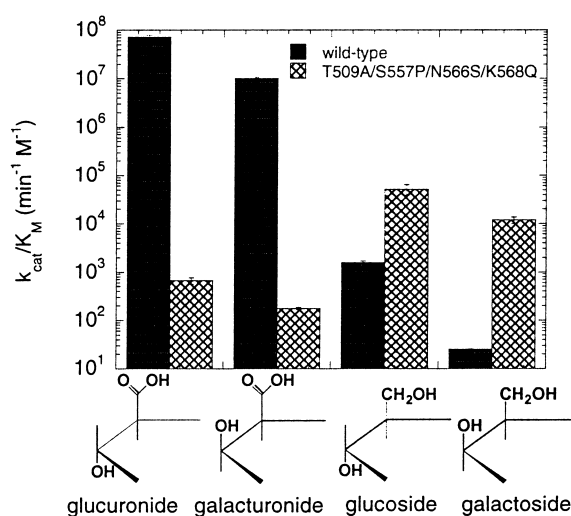


**Figure 2.** Second-order rate constants of GUS variants in reactions with *para*-nitrophenyl galactoside. Each GUS variant was expressed and purified, and reacted with varying concentrations of pNP-galactoside (shown in inset). The formation of the *para*-nitrophenol product was followed in a spectrophotometer, and the  $k_{cat}/K_m$  values were determined by fitting the data to the Michaelis-Menten equation as described in Materials and Methods.

for beta-galactosides (Figure 3). The magnitude of this change exceeds those of previous studies in which whole genes were randomly mutated.

Most of the adaptation occurred in residues that interact with the C5 substituent. The T509A/S557P/N566S/K568Q GUS variant is 33-fold more reactive with pNP-beta-glucoside than the wild-type (Figure 3), even though GUS had not encountered this substrate during the course of *in vitro* evolution. The stereospecificity of GUS for the C4 hydroxyl group diminished as the enzyme evolved. The wild-type enzyme favors glucosides (C4 hydroxyl down) over galactosides (C4 hydroxyl up) by a factor of 62, whereas the T509A/S557P/N566S/K568Q does so by only a factor of 4.3. The decrease in specificity was less pronounced when glucuronide and galacturonide utilization was compared, demonstrating that the enzyme's interactions with the substrate's C4 and C5 substituents are interdependent.

The specificities of the evolved variants for the C5 position clearly decreased during the first two rounds of selection, as assessed by examining ratios of  $k_{cat}/K_m$  values. While the wild-type enzyme preferred the glucuronide over the glucoside analogue by a factor of 46,000, the round 2 variant (T509A/D531E/S557P/N566S) preferred the glucuronide by a factor of only 750 (Figure 4). The round 2 variant appeared to be an enzymatic "generalist", in that it retained parental catalytic activity in reactions with pNP-glucuronide and pNP-galacturonide, but was significantly more active than the wild-type in reactions with pNP-glucoside and



**Figure 3.** Second-order rate constants of the wild-type and round 3 GUS in reactions with various substrates. The wild-type (black) and T509A/S557P/N566S/K568Q GUS (hatched) variants were expressed and purified, and reacted with varying concentrations of pNP-glucuronide, pNP-galacturonide, pNP-glucoside and pNP-galactoside (C4 and C5 substituents shown below the x-axis). The  $k_{cat}/K_M$  values were calculated as described in the legend to Figure 2.

pNP-galactoside. It is most impressive that the round 2 variant catalyzed the hydrolysis of two saccharides with novel C5 substituents, pNP-fucoside (C5 methyl group) and oNP-galactoside-6-phosphate (C5 hydroxymethyl phosphate). The wild-type enzyme exhibited no detectable activity on either substrate.

While the *in vitro* evolution of substrate specificity initially increased the range of substrates that could be utilized, further selection narrowed specificity. In the most evolved third round clone, reactivity with pNP-glucuronide was considerably reduced and no detectable activity on oNP-galactoside-6-phosphate remained (Figure 4). Interestingly, the evolved variants look as though they are moving towards the specificity profile of the *lacZ* beta-galactosidase. The natural enzyme also cannot utilize galactoside 6-phosphate analogues, but exhibits 30-fold greater catalytic efficiency in reactions with fucosides than our artificially evolved variant (Zhang *et al.*, 1997).

## Discussion

### Model of the wild-type complex

We inferred the structure of the wild-type enzyme-substrate complex by re-examining the crystal structure of the human GUS (Jain *et al.*, 1996) in light of our kinetic results. The mutations that affected substrate specificity, T509A, S557P, N566S and K568Q, occurred in a region that is well conserved among the five sequenced wild-

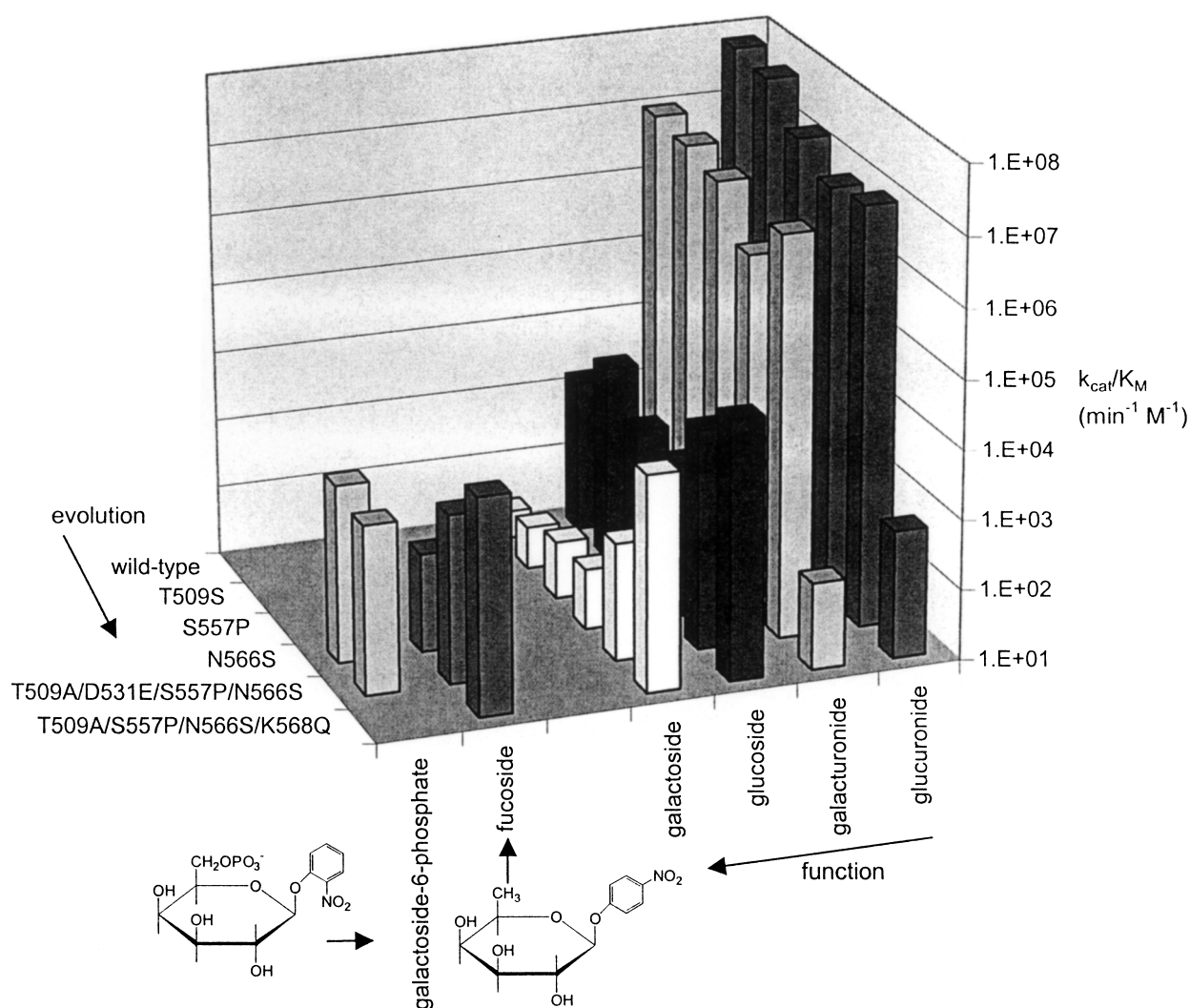
type *gusA* genes (D'Amore *et al.*, 1988; Nishimura *et al.*, 1986; Oshima *et al.*, 1987; Ray *et al.*, 1998; Schlaman *et al.*, 1994). We fit a beta-glucuronic acid into the active-site of the wild-type human GUS without otherwise altering the crystal structure (Figure 5). The model predicts that the putative catalytic residues, E413 and E504 (Islam *et al.*, 1999), are within 3 Å of the glycosidic bond oxygen. Eight intermolecular hydrogen bonds form between the substrate and seven conserved GUS residues D163, Y468, E504, Y549, R562, N566 and K568. The C5 carboxylate participates in six hydrogen bonds, and the C4 participates in one other. This would explain the high specificity of the wild-type enzyme for beta-glucuronide substrates.

### Limits to *in vitro* evolution

We did not observe any additional improvements in the fourth round of *in vitro* evolution, and have considered the possible obstacles to further evolution. First, the T509A/S557P/N566S/K568Q *gusA*-pET28 (round 3 clone) may be at the upper limit of the dynamic range of the X-gal colony assay, since this construct turns colonies blue almost as fast as the much more catalytically efficient *lacZ*-pET28 control. This problem could potentially be solved by decreasing the protein expression level to compensate for the increased activity. Second, it is possible that the frequency of beneficial mutations in local sequence space decreased during the course of evolution. Mutations that destabilize the conformation of the active-site likely occur at a higher frequency than those that introduce novel interactions with the substrate. This obstacle could potentially be overcome by increasing the number of colonies screened, employing a less biased method for random mutagenesis, or by focusing random mutagenesis on the likely active-site residues, as shown in Figure 5.

### Mechanism of adaptation

Other attempts to evolve enzyme specificity *in vitro* have generated variants with broadened specificity (Graham *et al.*, 1993; Yano *et al.*, 1998; Zhang *et al.*, 1997). Our results confirm that the conversion of natural "specialists" into generalists requires relatively few mutations. Such mutations apparently partially disassemble protein architectures optimized for a particular function. For example, the T509A and S557P mutations might destabilize the active-site loop of GUS (Figure 5), thereby decreasing the specificity of the binding pocket. Further improvements in catalytic efficiency might require synergistic mutations that collectively introduce new interactions with the substrate and stabilize the new, catalytically productive active-site conformation. Since step-wise screening for function reduces the frequency of mutations that are selectively neutral in a given generation, many directed evolution experiments



**Figure 4.** Second-order rate constants of GUS variants in reactions with various glycoside substrates. Purified enzymes were reacted with varying concentrations of each substrate, which vary at the C4 positions (hydroxyl group up or down) and C5 positions (carboxylate, hydroxymethyl, methyl, or hydroxymethyl phosphate) as shown in Figure 3 or below the graph. The remainder of each substrate is identical with pNP-galactoside (see Figure 2, inset), except for oNP-galactoside-6-phosphate, which has an *ortho*-nitrophenol aglycone leaving group.

may concomitantly reduce the likelihood that synergistic mutations will arise. Methods such as family shuffling (Cramer *et al.*, 1998) that incorporate large numbers of neutral mutations into directed evolution experiments at the outset may more readily find unanticipated, synergistic interactions.

### Patchwork hypothesis

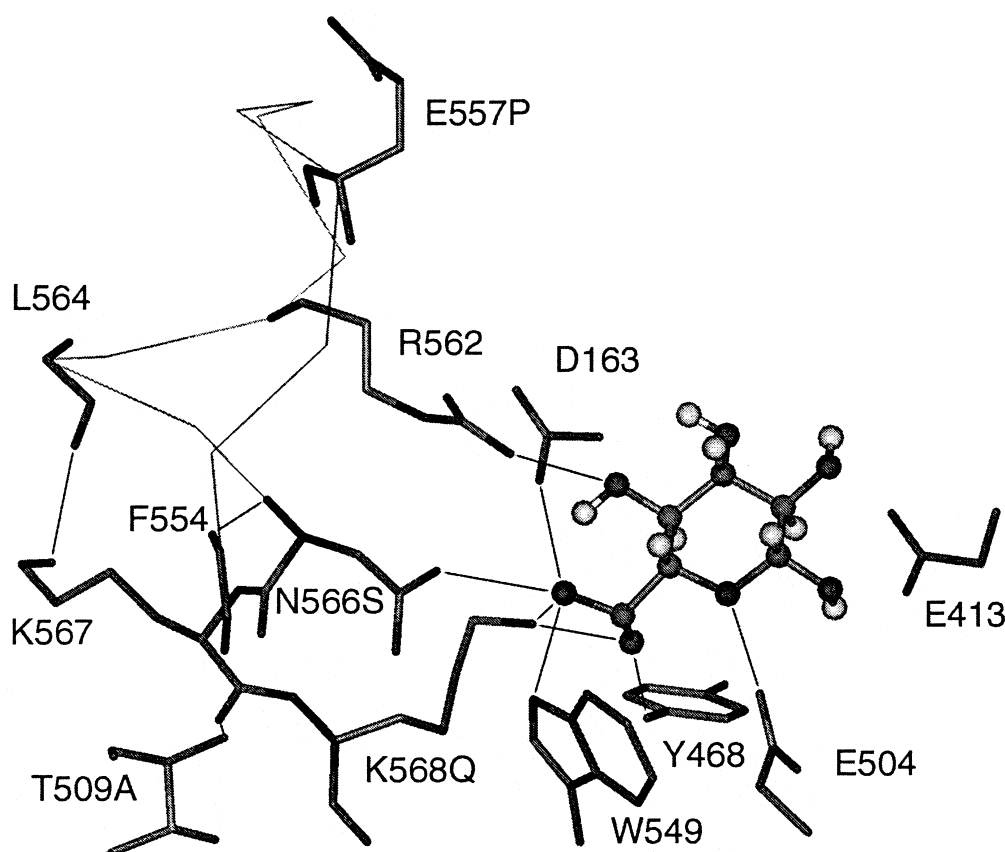
On an evolutionary time-scale, however, enzymes could fluctuate between specialist and generalist states in response to changing selective pressures. The "patchwork" hypothesis (Jensen, 1976) postulates that "primitive enzymes possessed a very broad substrate specificity", and that gene duplication led to "the luxury of increased specialization and ... improved metabolic efficiency." Others have proposed similar hypotheses (Waley, 1969; Ycas, 1974). We propose that in more recent

times, modern enzymes may adapt to recognize new substrates by passing through a functional intermediate similar to their ancestral states. This hypothesis will be further tested through further *in vitro* evolution experiments and characterization of evolutionary intermediates.

## Materials and Methods

### Materials

The lambda DE3 lysogenization kit, which was used to convert the *E. coli* strain DH5ΔLAC (Ruether & Sauer, 1989) into a DE3 lysogen for T7 RNA polymerase expression, and the *lacZ*-pET28 control vector (induction control E) were from Novagen (Madison, WI). The chromogenic substrate analogues were from Sigma Chemicals. The cloning of *gusA* into pET28a+, and the sources of the other reagents, were previously described (Matsumura *et al.*, 1999).



**Figure 5.** Model of complex between the wild-type human GUS and beta-glucuronic acid. The structure of glucuronate was reconstructed as a ball and stick model using the Biopolymer module of Insight II (Molecular Simulations, San Diego, CA), and overlaid on the same scale onto the unmodified crystal structure of human GUS (Jain *et al.*, 1996). The residues are labeled with the number of the homologous *E. coli* residue. All amino acids shown are conserved, except for E557 and L564, which are Asn and Gly residues, respectively in the *E. coli* enzyme. The arrows between residues indicate hydrogen bonding interactions; the predicted intermolecular bonding distances are all less than 3 Å.

### Library construction

For the first round of screening, random mutations were introduced into the GUS gene *via* mutagenic PCR that included either manganese (Cadwell & Joyce, 1992) or the nucleotide analogue, 8-oxo-dGTP (Zaccolo *et al.*, 1996). The mutagenic amplification reactions included amplification primers (100 nM 5'-AGATCTCGATCC-CGCGAAATTAATACGA-3' and 5'-CGGGCTTTGTTAG-CAGCCGGATCTC-3'), which annealed to the pET28a(+) vector outside of the boundaries of the *gusA* insert, 60 mM Tris-HCl (pH 8.5), 15 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2.5 mM MgCl<sub>2</sub>, 0.2 mM dNTPs, plus either (0.1 mM dCTP, 0.1 mM dTTP, 0.6 mM MgCl<sub>2</sub>, 0.0625 mM MnCl<sub>2</sub>) or (0.01 mM 8-oxo-dGTP), and were thermally cycled 25 times between 94 °C for 30 seconds and 72 °C for two minutes. The two libraries were separately subcloned into pET28a(+) and transformed into DH5ΔLAC(DE3), but turned out to be very similar: 73% of the library of manganese-induced GUS mutants and 65% of the 8-oxo-dGTP induced mutants failed to turn colonies green when expressed in the presence of the chromogenic GUS substrate 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide (X-gluc). In the first round of screening, eight clones were isolated from manganese library and nine from the 8-oxo-dGTP library. These collectively contained 41 tran-

sitions and ten transversions, an average of 2.9(±1.4) mutations/allele.

### Protein purification

Fragments of *gusA* containing one or more mutations of interest were amplified by PCR and subcloned into otherwise wild-type *gusA* genes. Each recombinant *gusA* variant was separately subcloned behind sequences encoding an N-terminal six histidine-tag in pET28a(+), transformed into DH5ΔLAC(DE3)/pLysS cells. Addition of the tag did not affect the phenotypes of any of the clones when they were induced on plates containing 0.5 mM IPTG and 4.5 mM 5-bromo-4-chloro-3-indolyl-beta-D-galactoside (X-gal). The transformants were propagated to mid-log phase in liquid LB culture (25 μg/ml kanamycin, 34 μg/ml chloramphenicol), induced with 0.5 mM IPTG, purified to homogeneity (as determined by subsequent SDS-PAGE) by nickel chelate affinity chromatography, dialyzed in GUS buffer (50 mM sodium phosphate (pH 7.0), 5 mM beta-mercaptoethanol, 1 mM EDTA) and quantified by the Bradford protein assay (data not shown) as previously described (Matsumura *et al.*, 1999). The wild-type protein was

stable when stored at concentrations greater than 1 mg/ml at 4 °C.

### Enzyme kinetics

Purified GUS protein (2 nM-1 μM) was added to 1 ml of GUS buffer containing varying concentrations (5 μM-10 mM) of substrate. The absorption extinction coefficients of *p*-nitrophenol (pNP) and *o*-nitrophenol (oNP) at 405 nm under these conditions are 11.50 and 2.323 mM<sup>-1</sup> cm<sup>-1</sup>, respectively. The formation of the pNP or oNP product at 23 °C for one to 60 minutes was followed in a Shimadzu UV-1601 spectrophotometer. The kinetic parameters of the wild-type and mutant enzymes were calculated by fitting the steady-state initial velocity values to the Michaelis-Menten equation using a least-squares method and the application Kaleidagraph 3.0.5 (Adelbeck software, Reading, PA). The solubility of the substrates (~10 mM) was too low to saturate most of the variant active sites, so we were unable to reliably derive separate values for  $k_{cat}$  and  $K_M$ , although the value of  $k_{cat}/K_M$  could still be independently determined. Each of the  $k_{cat}/K_M$  values was derived from at least nine independent reactions.

### Note added in proof

While this manuscript was in press, Alejandro Hochkoeppler and his colleagues (personal communication) directed the evolution of beta-galactosidase variants with increased beta-glucosidase activity. The substrate specificities of these variants were broadened but not inverted.

### Acknowledgments

We are grateful for receipt of grant N0014-98-1-0883 from the Office of Naval Research. I.M. was supported by a National Science Foundation/Alfred P. Sloan Postdoctoral Research Fellowship in Molecular Evolution. We thank Dr Brian Sauer of DuPont (Wilmington, DE) for sending us strain DH5ΔLAC and Ms Sabine Bell for synthesizing oligonucleotides. We thank members of the Ellington group for helpful discussion and Professor James Bull, Professor Tony Dean and Mr Jay Hesselberth for useful suggestions on the manuscript. Correspondence and requests for materials should be sent to A.E.

### References

Altamirano, M. M., Blackburn, J. M., Aguayo, C. & Fersht, A. R. (2000). Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature*, **403**, 617-622.

Cadwell, R. C. & Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28-33.

Carter, P. & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature*, **332**, 564-568.

Chen, R., Greer, A. & Dean, A. M. (1996). Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc. Natl Acad. Sci. USA*, **93**, 12171-12176.

Cramer, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, **391**, 288-291.

D'Amore, M. A., Gallagher, P. M., Korfhagen, T. R. & Ganschow, R. E. (1988). Complete sequence and organization of the murine beta-glucuronidase gene. *Biochemistry*, **27**, 7131-7140.

el Hawrani, A. S., Sessions, R. B., Moreton, K. M. & Holbrook, J. J. (1996). Guided evolution of enzymes with new substrate specificities. *J. Mol. Biol.* **264**, 97-110.

Forney, L. J., Wong, D. C. & Ferber, D. M. (1989). Selection of amidases with novel substrate specificities from penicillin amidase of *Escherichia coli*. *Appl. Environ. Microbiol.* **55**, 2550-2555.

Graham, L. D., Haggett, K. D., Jennings, P. A., Le Brocque, D. S., Whittaker, R. G. & Schober, P. A. (1993). Random mutagenesis of the substrate-binding site of a serine protease can generate enzymes with increased activities and altered primary specificities. *Biochemistry*, **32**, 6250-6258.

Hall, B. G. (1999). Experimental evolution of Eb enzyme provides clues about the evolution of catalysis and to evolutionary potential. *FEMS Microbiol. Letters*, **174**, 1-8.

Harris, J. L. & Craik, C. S. (1998). Engineering enzyme specificity. *Curr. Opin. Chem. Biol.* **2**, 127-132.

Hedstrom, L., Szilagyi, L. & Rutter, W. J. (1992). Converting trypsin to chymotrypsin: the role of surface loops. *Science*, **255**, 1249-1253.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309-316.

Islam, M. R., Tomatsu, S., Shah, G. N., Grubb, J. H., Jain, S. & Sly, W. S. (1999). Active site residues of human beta-glucuronidase. Evidence for Glu(540) as the nucleophile and Glu(451) as the acid-base residue. *J. Biol. Chem.* **274**, 23451-23455.

Jacobson, R. H., Zhang, X. J., DuBose, R. F. & Matthews, B. W. (1994). Three-dimensional structure of beta-galactosidase from *E. coli*. *Nature*, **369**, 761-766.

Jain, S., Drendel, W. B., Chen, Z. W., Mathews, F. S., Sly, W. S. & Grubb, J. H. (1996). Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nature Struct. Biol.* **3**, 375-381.

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409-425.

Joo, H., Lin, Z. & Arnold, F. H. (1999). Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature*, **399**, 670-673.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.

Matsumura, I., Wallingford, J. B., Surana, N. K., Vize, P. D. & Ellington, A. D. (1999). Directed evolution of the surface chemistry of the reporter enzyme beta-glucuronidase. *Nature Biotechnol.* **17**, 696-701.

May, O., Nguyen, P. T. & Arnold, F. H. (2000). Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. *Nature Biotechnol.* **18**, 317-320.

Miyazaki, K. & Arnold, F. H. (1999). Exploring non-natural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J. Mol. Evol.* **49**, 716-720.

Moore, J. C. & Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvent. *Nature Biotechnol.* **14**, 458-467.

Nishimura, Y., Rosenfeld, M. G., Kreibich, G., Gubler, U., Sabatini, D. D., Adesnik, M. & Andy, R. (1986). Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and *in vitro* insertion of its

- encoded polypeptide into microsomal membranes. *Proc. Natl Acad. Sci. USA*, **83**, 7292-7296.
- Onuffer, J. J. & Kirsch, J. F. (1995). Redesign of the substrate specificity of *Escherichia coli* aspartate aminotransferase to that of *Escherichia coli* tyrosine aminotransferase by homology modeling and site-directed mutagenesis. *Protein Sci.* **4**, 1750-1757.
- Oshima, A., Kyle, J. W., Miller, R. D., Hoffmann, J. W., Powell, P. P. & Grubb, J. H., *et al.* (1987). Cloning, sequencing, and expression of cDNA for human beta-glucuronidase. *Proc. Natl Acad. Sci. USA*, **84**, 685-689.
- Oue, S., Okamoto, A., Yano, T. & Kagamiyama, H. (1999). Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *J. Biol. Chem.* **274**, 2344-2349.
- Palzkill, T. & Botstein, D. (1992). Identification of amino acid substitutions that alter the substrate specificity of TEM-1 beta-lactamase. *J. Bacteriol.* **174**, 5237-5243.
- Ray, J., Bouvet, A., DeSanto, C., Fyfe, J. C., Xu, D. & Wolfe, J. H., *et al.* (1998). Cloning of the canine beta-glucuronidase cDNA, mutation identification in canine MPS VII, and retroviral vector-mediated correction of MPS VII cells. *Genomics*, **48**, 248-253.
- Ruether, J. & Sauer, B. (1989). A screen for SVgpt in *E. coli* DH5 delta lac and DH5 alpha: small colony phenotype. *Nucl. Acids Res.* **17**, 10499.
- Schlaman, H. R., Risseuw, E., Franke-van Dijk, M. E. & Hooykaas, P. J. (1994). Nucleotide sequence corrections of the uidA open reading frame encoding beta-glucuronidase. *Gene*, **138**, 259-260.
- Stemmer, W. P. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747-10751.
- Stemmer, W. P. (1994b). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, **370**, 389-391.
- Waley, S. G. (1969). Some aspects of the evolution of metabolic pathways. *Comput. Biochem. Physiol.* **30**, 1-11.
- Yano, T., Oue, S. & Kagamiyama, H. (1998). Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl Acad. Sci. USA*, **95**, 5511-5515.
- Ycas, M. (1974). On earlier states of the biochemical system. *J. Theor. Biol.* **44**, 145-160.
- Zaccolo, M. & Gherardi, E. (1999). The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775-783.
- Zaccolo, M., Williams, D. M., Brown, D. M. & Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J. Mol. Biol.* **255**, 589-603.
- Zhang, J. H., Dawes, G. & Stemmer, W. P. (1997). Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl Acad. Sci. USA*, **94**, 4504-4509.

Edited by J. A. Wells

(Received 17 July 2000; received in revised form 19 October 2000; accepted 19 October 2000)



<http://www.academicpress.com/jmb>

Supplementary material comprising two Tables is available for IDEAL.