Directed evolution combined with rational design increases activity of GpdQ toward a non-physiological substrate and alters the oligomeric structure of the enzyme

Sylvia H-C.Yip1, Jee-Loon Foo1, Gerhard Schenk2,3, Lawrence R.Gahan2, Paul D.Carr1 and David L.Ollis1,4

1Research School of Chemistry, Australian National University, Canberra, Australia, 2School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia and 3Department of Chemistry, National University of Ireland—Maynooth, Maynooth, Ireland

Received August 25, 2011; revised August 25, 2011; accepted August 26, 2011

Edited by Alan Fersht

Directed evolution was used to enhance the activity of the glycerophosphodiesterase enzyme from Enterobacter aerogenes, GpdQ, toward bis(para-nitrophenol) phosphate (BpNPP), a substrate that is frequently used to assay phosphodiesterases. Native GpdQ has a low level of activity toward BpNPP while the evolved enzymes exhibited $k_{cat}$ values that were well over 100 times better while improvements in $k_{cat}/K_m$ of around 500 times were observed along with improved activity we observed a change in the oligomeric structure in the evolved enzymes. The native enzyme is a hexamer with tightly associated dimers related by a 3-fold axis. The stability of the dimer was attributed in part to the cap domain that forms a disulfide bond with its 2-fold-related subunit and in part due to the fact that dimerization results in burying 23.6% of the monomer’s accessible surface area. The cap domain also forms the top of the active site and contributes an essential part of the interface between 3-fold-related molecules. The evolved proteins quickly lost one of the cysteine residues that formed the disulfide bond and other mutations that might stabilize the cap domain. The likely effect of these mutations was to open up the active site for the new substrate and to favor the formation of dimeric molecules. The breakdown of the oligomeric structure was accompanied by a reduction in the thermal stability of the protein—as monitored by the residual activity of the native and mutant proteins following pre-incubation at elevated temperatures. A discussion on the evolutionary implications of these studies is presented.

Introduction

Understanding how enzymes evolve to process new substrates is a challenge with significant practical implications. Although enzymes are traditionally considered to have a single substrate, it is now appreciated that many enzymes are promiscuous and that these promiscuous activities can be enhanced in the laboratory. In this manuscript we follow the changes that occur in the glycerophosphodiesterase enzyme from Enterobacter aerogenes, GpdQ, as it is evolved to have enhanced activity toward a non-physiological substrate bis(para-nitrophenol) phosphate (BpNPP). We were interested in seeing how the active site was altered during the course of evolution and what effect these alterations had on the activity of the enzyme with other substrates. We were also interested in monitoring how the stability of the protein changed during the experiment.

GpdQ is a phosphohydrolase from E.aerogenes whose rare ability to hydrolyze a variety of simple phosphate diesters first caught the attention and interest of researchers in the 1970s (Gerlt and Whitman, 1975; Cook et al., 1978) but it was not until 2004 that McLoughlin and co-workers isolated the gpdQ gene as part of a 5.6-kb DNA fragment from the E.aerogenes genome (McLoughlin et al., 2004). BLAST analysis of the operon’s sequence indicated that the fragment was largely homologous to the ugp operon of Escherichia coli. This operon is responsible for the uptake and breakdown of glycerol-3-phosphate esters derived from the degradation of the glycerophospholipids found in cell membranes (Raetz, 1986; Brzoska et al., 1994). The E.coli operon contains five genes ugpA, ugpB, ugpC, ugpE and ugpQ. The E.aerogenes operon contains homologous genes for ugpA, ugpB, ugpC and ugpE plus a fifth gene, gpdQ. However, the amino acid sequence of GpdQ bore no similarity to that of UgpQ, the glycerophosphodiesterase expressed by the E.coli ugp operon (Kasahara et al., 1989). Instead, sequence analysis revealed a conserved sequence motif that is found in a diverse collection of metallo-phosphodiesterases/phosphatases that include purple acid phosphatases, Ser/Thr protein phosphatases and calcineurin (Mitic et al., 2006). This collection of proteins has catalytic domains that display an $\alpha$β$\alpha$-fold. A motif with the active sites located between the two central $\beta$ sheets (Koonin, 1994). The conserved residues are found clustered around the active site and form binuclear metal centers that coordinate divalent metal ions that include Mn$^{2+}$, Co$^{2+}$, Ni$^{2+}$, Zn$^{2+}$ and also Fe$^{3+}$ (Strater et al., 1995; Knoefel and Strater, 1999; Rusnak and Mertz, 2000; Schenk et al., 2005; Shenoy et al., 2007).

Keywords: directed evolution/glycerophosphodiesterase/oligomeric state/rational design
Most glycerophosphodiesterases show activity with a range of phosphodiesters (Shang et al., 1997; Zheng et al., 2003; Santelli et al., 2004; Rao et al., 2006; Shi et al., 2008). In fact, most have activity toward model phosphodiester substrates like bis(p-nitrophenyl) phosphate (BpNPP) and O-ethyl(p-nitrophenyl) phosphate. GpdQ acts on these substrates as well as a number of others. GpdQ is unusual in that it will act on simple alkyl phosphodiester like dimethyl phosphate (DMP) and diethyl phosphate (DEP). Additionally, GpdQ shows low activity toward phosphomonoesters, phosphotriesters, phosphonate monoesters and notably, an analog of EA 2192, a highly toxic and stable hydrolysis product of nerve agent VX (McLoughlin et al., 2004; Ghanem et al., 2007; Hadler et al., 2008).

During the course of directed evolution aimed at enhancing the activity toward a non-physiological substrate one would expect changes to occur in the active site. For this reason, a short description of the GpdQ structure and its active site will be given. The protein consists of six subunits that aggregate with $D_3$ symmetry (Jackson et al., 2007; Hadler et al., 2008) that is best described as a trimer of dimers. There is a tight association of subunits about the 2-fold axis and a loose interaction between the dimers related by the perpendicular 3-fold axis. The individual subunits can be divided into three parts: the catalytic domain, a dimerization domain and a 'cap' domain as shown in Fig. 1. The $\alpha/\beta$ sandwich fold of the catalytic domain is not observed in any other glycerophosphodiesterase. All previously characterized glycerophosphodiesterases, including those from *Thermotoga maritima* (Santelli et al., 2004), *Agrobacterium tumefaciens* (Rao et al., 2006), *Thermoanaerobacter tengcongensis* (Rao et al., 2006), and an ortholog of *E.coli* GlpQ found in *Borrelia hermsii* (Shang et al., 1997), possess a central ($\alpha/\beta$)$_8$, or triosephosphateisomerase barrel structure. The four strands of the GpdQ dimerization domain extend the two central sheets of the catalytic domain and link the 2-fold-related subunits. The cap domains further strengthen the link between 2-fold-related subunits—they protrude from one subunit and form part of the active site of an adjacent subunit and are linked to the adjacent subunit by a disulfide bond. The cap peptide overhangs the active site and is also involved in some weak interactions at the interface between GpdQ dimer units that are related by the 3-fold axis as shown in Fig. 1. It is therefore evident that the cap domain has the potential to link substrate specificity and oligomeric structure.

The aim of our work was to evolve the GpdQ enzyme to have enhanced activity toward BpNPP—a non-physiological,
phosphodiester substrate that gives rise to a low level of activity with wild-type GpdQ. It has been noted that catalytic promiscuity is an excellent starting point for the laboratory evolution of enzymes to develop novel functions or activities (O’Brien and Herschlag, 1999; Khersonsky et al., 2006; O’Loughlin et al., 2006). However, there are no general rules as to how enzyme activity can be enhanced. In the experiments described herein, a dramatic change in the oligomeric structure of the protein was observed. After four rounds of evolution, gel filtration analysis indicated that the selection pressure had produced proteins existing as either hexamers or dimers and subsequent kinetic analysis revealed that high catalytic activity was associated with the dimer component. Several rationally designed GpdQ mutants were then constructed with the aim of producing a stable GpdQ dimer or promoting the oligomeric dissociation. From the fifth generation onward, a semi-rational approach was adopted in the engineering of GpdQ’s BpNPP activity where the rationally designed mutants were incorporated as template genes during recombinant PCR. At the end of eight rounds of evolution, selecting for BpNPP activity, mutant forms of GpdQ dimers were characterized with respect to their activity toward BpNPP and also their residual activity toward DEP, p-nitrophenyl phosphate (pNPP—a phosphomonoester) and p-nitrophenyl acetate (pNPA—a carboxylic ester). We also compared the thermostability of the mutants with the native GpdQ.

Materials and methods

Materials

Geneworks synthesized the oligonucleotides used in this work. Library efficient Escherichia coli DH5α cells were purchased from Invitrogen. Taq DNA polymerase and dNTP mix were purchased from Roche while all other enzymes used in molecular cloning were obtained from New England Biolabs. BpNPP sodium salt and all other reagents were supplied by Sigma unless otherwise specified. The preparation of the recombiant pCY76 plasmid containing the gpdQ gene (gpdQ-pCY76) has been described previously (Yang et al., 2003).

gpdQ genetic library creation

The gpdQ gene was amplified from the gpdQ-pCY76 plasmid using the forward primer, 5’ GGA GAT ATA CAT ATG CTG TTA GGC 3’ and the reverse primer, 5’ GCC ATG CTA ATT CCT TAT TAG CGC 3’ with the Ndel (in the forward primer) and EcoRI (in the reverse primer) cloning sites underlined. The 50-μl mixtures contained 0.5 ng template DNA (gpdQ-pCY76), 0.5 μM of each primer, 1.25 mM dNTP mix, 10 mM MgCl₂, 10 mM Tris-HCl (pH 8.3 at 20°C), 50 mM KCl and 0.2 mM MnCl₂ and 0.1 U Taq DNA polymerase.

Libraries were generated with error-prone polymerase chain reaction (epPCR) (Cadwell and Joyce, 1994) and a modified staggered extension process (StEP) (Zhao et al., 1998; Zhao, 2004; Zhao and Zha, 2006) as described in the literature (Stevenson et al., 2008) and will only be briefly described here. Conditions for epPCR were as follows: 94°C for 2 min, then 35 cycles of 45°C for 30 s, 72°C for 2 min, 94°C for 30 s, followed by 45°C for 2 min and a final extension at 72°C for 5 min. The ~0.825-kb PCR amplification product was purified, digested with the necessary restriction endonucleases and ligated into the Ndel and EcoRI cloning sites of pCY76 that was also pre-treated with the same restriction endonucleases and calf intestinal alkaline phosphatase. StEP reaction mixtures consisted of 0.5 ng template DNA (an equimolar mixture of selected variants), 1 μM of each previously described primer, 1.25 mM of dNTP, 1.5 mM MgCl₂, 3% DMSO and 0.1 U Taq DNA polymerase. Cycling parameters for StEP-PCR were 94°C for 2 min, 80 cycles of 94°C for 30 s and 55°C for 1 s and a final extension step at 72°C for 5 min.

In vitro preliminary random screening and secondary screening

Stevenson et al. (2008) have given a detailed description of the screening process used in these experiments and only the particulars of the current set of experiments will be given here. The mutant library was transformed into electrocompeptent library-efficient E. coli DH5α cells. After post-electroporation recovery incubation at 37°C for 3 h, the transformants were diluted with Luria–Bertani–ampicillin (LBA) medium, dispensed into 96-well round-bottomed culture plates (Sarstedt) at a density of 2 cells per well and incubated at 37°C overnight. Subsequently, 20 μl of the 200 μl culture in each well was aliquoted into 96-well flat bottom assay plates (Sarstedt), lysed with a solution containing 1X BugBuster (Novagen) and MnCl₂ (final concentration of 0.5 mM in the 100-μl assay reaction) and assayed with 1 mM BpNPP in 50 mM CHES pH 9.0. Production of p-nitrophenol from the catalytic hydrolysis of BpNPP was measured spectrophotometrically in a 96-well plate reader (Labsystems Multiskan Ascent) at 405 nm and room temperature for up to 10 min. The clones responsible for elevated levels of activity were isolated after screening—this avoided the need to pick colonies prior to screening.

The cultures from 50 wells displaying the highest activities were streaked on LBA plates and incubated overnight at 37°C. Four single colonies from each streaking were then picked to inoculate 200 μl fresh LBA medium and grown overnight at 37°C. Secondary screening of the single-colony cultures was conducted using exactly the same methodology described above for preliminary screening, with the wild-type GpdQ as the reference. A 5 μl aliquot of the solution containing the fittest mutants from each generation was used to inoculate 5 ml LBA medium. The cultures were grown overnight in a 37°C shaker. The plasmid DNA was then isolated and sequenced.

Site-directed mutagenesis (SDM)

Primers for SDM were designed according to guidelines recommended by Stratagene. Single codon changes were introduced to wild-type GpdQ in PCR mixtures that contained 0.5 ng gpdQ-pCY76 template DNA, 25 μM of each of the forward and reverse mutagenic primers, 1.25 mM dNTP, 5.0 μl of 10X Native Plus buffer and 0.05 U Native Pfu DNA polymerase (Stratagene). Conditions for SDM-PCR were as follows: 30 s at 95°C, followed by 25 cycles of 30 s at 95°C, 1 min at 55°C and 8 min at 68°C. The template DNA was digested through DpnI treatment before 1 μl of the PCR mixture was used to transform E. coli DH5α cells. The transformants were plated on ampicillin for selection and grown overnight at 37°C. The plasmid DNA of four single colonies from the LBA plate was isolated and mutagenesis was confirmed through DNA sequencing.
**Protein expression and purification**

All GpdQ mutant proteins were constitutively expressed in the same vector as the native protein. Terrific-Broth–ampicillin (TBA) medium (5 ml) was inoculated with two single colonies of *gpdQ*-pCY76-transformed *E.coli* DH5α cells. The starter culture was grown in a 37°C shaker for ~6 h until the OD_{605} ~ 0.20. TBA medium (500 ml) was then inoculated with the starter culture and grown at 37°C overnight for 16 h before the cells were harvested by centrifugation.

All purification steps were carried out at 4°C. The published protocols of McLoughlin (McLoughlin et al., 2004) and Ghanem (Ghanem et al., 2007) were modified for the purification of GpdQ (Hadler et al., 2008). These protocols use a Q Sepharose anion exchange column, a Phenyl Sepharose hydrophobic interaction column and Superdex 200 size exclusion column. For proteins produced in rounds 1–4 the purification involved the anion exchange step and hydrophobic column. The different oligomeric forms of the GpdQ mutants produced during rounds 5–8 were isolated with the size exclusion column after anion exchange chromatography. In summary, pelleted cells were resuspended with 50 mM Tris-HCl pH 8.0 and lysed in a French pressure cell press operated at 750 psig. The cell debris was removed by centrifugation. The supernatant was loaded onto a 120 ml Superdex 200 prep grade column (GE Healthcare) and eluted with 1 ml elution buffer (50 mM Tris-HCl, pH 8.0) and GpdQ concentration and the least impurities were pooled for the next purification step. Ammonium sulfate solution (2 M) was added to the chosen fractions to a final concentration of ~0.3 M of the salt. The protein solution was incubated for an hour at 4°C and centrifuged. The supernatant was loaded onto a Phenyl Sepharose hydrophobic interaction column and GpdQ was eluted with a 0–0.6 M sodium chloride gradient applied over three column volumes. GpdQ proteins typically eluted between 0.3 and 0.5 M NaCl. The eluted fractions were assayed for phosphodiesterase activity with BpNPP as the substrate. Fractions that displayed the highest activity were assayed for phosphodiesterase activity with BpNPP as the substrate. Fractions that displayed the highest activity were assayed for phosphodiesterase activity with BpNPP as the substrate. Fractions that displayed the highest activity were assayed for phosphodiesterase activity with BpNPP as the substrate. Fractions that displayed the highest activity were assayed for phosphodiesterase activity with BpNPP as the substrate.

In summary, pelleted cells were resuspended with 50 mM Tris-HCl pH 8.0 and lysed in a French pressure cell press operated at 750 psig. The cell debris was removed by centrifugation. The supernatant was loaded onto a 120 ml Superdex 200 prep grade column (GE Healthcare) and eluted with 1 ml elution buffer (50 mM Tris-HCl, pH 8.0) and GpdQ concentration and the least impurities were pooled for the next purification step.

**Enzyme kinetic analysis**

A Varian Cary 1E UV–vis spectrophotometer and Molecular Devices SpectraMax M2e were used for the kinetic assays conducted in this study. All assay reactions were run at 20°C and contained 50 mM CHES pH 9.0, 0.5 mM MnCl₂ and 0.1 mg/ml bovine serum albumin. All buffers, substrate and metal solutions were filtered. Esterase activity with BpNPP, pNPA and pNPP was measured by monitoring the production rate of *p*-nitrophenol spectrophotometrically at 405 nm (ε₄₉₅ = 17 000/M/cm) with a range of concentrations of the substrate (BpNPP: 50 µM to 20 mM; pNPP: 1–25 mM; and pNPA: 1 mM in 7.5% MeOH). For pNPA, blank reactions without GpdQ were run in parallel to determine the background hydrolysis. Hydrolysis of DEP was coupled to alcohol dehydrogenase (10 units per reaction) in the presence of 1 mM NAD⁺ and the rate of ethanol formation was determined by monitoring the rate of production of NADH at 340 nm (ε₃₄₀ = 6.22 mM/cm).

The kinetic parameters, V_{max} and K_{m}, were determined by fitting the kinetic data to the following Michaelis–Menten equation using KaleidaGraph v4.0.2:

\[

v = \frac{V_{max}[S]}{(K_{m} + [S])}

\]

where \( v \) is the initial velocity, \( V_{max} \) is the maximum velocity, \([S]\) is the substrate concentration and \( K_{m} \) is the Michaelis constant. The catalytic turnover rate, \( k_{cat} \), is derived from the following equation:

\[

k_{cat} = \frac{V_{max}}{[E]}

\]

where \([E]\) is the enzyme concentration.

**Dependence of the residual activity of GpdQ on pre-incubation temperature**

GpdQ samples were diluted in 20 mM MOPS pH 7.0 and heated at the desired temperature (40–85°C) for 30 min in a PCR thermal cycler (iCycler, BioRad). The samples were cooled to 25°C and assayed with 5 mM BpNPP in 50 mM CHES pH 9.0 and 0.5 mM MnCl₂. The activities were measured as described above for BpNPP and the results were fitted to the following equation with KaleidaGraph:

\[

a = a_0 - a_0 T^h/(T^h + T_1/2^h)

\]

where \( a_0 \) is the initial activity of the sample, \( a \) is the activity
of the sample after being exposed to temperature $T$ for 30 min, $T_{1/2}$ is the temperature at which half of the GpdQ has been inactivated and $h$ is the Hill coefficient for the thermal inactivation process. For wild-type GpdQ, only data in the 60–85°C range were used to determine $T_{1/2}$ due to thermal activation at lower temperatures. Wild-type GpdQ was heated at 60°C for 30 min and analyzed on a Superdex 200 (GE Healthcare) size-exclusion column to determine the oligomeric composition of the sample after heat treatment, as described in the previous section.

**GpdQ structure modeling**

The program COOT (Emsley and Cowtan, 2004) was used to display and modify the coordinates of GpdQ (PDB code 3D03). The coordinates had been refined against 1.9 Å X-ray diffraction as described by Hadler et al. (2008). When modeling mutant proteins the side chains were built with the most commonly occurring rotomers that did not result in steric clashes with the known wild-type structure. Inspection of the inter-subunit interactions within the hexamer allowed the rational design of mutants Y257stop and F21K. Inspection of the environment around the inter-chain disulfide bond suggested that N53 was a likely residue to mutate in conjunction with changes to C54, C269 or both. Structural figures in this paper were generated using PyMOL (DeLano and Lam, 2005).

**Accessible surface area calculations**

The total accessible surface area of a GpdQ monomer, dimer and hexamer were calculated using the CCP4 program AREAIMOL (http://www.ccp4.ac.uk/html/areaimol.html) that uses the method of Lee and Richards (1971). The GpdQ coordinates listed in the Protein Data Bank (entry 3D03.pdb) stripped of water molecules were used for the calculations.

**Results and discussion**

Eight generations of mutagenesis/recombination and screening were completed as summarized in Table I. The libraries generated in each round of evolution were all of ~100 000 cfu (colony forming units) in size of which about 10 000 were screened for each round of evolution. The epPCR method used in this study had a mutation rate of ~5–10 nucleotide changes, of which two to five resulted in changes to the encoded amino acids while the rest were silent or unexpressed mutations. DNA sequencing of 10 colonies selected at random was used to determine the mutation rate. The recombinatory StEP-PCR is known for its ability to introduce random point mutations (Zhao et al., 1998). Although the mutation rate of libraries constructed using StEP was never evaluated, we observed novel point mutations during DNA sequence analysis of the selected variants in the sixth and seventh generations. For each generation up to 30 mutants were selected for secondary screening and sequenced (Table I, Tables S1 and S2).

**Generations 1–4**

The first four rounds of evolution were screened using a substrate concentration of 5 mM BpNPP—a concentration that was above the $K_m$ of the native enzyme. There were 36 changes in the genes selected as parents for Generation 2. No new changes were observed in rounds 2–4.

By the fourth round there were only four changes (C54G, H217R, Y221H and G259R) to be found in six mutant proteins. This set of experiments was repeated with substrate concentrations below the $K_m$ of the native enzyme. For the first round a concentration of 1 mM was used while in the second round the concentration was reduced to 0.5 mM while in rounds 3 and 4 a concentration of 0.25 mM was used. However, the resulting mutant proteins were essentially the same after four rounds—they contained various combinations of the changes: C54G, H217R, Y221H and G259R. Six mutant proteins were purified using anion exchange and hydrophobic interaction chromatography as described in the methods section. The final protein solutions were obtained by pooling fractions on the basis of their activity and these preparations were used to produce the kinetic data given in Table II. At this point it was clear that changes in the sequence had resulted in mutant proteins with increased $k_{cat}$ and reduced $K_m$. The four individual mutations were made site specifically and the resulting mutants purified using the same protocol as the evolved proteins. The kinetic constants for the site-specific mutants each showed increased $k_{cat}$ and reduced $K_m$ compared with the native enzyme. All the mutations appeared to contribute to the increased activity of the mutant proteins. The best variant to have emerged from the first four generations of directed evolution was mutant 4-6 with the three substitutions, C54G, H217R, Y221H and G259R. The overall $k_{cat}/K_m$ of mutant 4-6 was more than 100-fold higher than that of the wild type. Two of the mutations in 4-6 had the potential to affect the structure of the cap domain.

### Table I. Outline of directed evolution generations

<table>
<thead>
<tr>
<th>Generation</th>
<th>Method used to generate diversity</th>
<th>Number of mutants selected and sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>epPCR</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>StEP</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>StEP</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>StEP</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>epPCR</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>StEP</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>StEP</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>StEP</td>
<td>14</td>
</tr>
</tbody>
</table>

**The effects of mutations**

As noted in the introduction a disulfide bond links C269 to C54 of the 2-fold-related chain in the wild-type protein so that the C54G change would result in the loss of a covalent link between the two subunits. This mutation by itself was not likely to cause a breakdown of the 2-fold symmetry, but it was likely to have an effect on the conformation of the peptide of the cap domain. Since this peptide forms part of the active site, such a mutation could open up the active site so that large substrates such as BpNPP could gain better access—this was consistent with the increased activity of the 4-6 mutant. There were other mutations that were likely to affect the cap domain. Glycine and proline residues have unusual conformational abilities and are commonly found in turns or bends in protein secondary and tertiary structures (Richards and Kundrot, 1988; Frishman and Argos, 1995). G259, together with its adjacent residue, P260, form a bend
in the cap domain of GpdQ (Fig. 2a). A mutation at this point would almost certainly alter the orientation of the domain due to steric hindrance between the side chains of R259 and W261. The effects of this mutation, like C54G, are likely to include destabilization of the dimer and also the latter interaction is not likely to be preserved with the H217R change—the side chain of the arginine is much larger than that of histidine. The H217R change, when made in isolation, does not improve the catalytic properties of GpdQ suggesting that it only enhances the activity of the enzyme in the presence of the C54G mutation.

**Size exclusion chromatography**

Size exclusion chromatography of the 4-6 mutant revealed partial dissociation of the hexameric enzyme to dimers as can be seen in Fig. 3. On the analytical size exclusion column, native GpdQ protein elutes at a volume of 190 kDa as is consistent with a hexamer of 31 kDa subunits. The most abundant species in the 4-6 mutant elutes at a volume that gave a molecular weight of 64 kDa, suggesting a dimer. Similar results were obtained with the preparative scale column—the native protein eluted with a volume consistent with a hexamer of 150 kDa. The 4-6 mutant eluted with a volume consistent with a molecular weight of 54 kDa.

The kinetic characterization of mutant 4-6 was repeated by assaying the hexamers and the dimers separately. Two sets of activities corresponding to the two oligomeric forms were measured and the kinetic data revealed that high activity was associated with the dimer (Table II) and that the hexameric
form of the mutant enzyme had very little activity. The hexamer → dimer dissociation would result in the disruption to the 3-fold symmetry leaving the C$_2$ symmetry of the dimer intact. It appeared that the GpdQ dimer was optimal for the hydrolysis of BpNPP and attempts were then made to identify changes that would break down the 3-fold symmetry of GpdQ so that a stable dimer would be formed.

The accessible surface area was calculated for a GpdQ monomer, dimer and hexamer as 12,894, 19,696 and 52,046 Å$^2$ respectively. This implies a burying of 3046 Å$^2$ (23.6%) per monomer on dimer formation, and a further burying of 2347 Å$^2$ (11.9%) per dimer upon trimerization. According to Ponstingl et al. (2005), based on a study of 163 non-redundant oligomeric protein structures, as a rule of thumb $\sim$18% of a subunit’s accessible surface area would normally be buried upon oligomer formation. The fact that GpdQ dimers bury considerably more than this is likely to indicate the dimer will be more stable than an average oligomeric protein, whereas GpdQ hexamers, which bury less than this benchmark, would be less stable.

**Rational design of GpdQ**

The structure of GpdQ was examined to identify mutations that could promote the dissociation of the hexamer. Seven potential mutants were identified and these are indicated with an asterisk in Table II. The rationale for their selection and their effects on enzyme kinetics are detailed in the following paragraph. These proteins were made and purified using ion exchange and size exclusion chromatography so that changes in the oligomeric structure could be monitored. In all, but for one, case evidence for the formation of dimeric species could be observed. The W261R mutant remained in the hexameric form while the F21K mutation produced predominantly dimeric enzyme. The remaining mutants were predominantly hexameric with some dimer present.

The values for the catalytic parameters for the designed mutants are given in Table II. Increased activity was observed for all mutations. The C269S and C269A mutations were designed to have the same effect as the C54G change: to eliminate the disulfide bond. It was surprising that both changes produced higher $k_{cat}$ and $K_m$ values than did the C54G mutation. The F21 residue (of Chain A) and W261 (of Chain B) of one dimer make hydrophobic contacts with the side chains of H256 (of Chain D) and the aliphatic atoms of the side chain of D30 (of Chain D) that belong to another dimer, as shown in Fig. 4. The intent of making mutants F21K and W261R was to disrupt the hexamer-stabilizing hydrophobic interactions between

---

**Fig. 2.** (a) The Cap domain of GpdQ is shown with residues 259–261 labeled. (b) Residue 217 is shown along with the metals and metal coordinating residues of GpdQ. (c) The proposed structure of R217 as found in the 4-6 mutant.

**Fig. 3.** (a) Elution of three different GpdQ proteins from the Superdex 200 size exclusion column. The wild type, elutes as a single peak (hexamer); mutants 4-6 (upper curve), and 8-7 (lower curve), elute as dimer predominantly. The column was calibrated with protein standards. (b) Elution profile of F21K mutant showing predominantly dimer.
3-fold-related subunits. This approach succeeded in the case of the F21K, but not in the case of the W261R that may have resulted in the formation of a salt link with D30 (of Chain D). The F21K mutant was predominantly dimeric and had a better than 6-fold increase in $k_{\text{cat}}$ and a 5-fold reduction in $K_m$ to give an almost 40-fold increase in $k_{\text{cat}}/K_m$. The amide group of N53 residue forms a hydrogen bond with side chain of S268. The alterations made to N53 gave rise to modest changes in the catalytic properties of the enzyme. The Y257stop mutation resulted in mixture of hexamer and dimer—mostly hexamer. This change resulted in an enzyme with no Cap domain so that the active site was made more accessible. The enzyme was significantly more active with a 10-fold increase in the catalytic turnover rate but a small decrease in $K_m$ compared with wild type.

**Fig. 4.** The environment of F21. Hydrophobic interactions form between the side chains of F21 of Chain A, W261 of Chain B, H256 of Chain D and the aliphatic part of the side chain of D30 from Chain D. A hydrogen bond exists between OD2 of D30 and NE2 of H256 of Chain D orienting the side chain of D30 appropriately to interact with F21 of Chain A.

**Generations 5–8**

From the sixth generation onwards, rationally designed GpdQ mutants were included as template DNA together with the best mutants from each round of directed evolution for gene recombination or shuffling. N53D, N53A, C269S and C269A were introduced into the GpdQ mutant library at the sixth generation, while Y257stop and F21K were introduced at the seventh and eighth generations, respectively. At the end of each round the best 30 mutants were selected and retested and the best 15 were sequenced. Representative sequence changes observed for Generations 5–8 are given in Table S2 while Table III gives all the sequence changes and relative activities for the 14 best mutants obtained from the eighth generation. The most common change observed was H217R, which first arose in Generation 1. When made in isolation this mutation had a modest effect on both $k_{\text{cat}}$ and $K_m$ and did not have a dramatic effect on activity when combined with the C54G change as found in mutant 4-2. However, it has survived and is found in all the most active mutants while the Y221H change has all but disappeared from the most active proteins; it is found in only 1 of the 15 most active proteins in Generation 8. The Y221H and H217R are not found simultaneously in an evolved mutant—probably because they are too close to be recombined by the StEP shuffling process. Most of the more active mutants identified in Generation 8 (Table III) had changes at either 54 or 269 that would have prevented the formation of the disulfide bond. Of the two that did not have a mutation at either of these positions, one (8-4) had the Y221H change but the remaining mutant (8-1) had the G259R change that may have prevented the cap folding correctly so that the disulfide bond could not form. The S127A, N103D and R12L changes were first observed in either the fifth and sixth rounds of evolution and were common among the best mutants found in generation 8 (Table III). Single-site mutants R12L, N103D and S127A were also produced and assayed (Table II). The three mutants remained associated as hexameric proteins and only one (R12L) showed significantly improved catalytic

**Table III.** Showing the occurrence of individual amino acid changes within the eighth-round mutants, their frequency, $f$, and percentage of total number of mutations

<table>
<thead>
<tr>
<th>R12L</th>
<th>F21K</th>
<th>C54G</th>
<th>N103D</th>
<th>S127A</th>
<th>H217R</th>
<th>Y221H</th>
<th>Y257 stop</th>
<th>G259R</th>
<th>C269A</th>
<th>Activity improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>8-1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>15.3</td>
</tr>
<tr>
<td>8-2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>15.2</td>
</tr>
<tr>
<td>8-3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>14.5</td>
</tr>
<tr>
<td>8-4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>13.6</td>
</tr>
<tr>
<td>8-5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>13.4</td>
</tr>
<tr>
<td>8-6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>12.5</td>
</tr>
<tr>
<td>8-7</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>12.4</td>
</tr>
<tr>
<td>8-8</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>12.2</td>
</tr>
<tr>
<td>8-9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11.7</td>
</tr>
<tr>
<td>8-10</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11.6</td>
</tr>
<tr>
<td>8-11</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11.6</td>
</tr>
<tr>
<td>8-12</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11.4</td>
</tr>
<tr>
<td>8-13</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11.3</td>
</tr>
<tr>
<td>8-14</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>11.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$f$</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>6</th>
<th>12</th>
<th>1</th>
<th>1</th>
<th>7</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of $f$ (%)</td>
<td>14.3</td>
<td>35.7</td>
<td>35.7</td>
<td>42.9</td>
<td>42.9</td>
<td>85.7</td>
<td>7.1</td>
<td>7.1</td>
<td>50.0</td>
<td>42.9</td>
</tr>
</tbody>
</table>

The activity improvement indicates the relative $V_{\max}$ (wild type = 1.0) based on a 96-well plate reader measurements during secondary screening. Some wells contained mixtures of different oligomeric species.
properties over the wild-type enzyme. R12 forms a salt link to OE2 of E59. This residue in turn is hydrogen bonded via OE1 to the backbone nitrogen atoms of C54 and G55. These hydrogen bonded interactions may have a role in orienting the backbone in a suitable conformation for disulfide bond formation between C54 and C269. The absence of this interaction in an R12L mutant, together with the C269A sequence change, as seen in 8-2 and 8-12, is likely to destabilize the conformation of the cap domain and facilitate access to the active site for large substrates. The effects of the N103D and S127A mutations are more difficult to explain. N103 is a surface residue whose side chain points into the solvent. Figure 5 demonstrates how this residue is situated at the interface between an acidic and basic patch of surface charge and how the mutation is likely to affect the charge distribution. The surface charge may have a role in guiding the incoming substrate or helping it adopt a suitable orientation before reaching the active site. The side chain S127 is 3.62 Å away from N80. N80 is the ligand responsible for the lower metal binding affinity at the β-site (Hadler et al., 2008, 2009, 2010). Despite the proximity, a hydrogen bond interaction between S127 and N80 was unlikely because of the conformation of S127, whose side chain hydroxyl group points away from the amide group of N80. An alanine substitution at this position might allow N80 more flexibility, thus enhancing the ability of the residue to bind and release the β-metal more easily. However, the S127A enzyme (Table II) gave no improvement in either $k_{\text{cat}}$ or $K_m$ when present as a single-site mutant.

Mutants 8-1, 8-2, 8-3, 8-5 and 8-7 were expressed, purified and characterized for BpNPP activity (Table I). The purified proteins were all predominantly dimeric. Small amounts of 8-1 and 8-2 formed monomers and about 10% of 8-3, 8-5 and 8-7 forming a higher-molecular-weight species. Both monomeric and dimeric forms of the mutants were purified and characterized. The monomeric forms of 8-1 and 8-2 were more active than the dimers, but their activity was far below that of the dimeric forms of the remaining enzymes. The three most catalytically active variants were the dimeric forms of 8-3, 8-5 and 8-7. As can be seen in Table II, compared with the native enzyme these mutant proteins have 25-fold higher $k_{\text{cat}}$ values and 20-fold lower $K_m$ values; giving a 500-fold increase in the $k_{\text{cat}}/K_m$—at $10^{-6}/\text{M/s}$ they are starting to approach the diffusion limit. It is noteworthy that these three mutants all carried the C269A mutation, which was first induced through SDM. Additionally, mutant 8-7 bore the F21K mutation that was also rationally designed. These results highlighted the synergy between directed evolution and rational design in engineering GpdQ’s BpNPP activity.

**Substrate specificity of GpdQ variants**

The activities of the best three GpdQ variants toward other substrates were investigated. Native GpdQ shows a low level of activity toward the phosphomonoester substrate pNPP. The three best mutants all showed significant enhancement in activity toward pNPP though to a lesser extent than those observed for BpNPP (Table IV). Mutants 8-3 and 8-5 show large decreases in $K_m$ and these combined with about a 4-fold increase in $k_{\text{cat}}$ result in a better than 20-fold improvement in $k_{\text{cat}}/K_m$. pNPP bears some similarity to the BpNPP in that both have a nitrophenoate leaving group and it is not surprising that the mutant GpdQ enzymes have increased activity toward pNPP. However, the same leaving group in a simple ester such as pNPA displayed a very low level of activity with both native and mutant enzymes. The activities of mutants 8-3 and 8-5 toward the carboxylic ester pNPA are very similar to that of wild type while 8-7 hydrolyses the substrate at a markedly reduced rate (Table IV).

With substrates that did not have a p-nitrophenoate leaving group the results were quite different compared to those noted above. The kinetic constants of native GpdQ measured with diethylphosphate (DEP) are not impressive. The $k_{\text{cat}}$ is $2.8 \times 10^{-3} \text{s}^{-1}$ while the $K_m$ is $15.7 \text{mM}$ giving a $k_{\text{cat}}/K_m$ of $0.18/\text{M/s}$. None of the best three mutants (8-3, 8-5 and 8-7) had detectable activity toward this substrate. In this case the leaving group is quite poor and the native enzyme may be able to provide a group capable of protonating it. This group is not required for the hydrolysis of BpNPP and it has been lost during the course of evolution. BpNPP is a relatively large substrate that is easy to hydrolyze and it would seem that increased activity was achieved by simply making the active site more accessible to the substrate. This process has clearly had a detrimental effect on the ability of the evolved proteins to process DEP.

---

**Fig. 5.** Electostatic surface charge diagrams for above wild-type GpdQ (3D03.pdb) and below a simulated mutation of N103D in this structure. The residue lies at the interface between surface acidic and basic regions. The mutation alters the relative sizes of the two charged surface regions.
Temperature dependence of the catalytic activity of GpdQ

The residual activities of the native and mutant enzymes were measured subsequent to being subjected to pre-incubation at temperatures between 40 and 85°C to gain some indication of their stability as shown in Fig. 6. Measurements were made for native GpdQ and dimeric forms of the F21K, 8-3, 8-5 and 8-7 mutants. It should be noted that these measurements were made at pH 7 where the difference between native and mutant activity is small. All the GpdQ variants lose activity with increasing pre-incubation temperature; probably due to irreversible denaturation. However, the wild type shows an anomalous surge in activity with increasing pre-incubation temperatures, reaching maximum residual activity at ~60°C, before thermal inactivation ensues. The elution profile from a size-exclusion column of a wild-type GpdQ sample heated at 60°C indicates the presence of denatured and hexameric GpdQ along with a small amount of the enzyme in the dimeric form. This suggests that the thermal activation observed for the wild-type enzyme is due to dissociation of the weakly associated hexamer (a trimer of dimers) to form dimers. The net increase in activity despite the small amount of dimeric GpdQ relative to the denatured form suggests that the dimer is highly active and is able to compensate for the activity loss due to irreversible thermal denaturation. These results are consistent with our proposal that dimeric GpdQ allows easier access of the bulky substrate BpNPP to the active site and leads to its enhanced hydrolysis. The eighth generation GpdQ variants evidently have lower thermostability relative to the wild type, as indicated by their lower $T_{1/2}$ (see caption of Fig. 6). The near-identical $T_{1/2}$ of F21K and wild-type GpdQ implies that the dimeric form of GpdQ is not inherently less thermostable than the hexamer. The lowered $T_{1/2}$ of the eighth-generation mutants is therefore probably a consequence of multiple mutations having a deleterious effect on the thermostability of the enzymes. Mutations that have resulted in increased activity have resulted in mutant enzymes with decreased stability.

Concluding remarks

We have used directed evolution to produce mutants of the GpdQ enzyme that have enhanced catalytic properties toward BpNPP at high pH. The selection conditions were not demanding; the substrate has a good leaving group and one would expect catalysis to occur rapidly at high pH. Enhanced activity is achieved by making the active site more accessible to the new and bulkier substrate by altering the conformation of the cap domain. In so doing, the interaction between 3-fold-related dimers was compromised and the oligomeric structure of the protein was broken down. Higher oligomeric forms of fourth round mutants were found to have little activity with BpNPP and further evolution with selection for increased catalytic activity resulted in further breakdown of the oligomeric structure with the appearance of monomeric forms of the enzyme in the eighth round. However, a higher oligomeric form of the enzyme was never totally eliminated—even in the eighth round. These observations suggest that there were some selective advantages associated with the hexameric form of the enzyme. The thermal stability measurements indicate that the eighth-round mutants are considerably less stable than the native enzyme. The hexameric structure of the native molecule can be easily broken down by heat treatment, but it appears that the resulting dimeric molecule is more stable than the dimeric forms of the evolved enzymes containing multiple mutations. Evolution appears to be a compromise between altering the conformation of the cap domain to give increased activity and a loss of stability that results from the breakdown of the oligomeric structure.

There is some evidence that the cap domain continues to play a role in catalysis even toward the end of the evolutionary process. If removing the cap domain were all that were necessary to improve hydrolysis of BpNPP, the Y257stop
mutant would be the most active. It is not and its $K_m$ is an order of magnitude higher than the highly evolved mutants obtained in the eighth round that still possess a modified cap domain. This suggests that the cap domain changes conformation so that it can still interact with the substrate or that it stabilizes the mutant proteins. We have selected for mutants that show enhanced activity toward a large compound that is easily broken down at high pH. The improvement is modest at pH 7 where the stability measurements were made. Furthermore, the final mutants showed no detectable ability to degrade smaller and more stable esters like DEP. Presumably, to improve the activity of GpdQ toward DEP requires the maintenance of elements of the native active site that are lost in the experiments with selective pressure toward the large substrate BpNPP described herein.

GpdQ appears to have evolved to degrade glycerophosphodiesters that are quite different to the compounds we have worked with. We have addressed questions related to how the native enzyme can be modified to accommodate a new substrate. It is also interesting to speculate how this process might explain how the native enzyme evolved. We could address this question by the structures of related proteins. As we noted in the introduction, GpdQ has an active site found in a number of phosphodiesterases—specifically, the secondary structure and the residues used to coordinate metals. However, the structural similarity between GpdQ and these proteins does not, with one exception, extend far beyond sections of the molecules that form their active sites. The one exception is the Rv0805 protein, which has quite a different function to GpdQ. It is a cyclic nucleotide phosphodiesterase (Class III) and is found in Mycobacterium tuberculosis (Shenoy et al., 2005). The sequence similarity between GpdQ and Rv0805 is only 22%, but the structural similarity is high: 260 out of 295 residues align with an rms displacement of 3.3 Å and a DALI Z score of 31.5. Rv0805 is a dimer with the same catalytic and dimerization domains as GpdQ. Crystal structures of Rv0805 are available for both the full-length protein (Podobnik et al., 2009) and a C-terminal truncation mutant (Shenoy et al., 2007) that essentially has the equivalent of the Cap domain removed. The catalytic activity of the latter is only two-thirds that of the native protein, but unlike the native protein consists of a mixture of monomers and dimers. It would appear that in Rv0805, as in GpdQ, the C-terminal domain contributes to both the catalytic activity and stability of the protein.

In Rv0805 the C-terminal peptide is considerably larger than its equivalent in GpdQ. In Rv0805 it consists of 82 residues, 62 of which are visible in the X-ray structure (Podobnik et al., 2009). The 20 residues that are not visible are presumably disordered and are the residues likely to be involved in membrane binding (Podobnik et al., 2009). This contrasts to the cap domain in GpdQ that has 18 residues, of which 15 are visible in the X-ray structure (Jackson et al., 2007). The only secondary structure present in the Rv0805 C-terminal domain is an amphiphilic $\alpha$-helix formed by residues 278–288. This helix forms hydrophobic interactions with the helix 178–183 from the dimer-related subunit, and so, as in GpdQ, constitutes a domain-swapped peptide likely to stabilize the dimer.

Inspection of a docked substrate model of AMP in Rv0805 (PDB entry 3IBB) shows that the C-terminal domain has the potential to effect substrate binding either directly or via interactions with the neighboring peptides that form the 178–183-helix or 239–244-loop regions.

In the case of GpdQ the stability of the dimer was significantly increased by the formation of a disulfide bond between residues 54 and 269. In addition the cap domain was used to form the top of the active site and links with other dimers that resulted in the formation of a hexameric molecule. Directed evolution essentially reversed this fine tuning—we produced an enzyme that had a more open active site and a simpler oligomeric structure than that of the native enzyme. Although we have gained enhanced activity toward compounds like BpNPP, we have lost stability and the ability to degrade small compounds like DMP. The mutants we obtained after eight generations of in vitro evolution looks a great deal like an enzyme precursor that nature has subsequently evolved—Rv0805.

It is interesting to compare the results obtained in this work with experiments in which an oligomeric protein was evolved for increased thermostability. Specifically, mutants of $\beta$-glucuronidase were selected for activity after being subjected to heat treatment for a fixed time. After four rounds of evolution, active mutants were identified that survived temperatures that denatured the native protein. The mutations were found not only to increase the stability of the oligomeric forms of the proteins, but also to degrade their catalytic performance with respect to their physiological substrate (Flores and Ellington, 2002). These experiments illustrate the relationship between oligomeric structure, stability and kinetic properties. Nature has evolved the protein so that these characteristics are adequate for the requirements of the host organism and attempts to increase stability (without a selection for the preservation of catalytic activity) has resulted in an increase in the stability of the oligomeric structure and a decrease in catalytic performance. Our experiments have taken the opposite direction, but they also illustrate the relationship between stability, oligomeric structure and catalytic activity. We have selected for mutants with increased activity toward a non-physiological substrate (without a selection for the preservation of enzyme stability) and found mutants with reduced oligomeric structure and stability. If we desire an enzyme that has both enhanced catalytic activity toward a larger, non-physiological substrate and good thermal stability, the most efficient means would be to include stability in the selection process. In the absence of such experiments it is hard to predict whether the stable oligomeric state or a modified cap structure would be the more dominant structural constraint.

Supplementary data
Supplementary data are available at PEDS online.

Funding
This research was supported by a grant from the Australian Research Council to D.O., G.S. and L.G.

References