Structure of Recombinant *Haemophilus Influenzae e* (P4) Acid Phosphatase Reveals a New Member of the Haloacid Dehalogenase Superfamily[†,‡]

Richard L. Felts,[§] Zhonghui Ou,[||] Thomas J. Reilly,[⊥,#] and John J. Tanner*[,§,||]

Departments of Chemistry, Biochemistry, and Veterinary Pathobiology and Veterinary Medicine Diagnostic Laboratory, University of Missouri—Columbia, Columbia, Missouri 65211

Received May 24, 2007; Revised Manuscript Received July 17, 2007

ABSTRACT: Lipoprotein *e* (P4) from *Haemophilus influenzae* belongs to the "DDDD" superfamily of phosphohydrolases and is the prototype of class C nonspecific acid phosphatases. P4 is also a component of a *H. influenzae* vaccine. We report the crystal structures of recombinant P4 in the ligand-free and tungstate-inhibited forms, which are the first structures of a class C phosphatase. P4 has a two-domain architecture consisting of a core α/β domain and a smaller α domain. The core domain features a five-stranded β-sheet flanked by helices on both sides that is reminiscent of the haloacid dehalogenase superfamily. The α domain appears to be unique and plays roles in substrate binding and dimerization. The active site is solvent accessible and located in a cleft between the two domains. The structure shows that P4 is a metalloenzyme and that magnesium is the most likely metal ion in the crystalline recombinant enzyme. The ligands of the metal ion are the carboxyl groups of the first and third Asp residues of the DDDD motif, the backbone carbonyl of the second Asp of the DDDD motif, and two water molecules. The structure of the tungstate-bound enzyme suggests that Asp64 is the nucleophile that attacks the substrate P atom. Dimerization appears to be important for catalysis because intersubunit contacts stabilize the active site. Analysis of the structural context of mutations engineered for vaccine studies shows that the most promising mutations are located in the dimer interface. This observation suggests a structure-based vaccine design strategy in which the dimer interface is disrupted in order to expose epitopes that are buried in dimeric P4.

Lipoprotein *e* (P4) is one of six highly conserved major outer membrane proteins of *Haemophilus influenzae*, a Gram-negative facultatively anaerobic coccobacillus. The organism is a common commensal inhabitant of the human nasopharynx and the etiologic agent of local and invasive infections in both pediatric and adult populations. *H. influenzae* strains cause a variety of infections including otitis media, sinusitis, bronchitis, conjunctivitis, and pneumonia. In particular, *H. influenzae* infections of the lower respiratory tract are a major cause of mortality in both infants and children in developing countries (*1*).

An obvious prelude to clinical infection is colonization of the host. Colonization and survival of the microbe on the mucosal surface is dependent on the activity of proteins located at the periphery of the bacterium, juxtaposed to the host's mucosal surface. These outer membrane proteins have been under intensive investigation as vaccine candidates and as targets for development of chemotherapeutic agents. Many of the 36 different protein species found in the *H. influenzae* envelope have been characterized with regard to immunogenicity, antigenicity, and protective attributes (*2*).

Lipoprotein *e* (P4), the subject of this study, is a highly conserved 28 kDa acid phosphatase found in all strains of *H. influenzae* (*1*, *3*). The immature protein has a lipoprotein signal peptide at the N-terminus, which results in anchoring of the mature protein in the outer membrane surface by a lipidated N-terminal cysteine. P4 is currently being investigated for use in a vaccine against nontypeable *H. influenzae* (*4*−*6*). Recombinant P4 and P4 mutants are highly immunogenic, and anti-P4 antibodies exhibit bacteriocidal activity against nontypeable *H. influenzae* strains (*6*). Moreover, in a nasal challenge mouse model, intranasal immunization with formulations containing recombinant P4 and other proteins resulted in greater than 100× reduction of nasal colonization of nontypeable *H. influenzae* strain SR7332 (*5*).

Wild-type *e* (P4) and the recombinant enzyme (denoted rP4)[1] exhibit phosphomonoesterase activity with aryl phosphate substrates including nicotinamide mononucleotide (NMN), tyrosine phosphate, phenyl phosphate, *p*-nitrophenyl phosphate, and 4-methylumbelliferyl phosphate (*3*, *7*). De-

[†] This research was supported by the NIH Grant U54 AI057160 to the Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Research (MRCE) and the University of Missouri Research Board.

[‡] Coordinates and structure factors have been deposited in the Protein Data Bank under Accession Numbers 2HLK and 2HLL.

* Address correspondence to this author. Tel: 573-884-1280. Fax: 573-882-2754. E-mail: tannerjj@missouri.edu.

[§] Department of Chemistry, University of Missouri—Columbia.

[||] Department of Biochemistry, University of Missouri—Columbia.

[⊥] Department of Veterinary Pathobiology, University of Missouri—Columbia.

[#] Veterinary Medicine Diagnostic Laboratory, University of Missouri—Columbia.

[1] Abbreviations: rP4, recombinant *e* (P4); NMN, nicotinamide mononucleotide; NSAP, nonspecific acid phosphatase; HAD, haloacid dehalogenase; PDB, Protein Data Bank.

10.1021/bi701016m CCC: $37.00
Published on Web 09/08/2007

phosphorylation of NMN by P4 is biologically relevant because *H. influenzae* lacks almost all of the enzymes necessary for biosynthesis of NAD^+ and therefore requires NAD^+, NMN, and nicotinamide riboside as necessary growth factors. The role of P4 in NAD^+ uptake appears to be the dephosphorylation of NMN to nicotinamide riboside, which is subsequently taken up by the organism (*7*). Additionally, P4 is important for uptake of hemin (*8*) under aerobic conditions, although the molecular mechanism of this function remains obscure.

Amino acid sequence analysis has provided insights into the relationship of P4 to other phosphoryl transfer enzymes. P4 belongs to the "DDDD" superfamily of phosphohydrolases, which is distinguished by a bipartite sequence motif consisting of two pairs of essential aspartic acid residues separated by a linker region (*9*). The DDDD motif of *e* (P4) corresponds to Asp64, Asp66, Asp181, and Asp185. The DDDD superfamily comprises diverse phosphatases from bacteria, archaea, and eukaryotes, including nonspecific acid phosphatases, phosphoglycolate phosphatases, histidinol phosphatases, phosphoserine phosphatases, and threalose-6-phosphatases (*9*).

Within the DDDD superfamily, P4 is the prototype of a group of secreted bacterial enzymes known as class C nonspecific acid phosphatases (NSAPs). Class C NSAPs are distinguished by the bipartite sequence motif [IV]-[VAL]-D-[IL]-D-E-T-[VM]-L-X-[NT]-X(2)-Y in the N-terminus and [IV]-[LM]-X(2)-G-D-[NT]-L-X-D-F in the C-terminus (*9*). Genes encoding class C NSAPs have been identified in *Chryseobacterium meningosepticum* (*OlpA*) (*10*), *Streptococcus equisimilis* (*LppC*) (*11*), *H. influenzae* (*hel*) (*12*), *Helicobacter pylori* (HP1285) (*13*), *Staphylococcus aureus* (*14*), *Francisella tularensis* (unpublished data), *Pasteurella multocida* (*15*), and *Bacillus anthracis* (*16*). Several of these class C enzymes have been purified and characterized, but three-dimensional structures have not been reported. To understand the structural basis of phosphatase activity and to aid in vaccine development, we have solved the crystal structure of rP4.

MATERIALS AND METHODS

Crystallization and Heavy Atom Derivative Preparation. Preparation of the plasmid encoding rP4 was described previously (*17*). Initially, the *hel* gene was cloned from *H. influenzae* strain Rd KW20 (*18*). The recombinant gene was subsequently modified to replace the N-terminal signal sequence with the *pelB* leader sequence from *Erwinia chrysanthemi* and to replace the N-terminal Cys with Met-Val (*17*). As a result, the rP4 protein used here for crystallization is free of lipid modification and targeted to the *Escherichia coli* periplasm.

Expression, purification, and crystallization of rP4 have been previously described (*19*). Briefly, crystals were grown in sitting drops using a reservoir solution of 0.2 M $MgCl_2$, 0.1 M Tris-HCl, pH 8.1−8.5, and 28−36% (w/v) PEG 4000. The space group is $P4_22_12$ with unit cell dimensions of $a = 65.6$ Å and $c = 101.4$ Å and one rP4 chain per asymmetric unit. The structure was solved using single isomorphous replacement with anomalous scattering phasing based on a tungstate (Na_2WO_4) derivative. The derivative used for phasing was obtained as follows. A native crystal was transferred into cryobuffer consisting of 0.2 M $MgCl_2$, 0.1 M Tris-HCl, pH 8.5, 40% (w/v) PEG 4000, and 15% (w/v) PEG 200. The crystal was then transferred into 20 μL of the cryobuffer supplemented with 50 mM Na_2WO_4. After soaking in the tungstate solution for 40 min, the crystal was picked up with a mounting loop and plunged into liquid nitrogen.

Table 1: Data Collection and Refinement Statistics[a]

	native	tungstate derivative
wavelength (Å)	1.0359	1.0052
space group	$P4_22_12$	$P4_22_12$
unit cell dimensions (Å)	$a = 65.6$, $c = 101.4$	$a = 65.8$, $c = 101.9$
resolution (Å)	42−1.70 (1.79−1.70)	47−2.00 (2.07−2.00)
no. of observations	158442	111384
no. of unique reflections	25076	15761
average redundancy	6.3 (3.6)	7.1 (7.1)
completeness (%)	99.8 (98.3)	99.7 (99.9)
mean $I/\sigma(I)$	17.5 (2.6)	9.1 (3.8)
$R_{merge}(I)$	0.071 (0.410)	0.117 (0.416)
no. of protein atoms	1941	1933
no. of water molecules	184	146
no. of PEG fragments	1	1
no. of Mg^{2+} ions	1	1
no. of tungstate ions	0	1
R_{cryst}	0.186 (0.262)	0.182 (0.220)
R_{free}[b]	0.222 (0.329)	0.225 (0.282)
rmsd[c]		
bond lengths (Å)	0.010	0.013
bond angles (deg)	1.14	1.31
Ramachandran plot[d]		
favored (%)	97.3	97.3
allowed (%)	3.7	3.7
average *B*-factors (Å^2)		
protein	24	19
Mg^{2+}	16	16
PEG fragment	38	28
water molecules	27	22
tungstate ion		10
PDB accession code	2HLK	2HLL

[a] Values for the outer resolution shell are in parentheses. [b] 5% random test set. [c] Compared to Engh and Huber parameters (*39*). [d] The Ramachandran plot was generated with RAMPAGE (*40*).

X-ray Diffraction Data Collection. Diffraction data were collected at Advanced Light Source beamline 4.2.2. Acquisition of a 1.7 Å resolution native data set was described previously (*19*). A data set for the tungstate derivative was collected with $\lambda = 1.0052$ Å ($E = 12335$ eV), which corresponds to an energy slightly higher than the L_I absorption edge of W ($E = 12100$ eV). We note that the theoretical value of f'' is approximately 12 electrons at $\lambda = 1.0052$ Å. The tungstate data set consisted of 90° of data collected with an oscillation angle of 1°, exposure time of 5 s/frame, and detector distance of 140 mm. The tungstate data set was processed to 2.0 Å resolution using d*TREK (*20*). Data collection statistics for the native and derivative data sets are listed in Table 1.

An anomalous difference Patterson map (15−3.5 Å resolution) was calculated to assess the suitability of the tungstate data set for phasing. The $w = 0.5$ Harker section displayed a 15σ peak at $(u, v) = (0.15, 0.33)$, suggesting the presence of one high-occupancy tungstate site per asymmetric unit.

Phasing and Refinement Calculations. Phasing calculations were performed with SOLVE (*21*) using the tungstate and native data sets. One heavy atom site was identified by SOLVE and resulted in a figure of merit of 0.27 for data to

2.0 Å resolution. Density modification and phase extension to 1.7 Å resolution were performed with RESOLVE (*21*). The figure of merit from RESOLVE was 0.60 for reflections to 1.7 Å resolution. The density-modified phases were input to ARP/wARP (*22*) for automated backbone tracing. GuiSIDE (*22*) was used to add side chains to the ARP/wARP backbone model. The resulting model included 211 residues with side chains out of the expected 254 residues in the asymmetric unit.

The model from ARP/wARP/GuiSIDE was improved through several rounds of model building in COOT (*23*) followed by refinement to 1.7 Å resolution with REFMAC5 (*24*). The final model includes residues 11−254, 1 Mg^{2+} ion bound in the active site, 184 water molecules, and 1 PEG fragment. Residues of the structure have been numbered so that the first two residues of the expressed protein are Met0 and Val1. Hence, the four aspartate residues of the DDDD motif have residue numbers 64, 66, 181, and 185, in agreement with the numbering convention of Green and coworkers (*3*). See Table 1 for refinement statistics.

The 1.7 Å resolution native structure, with selected active site residues, Mg^{2+}, and solvent omitted, served as the starting point for refinement of the tungstate complex at 2.0 Å resolution. The selection of test reflections for cross-validation was based on the native data set. The final model of the tungstate-bound enzyme includes residues 11−254, 1 Mg^{2+} ion, 1 tungstate ion, 146 water molecules, and 1 PEG fragment. Refinement statistics are listed in Table 1.

PyMol (*25*) and COOT were used to analyze structures. PyMol was used to prepare figures.

RESULTS

P4: A New Member of the HAD Superfamily. The 1.7 Å resolution structure of rP4 was solved using single isomorphous replacement with anomalous scattering phasing based on a high-occupancy single-site tungstate derivative. The structure of the tungstate-bound enzyme was subsequently solved at 2.0 Å resolution using the native structure as the starting model.

The rP4 structure comprises two domains: an α/*β* core domain and an α domain (Figure 1). The core domain contains the residues responsible for catalysis and binding the active site metal ion. The active site is located in a cleft between the two domains, as denoted by the tungstate and Mg^{2+} ions in Figure 1a,b. The core domain (residues 31−70, 106−216) consists of a central five-stranded parallel *β*-sheet flanked by three α-helices (αB, αE, αF) on one side and two α-helices (αG, αH) on the other side. The *β*-sheet has strand order of 3, 2, 1, 4, 5 (Figure 1c). Note that the topology of the α/*β* core domain resembles that of the Rossmann dinucleotide-binding fold.

The α domain is located above the carboxyl-terminal edge of the strands of the *β*-sheet and has three components (Figure 1). The first component is a long α-helix at the N-terminus of the polypeptide chain (αA, residues 11−30). This helix connects to another long helix (αB), which is part of the core domain, and together αA and αB form a long helical elbow that spans the entire length of rP4 (Figure 1b). The second component of the α domain is a helix−loop−helix substructure (αC and αD, residues 71−105) inserted between *β*1 and αE of the core domain. The third component

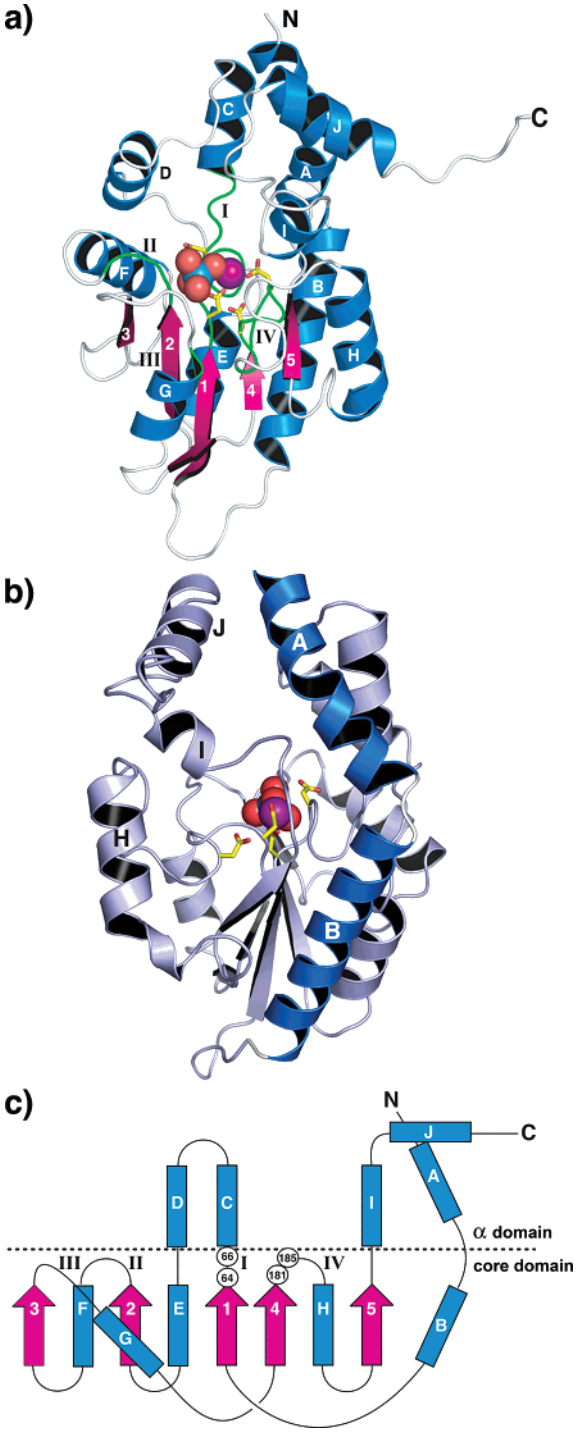

FIGURE 1: Overall structure of rP4. (a) Ribbon drawing of the rP4 subunit. The tungstate inhibitor (blue/red) and the active site Mg^{2+} ion (purple) are drawn as spheres. The aspartate side chains of the DDDD motif are shown in yellow. Strands are colored pink and labeled 1−5. Helices are colored blue and labeled A−J. Active site loops are colored green and labeled I−IV. The N- and C-termini are labeled. (b) Another view of the rP4 subunit highlighting the long helical elbow formed by αA and αB. Compared to (a), this view is rotated by about 90° around the vertical axis. (c) Secondary structure topology diagram of rP4. Circles denote the conserved aspartate residues of the DDDD superfamily motif. Secondary structure elements below the dotted line constitute the core domain. Those above the line constitute the α domain.

consists of 38 residues at the C-terminus (residues 217−254) that form two helices (αI, αJ) followed by a six-residue extension (residues 249−254). As discussed below, the α domain is involved in dimerization.

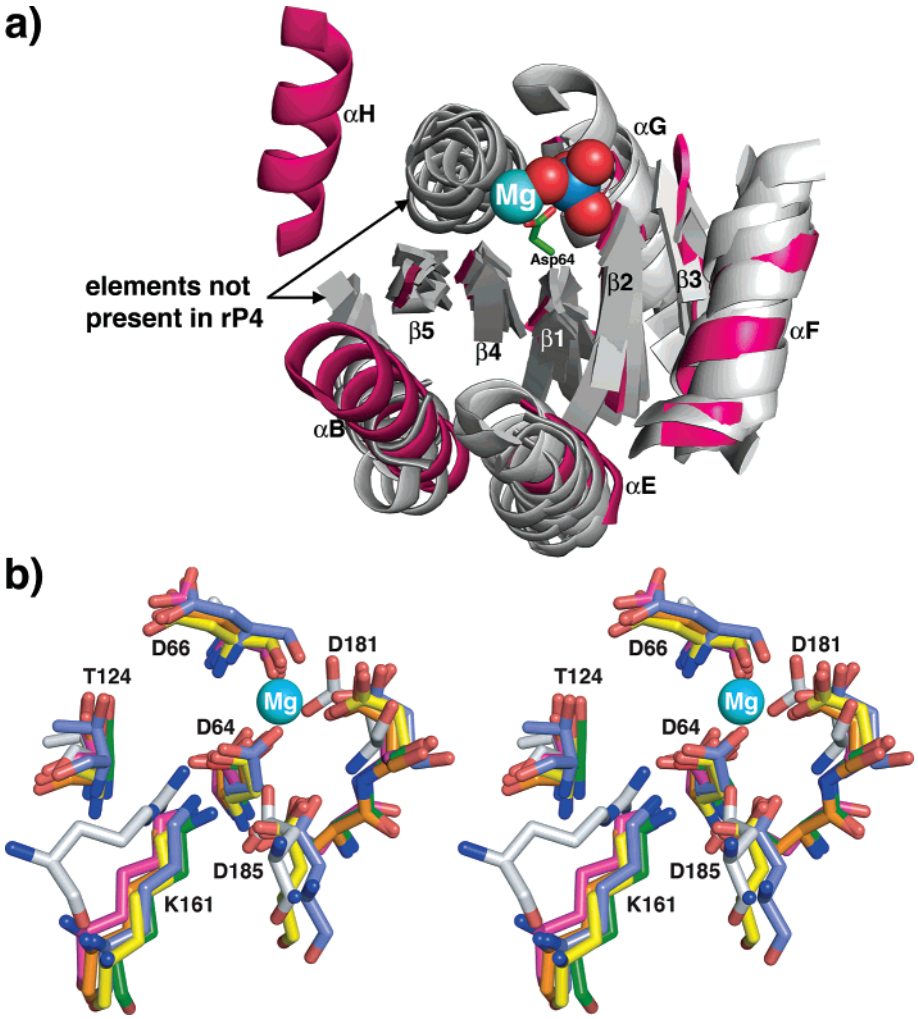

FIGURE 2: Comparison of rP4 with five other HAD superfamily enzymes. The structures shown are the class B NSAP AphA [PDB code 2B8J (*28*)], bacteriophage T4 polynucleotide kinase [PDB code 1LTQ (*29*)], mitochondrial 5′(3′)-deoxyribonucleotidase 1 [PDB code 1Z4Q (*30*)], phosphotyrosine phosphatase MDP-1 [PDB code 1U7P (*31*)], and phosphonoacetaldehyde hydrolase [PDB code 1FEZ (*32*)]. (a) Ribbon drawing of the conserved core domains, with rP4 shown in pink and the other enzymes in silver. The Mg^{2+} ion (blue) and the tungstate (blue/red) of rP4 are drawn as spheres. The proposed nucleophile, Asp64, is shown in stick form and is located at the C-terminus of strand β1. The strands and helices are labeled in accordance with the topology diagram in Figure 1. (b) Stereoview of the active sites, with rP4 shown in yellow and the other enzymes shown in various colors. The Mg^{2+} ion (blue) and residue labels correspond to rP4.

Comparison of the rP4 structure with structures in the Protein Data Bank (PDB) (*26*) using DALI (*27*) revealed that rP4 is a member of the haloacid dehalogenase (HAD) enzyme superfamily. The closest structural homologues are HAD superfamily enzymes and include the class B NSAP AphA [DALI *Z*-score = 9.8; PDB code 2B8J (*28*)], bacteriophage T4 polynucleotide kinase [$Z = 9.7$; PDB code 1LTQ (*29*)], mitochondrial 5′(3′)-deoxyribonucleotidase [$Z = 8.1$; PDB code 1Z4Q (*30*)], phosphotyrosine phosphatase MDP-1 from *Mus musculus* [$Z = 7.6$; PDB code 1U7P (*31*)], and *Bacillus cereus* phosphonoacetaldehyde hydrolase [$Z = 7.6$; PDB code 1FEZ (*32*)]. Despite negligible overall amino acid sequence identity to rP4 (<19% for aligned residues), these five enzymes share with rP4 a common core domain (Figure 2a) and active site (Figure 2b). The structural homology is highest for strands 1−5 of rP4 and helices B, E, F, and G (Figure 2a).

There are also three notable differences between rP4 and the other HAD enzymes within the core domains. First, none of the rP4 homologues have a helix analogous to αH (Figure 2a). Second, rP4 lacks a helix found in AphA, T4 polynucleotide kinase, MDP-1, and phosphonoacetaldehyde hydrolase (Figure 2a). This helix follows the strand analogous to β4 of rP4 in the latter four enzymes and is replaced by a short loop in rP4 and the 5′(3′)-deoxyribonucleotidase. Third, three of the enzymes pictured in Figure 2a [AphA, the 5′-(3′)-deoxyribonucleotidase, and phosphonoacetaldehyde hydrolase] have one to two additional β-strands not found in rP4, T4 polynucleotide kinase, and MDP-1.

Structure of the Active Site. The active site of rP4 is located in a cleft between the core and α domains. The cleft has approximate dimensions of 29 Å × 25 Å × 22 Å and is accessible to solvent. In the native structure, the cleft is filled with a PEG fragment and over 2 dozen water molecules. The tungstate ion binds in the lower part of the cleft, atop the carboxyl-terminal edge of the β-sheet. There is a large space in the upper part of the cleft that is presumably the binding site for the leaving groups of substrates (Figure 1a). Comparison of rP4 to structures of human mitochondrial deoxyribonucleotidase complexed with nucleoside monophosphate substrates (*30*) suggests that this open space is large enough to bind substrates such as NMN.

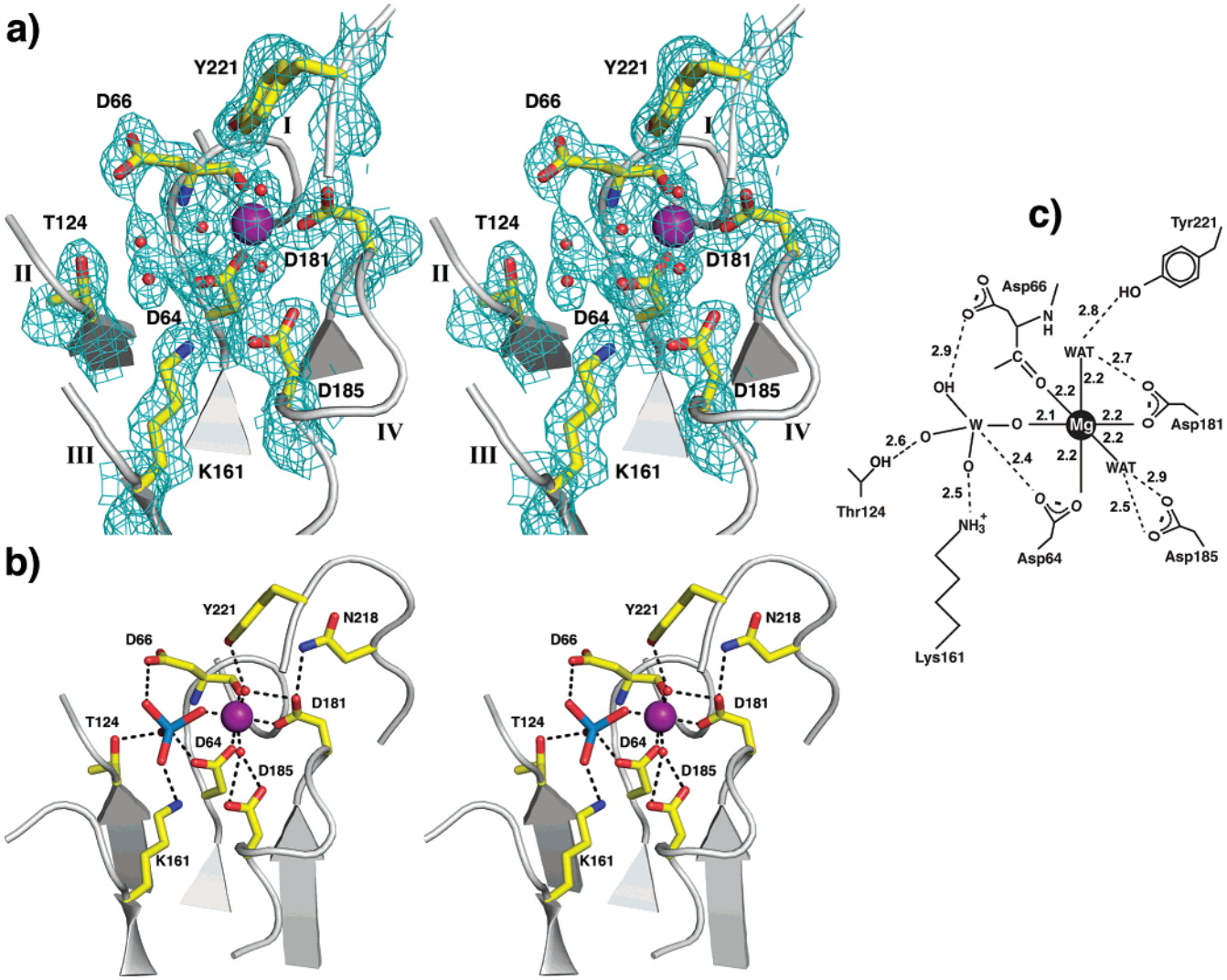

FIGURE 3: Active site of rP4. (a) Stereographic drawing of the active site of the native enzyme. Mg^{2+} is represented by the purple sphere. The mesh represents the experimental electron density map from RESOLVE after solvent flattening and phase extension to 1.7 Å. The contour level is 1σ. (b) Stereographic drawing of the active site of rP4 complexed with the inhibitor tungstate. The dashed lines indicate electrostatic interactions. The tungstate ion is shown in blue/red, and the Mg^{2+} ion is colored purple. (c) Schematic diagram of the active site.

Following the nomenclature used for HAD superfamily enzymes, the active site consists primarily of four catalytic loops, labeled I− IV (Figure 1a,c). The catalytic loops of rP4 connect secondary structure elements β1 and αC (loop I, residues 63−73), β2 and αF (loop II, residues 124−132), β3 and αG (loop III, residues 155−161), and β4 and αH (loop IV, residues 180−186).

Electron density maps suggested the presence of a 6-coordinate metal ion in the active site (Figure 3a), which is consistent with the fact that most HAD superfamily members are metalloenzymes (*33*). Mg^{2+} is a logical candidate because the crystallization solution contains 0.2 M $MgCl_2$. Also, $N = 6$ is the most common coordination number of Mg^{2+} in proteins (*34*). Structure refinement with Mg^{2+} modeled in the active site produced $F_o - F_c$ electron density maps having negligible residual density at the metal site and a reasonable *B*-value for the metal ion (16 $Å^2$). It is concluded that Mg^{2+} is the most likely candidate for the active site metal ion of crystalline rP4. We note that Mg^{2+} is the catalytic metal ion of AphA (*35*).

The Mg^{2+} ion has octahedral geometry, and the conserved DDDD motif plays a prominent role in metal binding (Figure 3). The protein provides three of the six ligands for the metal ion, which are the carboxyl groups of Asp64 and Asp181 and the main chain carbonyl of Asp66. All three residues are part of the DDDD motif. Water molecules serve as the other three ligands in the inhibitor-free conformation. These water molecules bridge the metal ion with Asp64, Asp181, Asp185, and Tyr221 (Figure 3a). We note that it has been reported that the activity of wild-type *e* (P4) purified from *H. influenzae* is optimal in the presence of $CuSO_4$ (*18*). The metal ion site described here is atypical for Cu, which tends to have coordination numbers of $N = 3-5$, with His, Cys, and Met as the most common ligands (*34*).

The tungstate-bound structure provides insights into the roles of active site residues because tungstate mimics the phosphate moiety of phosphomonoester substrates. The tungstate ion is bound by the Mg^{2+} ion and catalytic loops I−III. The oxygen atoms of the tungstate interact with the side chains of Asp66, Thr124, and Lys161 (Figure 3b,c), as

well as the backbone amine groups of Leu65, Asp66, and Asn125. The binding of tungstate displaces four water molecules but causes no major conformational changes in the protein. The four displaced water molecules occupy the same locations in the native structure as the tungstate oxygen atoms in the inhibitor-bound structure.

The rP4/tungstate structure suggests that Asp64 is the nucleophile that attacks the P atom of the substrate. A carboxyl oxygen atom of Asp64 is only 2.4 Å from the W atom of the inhibitor, which is the closest contact observed between the protein and the W atom (Figure 3b,c). Moreover, the other carboxyl oxygen atom of Asp64 is bound to the Mg^{2+} ion, which would depress the pK_a and increase the nucleophilicity of Asp64.

The constellation of residues in the active site of rP4 is similar to those of close homologues from the HAD superfamily, as shown in Figure 2b. In particular, visual inspection of superimposed structures shows that there is high structural similarity among these enzymes in the location of the metal ion and the four Asp residues of the DDDD motif. Moreover, two residues that interact with the substrate phosphoryl group, corresponding to rP4 residues Thr124 of catalytic loop II and Lys161 of catalytic loop III, are also structurally conserved by this group of enzymes.

Dimeric Structure of rP4. Purified rP4 forms an apparent homodimer in solution based on gel filtration chromatography (*18*). The asymmetric unit contains one rP4 chain, so interfaces generated by crystallographic symmetry were analyzed with the PISA server (*36*) to gain insight into the dimeric structure of rP4. The largest interface in the crystal lattice is generated by 2-fold rotation around the crystallographic *c*-axis. This interface buries 3053 Å^2 of surface area, while the other interfaces each bury less than 350 Å^2 of surface area. There are 44 hydrogen bonds and 4 ion pairs in this interface, implying substantial electrostatic complementarity. On the basis of this analysis, we propose that the assembly generated by 2-fold rotation around the *c*-axis represents the dimer formed in solution (Figure 4a).

The primary dimer interface is formed by αA, αB, αH, β5, αI, and αJ. These elements form a large groove containing two pronounced holes separated by a crosspiece (Figure 4b). The crosspiece is formed by αH and its preceding loop. In the dimer, the αA−αB elbow of one subunit packs into the groove of the other subunit (Figure 4c), which allows side chains of αA and αB to poke into the holes of the groove and interact with the active site of the other subunit. The side chains of Glu32 and Leu22 seem to be particularly important in this regard, as described below.

In addition to this large interface, there is a smaller intersubunit interface formed by interaction of the C-terminal extension (residues 248−254) of one subunit with the helix−loop−helix insert of the other subunit (Figure 4a). Here, Trp251 of the extension packs against αC and αD, making nonpolar contacts with the side chains of Tyr75, Trp78, Arg93, and Trp94.

Dimerization appears to be important for catalysis because side chains from αA and αB of one subunit poke through the holes in the dimer interface into the active site of the opposite subunit. Most notably, Glu32 forms direct and indirect electrostatic interactions with catalytic loop IV of the opposite subunit (Figure 5a). The direct interaction is a hydrogen bond between the carboxyl of Glu32 and the amine

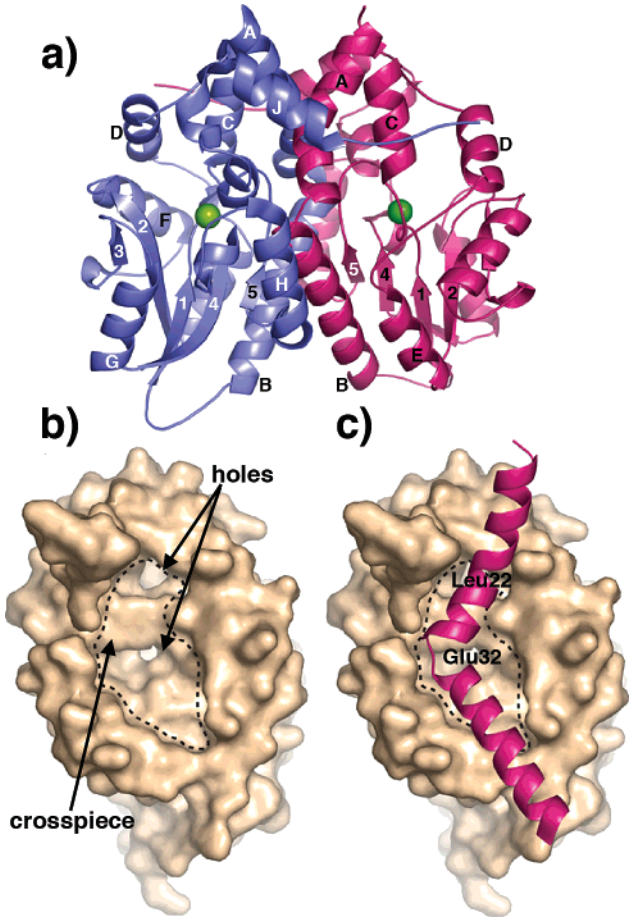

FIGURE 4: Dimeric structure of rP4. (a) Ribbon drawing of the rP4 dimer. The two subunits are colored blue and pink. Selected secondary structure elements are labeled. The green spheres represent Mg^{2+}. (b) Surface topography of the dimer interface. This view corresponds to the blue subunit from (a) rotated by 90° about the vertical axis so that the intersubunit surface faces the reader. The pink subunit has been removed to reveal the intersubunit surface. The dotted curve indicates the large groove that forms intersubunit contacts with the helical elbow of the opposite subunit. Note the two large holes in the groove, which lead to the active site. (c) Depiction of the helical elbow of one subunit fitting into the groove of the opposite subunit.

of Leu183 of loop IV. In the indirect interaction, Glu32 forms an intersubunit ion pair with Arg198, which, in turn, interacts with Asn182 and Asp184 of loop IV. We note that Glu32 and Arg198 are highly conserved among class C NSAPs. These interactions most likely help to define and stabilize the conformation of catalytic loop IV, which contributes two of the metal-binding ligands, Asp181 and Asp185.

The second place where dimerization appears to be important for catalysis is the upper part of the active site cleft in the region bounded by the helix−loop−helix insert on one side and by αA, αI, and αJ on the other side. In this region, Leu22 of one subunit pokes into the active site of the other subunit and fills a hole surrounded by αA, αI, and αJ (Figure 5b). As discussed above, it is possible that the upper part of the active site is involved in binding substrates and so this intersubunit contact may be important for substrate recognition.

Structural Context of rP4 Mutations Designed for Vaccine Studies. Recombinant P4 is a promising vaccine candidate against nontypeable *H. influenzae* infections, and several rP4 mutant enzymes have been generated in an effort to engineer

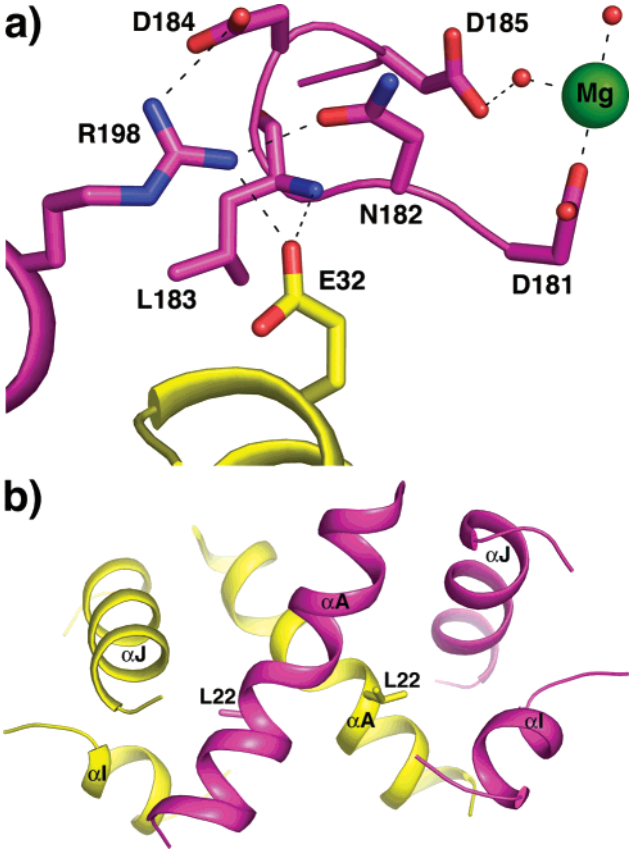

FIGURE 5: Stabilization of the active site through intersubunit interactions. (a) Intersubunit interactions formed between Glu32 and catalytic loop IV. Active site residues of one subunit are colored magenta. Residues of the adjacent subunit are colored yellow. (b) Intersubunit interactions involving Leu22. The two subunits are colored magenta and yellow.

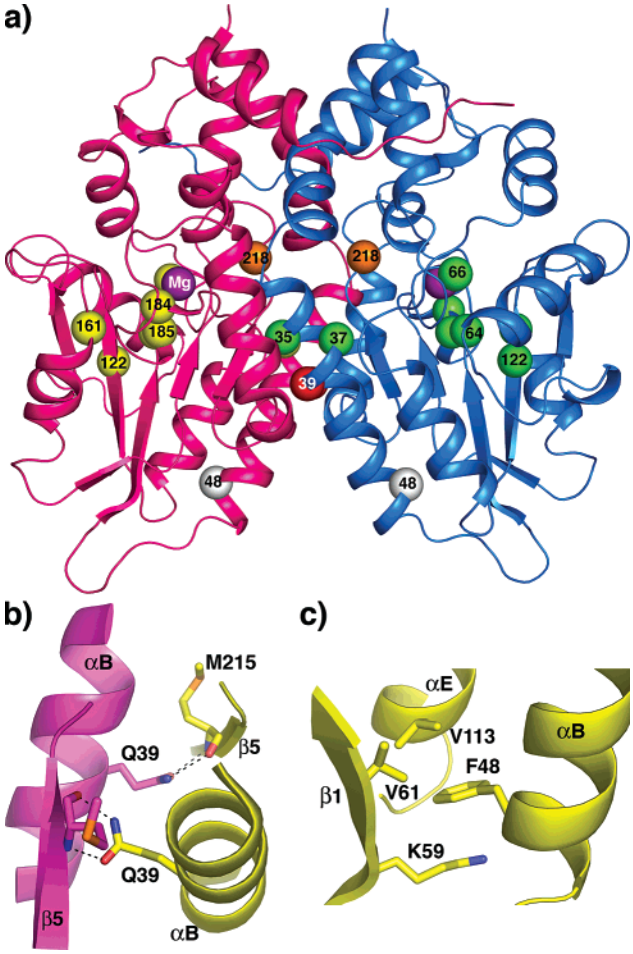

FIGURE 6: Structural context of mutations from Green and co-workers (*3*). (a) Locations of the 12 residues that were mutated. The C_α atoms of mutated residues are drawn as spheres. The yellow spheres belong to the pink subunit, and the green spheres belong to the blue subunit. The three most promising mutations for vaccine development are shown in orange (Asn218Gln), red (Gln39Glu), and white (Phe48Cys). The magnesium ion is shown as a purple sphere. (b) Close-up view of Gln39. Residues from the two subunits are colored pink and yellow. (c) Close-up view of Phe48.

catalytically inactive proteins for use in a vaccine. Green and co-workers recently measured the immune responses of mice vaccinated individually with 12 different rP4 mutants (*3*). The mutated residues are mapped onto the three-dimensional structure in Figure 6a. The structure shows that the mutations are largely confined to a swath passing through the upper part of the core domain, with none of the mutated residues appearing in the α domain (Figure 6a).

Three of the 12 mutants are of particular interest (Asn218Gln, Gln39Glu, Phe48Cys) because they displayed the best combination of reduced phosphatase activity and favorable immune response as measured by titer and bactericidal activity (*3*). Interestingly, the structure shows that these three residues do not participate directly in catalysis or binding the phosphoryl group of the substrate.

Asn218 is located in the dimer interface in the loop that connects β5 to αI (Figure 6a). It forms a hydrogen bond with Asp181, which is one of the Mg^{2+} ligands (Figure 3b). Thus, the role of Asn218 may be to help orient Asp181 for interaction with the metal ion. Mutation of Asn218 to Gln lengthens the side chain by one methylene, which may disrupt the hydrogen bond to Asp181 and adversely affect metal binding. Furthermore, the backbone carbonyl of Asn218 forms a hydrogen bond with the indole amine of Trp26 of the opposite subunit, which forms nonpolar intersubunit contacts in the dimer interface. Mutation of Asn218 may also disrupt interactions in the dimer interface.

Gln39 is located in the middle of αB (Figure 6a) and forms hydrogen bonds with β5 of the opposite subunit (Figure 6b). The side chain of Gln39 forms a bidentate intersubunit hydrogen bond interaction with the backbone of Met215 (Figure 6b). Mutation of Gln39 to Glu would disrupt this intersubunit interaction. In addition, because αB is close to and nearly parallel with the 2-fold axis of the dimer, Gln39 side chains of opposite subunits are only 5 Å apart (Figure 6b). Thus, mutation of Gln39 to Glu buries two carboxyl groups close together in the dimer interface, which is unfavorable. We suggest that the diminished activity of Gln39Glu results from disruption of the dimer interface.

Lastly, Phe48 is located at the C-terminus of αB (Figure 6a), and its side chain makes nonpolar contacts with Val61, Val113, and Lys59 (Figure 6c). The side chain of Lys59 reaches under the phenyl ring of Phe48 to form a hydrogen bond with the carbonyl of Phe48 thus capping αB. Mutation of Phe48 to Cys would certainly affect the tight packing in this region. Since αB is involved in dimerization, this mutation, like Gln39Glu and Asn218Gln, may diminish activity by disrupting the dimer interface.

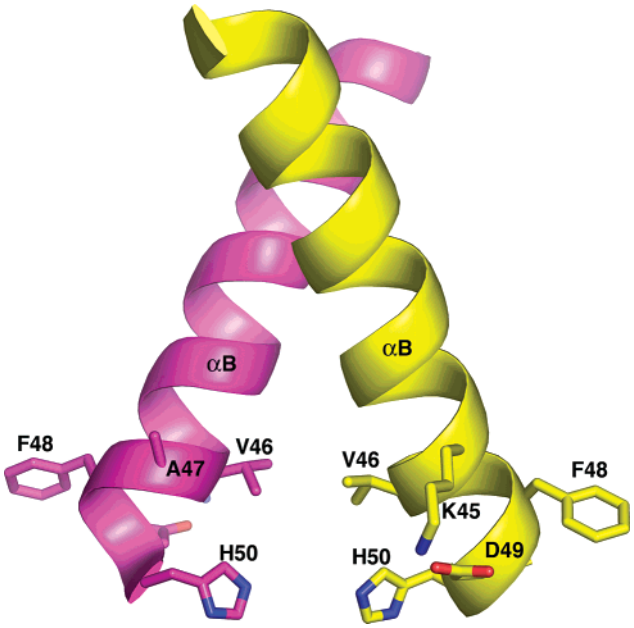

FIGURE 7: Structural context of the proposed KVAFDH heme binding motif. The location of this motif in the dimer can be seen in Figure 4a.

Structural Context of the Proposed KVAFDH Hemin Binding Motif. P4 is thought to be important for aerobic growth of *H. influenzae* and may mediate uptake of hemin from the environment (*37*, *38*). The hemin binding domain was originally proposed by Reidl and Mekalanos (*37*) as the six-residue sequence motif KVAFDH in residues 45−50 due to its similarity to the heme binding motifs found in hemoglobin, cytochrome c_3, HbpA, and HAP-1.

Residues 45−50 are located near the C-terminus of αB (Figure 7). Because this helix is near the 2-fold axis of the dimer, the αB helices of the two subunits pack against each other in the dimer interface (Figure 4a). As a result, the two KVAFDH motifs of the dimer form a V-shaped intersubunit cleft with the side chains of Val46 and His50 pointing into the cleft (Figure 7). The two His50 residues of the motif face each other with imidazole N atoms separated by 5.3 Å. It is tempting to speculate that a molecule of hemin could bind in this cleft edgewise and engage in bishistidinyl coordination to His50 in a manner similar to that found in cytochrome c_3.

DISCUSSION

The structure of rP4 revealed a two-domain architecture consisting of the quintessential HAD superfamily core domain and a smaller α-helical domain that is involved in dimerization. AphA, the prototype of class B NSAPs, also has a HAD superfamily core domain (*35*), and we found that AphA is a close structural homologue of rP4. Class B and C NSAPs are distantly related (<20% identity), but both feature the DDDD motif. In their paper describing the first structure of a class B NSAP, Calderone and co-workers suggested that the HAD fold might be conserved throughout the DDDD superfamily (*35*). The rP4 structure supports this idea.

The α domain distinguishes P4 from other HAD superfamily enzymes, including class B NSAPs. This domain mediates the unique dimeric assembly of P4. In particular, the protein−protein interaction surface observed in the rP4 dimer is not seen in AphA. Class B NSAPs form homotetramers in which the subunits interact via intermolecular β-sheet formation involving an N-terminal extension (*35*). It thus appears that the mode of self-association is a major structural difference between class B and C NSAPs.

The α domain might also function as a cap over the active site of the core domain. Two types of cap domains are commonly observed in HAD superfamily enzymes. HAD subclass I enzymes have a small α-helical bundle cap domain inserted between catalytic loops I and II, whereas subclass II enzymes have a larger β-sandwich cap domain inserted between loops II and III (*33*). The primary role of cap domains is to facilitate binding of the substrate leaving group and control accessibility of the active site to solvent (*33*). The helix−loop−helix substructure inserted between loops I and II of rP4 (αC, αD) is reminiscent of type I caps. However, this subdomain would only partially enclose the leaving group, based on comparison of rP4 to structures of human mitochondrial deoxyribonucleotidase complexed with nucleoside monophosphate substrates (*30*). It is possible that the insert of rP4 moves in response to substrate binding in order to increase interaction with the leaving group. Indeed, rigid body movements of cap domains have been described for subclass I HAD enzymes (*33*). Another possibility is that other parts of the α domain combine with the αC−αD insert to form a "split" cap domain. For example, helix I, and its adjacent loops, could potentially play this role.

Considerable effort has been focused on designing mutants of rP4 for use in a vaccine against nontypeable *H. influenzae* strains (*3*−*5*). An important design constraint is that the protein should lack catalytic activity because vaccination with an active phosphatase could affect host metabolism and signaling pathways. The four aspartate residues of the DDDD motif are obvious candidates for mutagenesis since these residues directly participate in catalysis and metal binding. Interestingly, Green and co-workers showed that such Asp mutants, while devoid of activity, do not elicit the best immune responses compared to the mutants Asn218Gln, Gln39Glu, and Phe48Cys (*3*).

We used the rP4 structures to help us understand why Asn218Gln, Gln39Glu, and Phe48Cys have diminished activity. All three residues are located in the dimer interface, and we reason that mutation of these sites might adversely affect activity by disrupting dimerization. Thus, the enhanced immunogenicity of these mutants may result from exposing epitopes that are buried in the dimer.

These results suggest a potential structure-based vaccine design strategy involving dimer-destabilizing mutagenesis. The dimer interface is large and offers many possibilities for mutagenesis. In fact, the entire α domain represents new territory to be explored, as shown in Figure 6a. Leu22 and Glu32 are two intriguing candidates since they stabilize the active site via intersubunit contacts. Also, Trp26 side chains from the two subunits interact with each other. Mutation of Trp26 to Ala would likely result in a destabilizing cavity buried in the dimer interface. Finally, a truncation mutation that removes the C-terminal extension may have a significant effect on dimerization and hence activity. An important prerequisite for implementing this strategy will be to verify that dimerization is compromised in rP4 mutants Asn218Gln, Gln39Glu, and Phe48Cys.

ACKNOWLEDGMENT

We thank Dr. Jay Nix of ALS beamline 4.2.2 for help with data collection and processing. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract DE-AC02-05CH11231.

REFERENCES

1. Foxwell, A. R., Kyd, J. M., and Cripps, A. W. (1998) Nontypeable *Haemophilus influenzae*: pathogenesis and prevention, *Microbiol. Mol. Biol. Rev. 62*, 294−308.
2. Loeb, M. R., and Smith, D. H. (1980) Outer membrane protein composition in disease isolates of *Haemophilus influenzae*: pathogenic and epidemiological implications, *Infect. Immun. 30*, 709−717.
3. Green, B. A., Baranyi, E., Reilly, T. J., Smith, A. L., and Zlotnick, G. W. (2005) Certain site-directed, nonenzymatically active mutants of the *Haemophilus influenzae* P4 lipoprotein are able to elicit bactericidal antibodies, *Infect. Immun. 73*, 4454−4457.
4. Hotomi, M., Ikeda, Y., Suzumoto, M., Yamauchi, K., Green, B. A., Zlotnick, G., Billal, D. S., Shimada, J., Fujihara, K., and Yamanaka, N. (2005) A recombinant P4 protein of *Haemophilus influenzae* induces specific immune responses biologically active against nasopharyngeal colonization in mice after intranasal immunization, *Vaccine 23*, 1294−1300.
5. Mason, K. W., Zhu, D., Scheuer, C. A., McMichael, J. C., Zlotnick, G. W., and Green, B. A. (2004) Reduction of nasal colonization of nontypeable *Haemophilus influenzae* following intranasal immunization with rLP4/rLP6/UspA2 proteins combined with aqueous formulation of RC529, *Vaccine 22*, 3449−3456.
6. Oropeza, V. C., Page, M. E., and Van Bockstaele, E. J. (2005) Systemic administration of WIN 55,212-2 increases norepinephrine release in the rat frontal cortex, *Brain Res. 1046*, 45−54.
7. Kemmer, G., Reilly, T. J., Schmidt-Brauns, J., Zlotnik, G. W., Green, B. A., Fiske, M. J., Herbert, M., Kraiss, A., Schlor, S., Smith, A., and Reidl, J. (2001) NadN and e (P4) are essential for utilization of NAD and nicotinamide mononucleotide but not nicotinamide riboside in *Haemophilus influenzae*, *J. Bacteriol. 183*, 3974−3981.
8. Reidl, J., Schlor, S., Kraiss, A., Schmidt-Brauns, J., Kemmer, G., and Soleva, E. (2000) NADP and NAD utilization in *Haemophilus influenzae*, *Mol. Microbiol. 35*, 1573−1581.
9. Thaller, M. C., Schippa, S., and Rossolini, G. M. (1998) Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily, *Protein Sci. 7*, 1647−1652.
10. Passariello, C., Schippa, S., Iori, P., Berlutti, F., Thaller, M. C., and Rossolini, G. M. (2003) The molecular class C acid phosphatase of *Chryseobacterium meningosepticum* (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5′-nucleotides, *Biochim. Biophys. Acta 1648*, 203−209.
11. Malke, H. (1998) Cytoplasmic membrane lipoprotein LppC of *Streptococcus equisimilis* functions as an acid phosphatase, *Appl. Environ. Microbiol. 64*, 2439−2442.
12. Green, B. A., Farley, J. E., Quinn-Dey, T., Deich, R. A., and Zlotnick, G. W. (1991) The e (P4) outer membrane protein of *Haemophilus influenzae*: biologic activity of anti-e serum and cloning and sequencing of the structural gene, *Infect. Immun. 59*, 3191−3198.
13. Godlewska, R., Bujnicki, J. M., Ostrowski, J., and Jagusztyn-Krynicka, E. K. (2002) The hppA gene of *Helicobacter pylori* encodes the class C acid phosphatase precursor, *FEBS Lett. 525*, 39−42.
14. du Plessis, E. M., Theron, J., Joubert, L., Lotter, T., and Watson, T. G. (2002) Characterization of a phosphatase secreted by *Staphylococcus aureus* strain 154, a new member of the bacterial class C family of nonspecific acid phosphatases, *Syst. Appl. Microbiol. 25*, 21−30.
15. May, B. J., Zhang, Q., Li, L. L., Paustian, M. L., Whittam, T. S., and Kapur, V. (2001) Complete genomic sequence of *Pasteurella multocida*, Pm70, *Proc. Natl. Acad. Sci. U.S.A. 98*, 3460−3465.
16. Read, T. D., Peterson, S. N., Tourasse, N., Baillie, L. W., Paulsen, I. T., Nelson, K. E., Tettelin, H., Fouts, D. E., Eisen, J. A., Gill, S. R., Holtzapple, E. K., Okstad, O. A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J. F., Beanan, M. J., Dodson, R. J., Brinkac, L. M., Gwinn, M., DeBoy, R. T., Madpu, R., Daugherty, S. C., Durkin, A. S., Haft, D. H., Nelson, W. C., Peterson, J. D., Pop, M., Khouri, H. M., Radune, D., Benton, J. L., Mahamoud, Y., Jiang, L., Hance, I. R., Weidman, J. F., Berry, K. J., Plaut, R. D., Wolf, A. M., Watkins, K. L., Nierman, W. C., Hazen, A., Cline, R., Redmond, C., Thwaite, J. E., White, O., Salzberg, S. L., Thomason, B., Friedlander, A. M., Koehler, T. M., Hanna, P. C., Kolsto, A. B., and Fraser, C. M. (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria, *Nature 423*, 81−86.
17. Reilly, T. J., and Smith, A. L. (1999) Purification and characterization of a recombinant *Haemophilus influenzae* outer membrane phosphomonoesterase e (P4), *Protein Expression Purif. 17*, 401−409.
18. Reilly, T. J., Chance, D. L., and Smith, A. L. (1999) Outer membrane lipoprotein e (P4) of *Haemophilus influenzae* is a novel phosphomonoesterase, *J. Bacteriol. 181*, 6797−6805.
19. Ou, Z., Felts, R. L., Reilly, T. J., Nix, J. C., and Tanner, J. J. (2006) Crystallization of recombinant *Haemophilus influenzae* e (P4) acid phosphatase, *Acta Crystallogr. F62*, 464−466.
20. Pflugrath, J. W. (1999) The finer things in X-ray diffraction data collection, *Acta Crystallogr. D55*, 1718−1725.
21. Terwilliger, T. C. (2003) SOLVE and RESOLVE: automated structure solution and density modification, *Methods Enzymol. 374*, 22−37.
22. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2002) ARP/wARP's model-building algorithms. I. The main chain, *Acta Crystallogr. D58*, 968−975.
23. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D60*, 2126−2132.
24. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. F53*, 240−255.
25. DeLano, W. L. (2002) The PyMOL Molecular Graphics System (http://www.pymol.org).
26. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res. 28*, 235−242.
27. Dietmann, S., Park, J., Notredame, C., Heger, A., Lappe, M., and Holm, L. (2001) A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3, *Nucleic Acids Res. 29*, 55−57.
28. Calderone, V., Forleo, C., Benvenuti, M., Thaller, M. C., Rossolini, G. M., and Mangani, S. (2006) A structure-based proposal for the catalytic mechanism of the bacterial acid phosphatase AphA belonging to the DDDD superfamily of phosphohydrolases, *J. Mol. Biol. 355*, 708−721.
29. Galburt, E. A., Pelletier, J., Wilson, G., and Stoddard, B. L. (2002) Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase, *Structure 10*, 1249−1260.
30. Wallden, K., Ruzzenente, B., Rinaldo-Matthis, A., Bianchi, V., and Nordlund, P. (2005) Structural basis for substrate specificity of the human mitochondrial deoxyribonucleotidase, *Structure 13*, 1081−1088.
31. Peisach, E., Selengut, J. D., Dunaway-Mariano, D., and Allen, K. N. (2004) X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily, *Biochemistry 43*, 12770−12779.
32. Morais, M. C., Zhang, W., Baker, A. S., Zhang, G., Dunaway-Mariano, D., and Allen, K. N. (2000) The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily, *Biochemistry 39*, 10385−10396.
33. Allen, K. N., and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold, *Trends Biochem. Sci. 29*, 495−503.
34. Harding, M. M. (2004) The architecture of metal coordination groups in proteins, *Acta Crystallogr. D60*, 849−859.
35. Calderone, V., Forleo, C., Benvenuti, M., Cristina, Thaller, M., Maria, Rossolini, G., and Mangani, S. (2004) The first structure of a bacterial class B Acid phosphatase reveals further structural heterogeneity among phosphatases of the haloacid dehalogenase fold, *J. Mol. Biol. 335*, 761−773.
36. Krissinel, E., and Henick, K. (2005) Protein interfaces, surfaces and assemblies service PISA at European Bioinformatica Institute

(http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), *Comp. Life*, 163−174.
37. Reidl, J., and Mekalanos, J. J. (1996) Lipoprotein e(P4) is essential for hemin uptake by *Haemophilus influenzae*, *J. Exp. Med. 183*, 621−629.
38. Reilly, T. J., Green, B. A., Zlotnick, G. W., and Smith, A. L. (2001) Contribution of the DDDD motif of *H. influenzae* e (P4) to phosphomonoesterase activity and heme transport, *FEBS Lett. 494*, 19−23.
39. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for x-ray protein structure refinement, *Acta Crystallogr. A47*, 392−400.
40. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation, *Proteins 50*, 437−450.

BI701016M