Integrating Structure, Bioinformatics, and Enzymology to Discover Function

BioH, A NEW CARBOXYLESTERASE FROM ESCHERICHIA COLI

Received for publication, April 14, 2003
Published, JBC Papers in Press, May 5, 2003, DOI 10.1074/jbc.M303867200

Ruslan Sanishvili, Alexander F. Yakunin, Roman A. Laskowski, Tatiana Skarina,
Elena Evdokimova, Amanda Doherty-Kirby, Gilles A. Lajoie, Janet M. Thornton,
Cheryl H. Arrowsmith, Alexei Savchenko, Andrzej Joachimiak, and Aled M. Edwards

From the *Biosciences Division, Argonne National Laboratory, Argonne, Illinois, 60439, the †Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada, the ‡Clinical Genomics Centre/Proteomics, University Health Network, Toronto, Ontario M5G 1L7, Canada, the ‡European Bioinformatics Institute, Welcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom, and the †Department of Biochemistry, University of Western Ontario, London, Ontario N6A 5C1, Canada

Structural proteomics projects are generating three-dimensional structures of novel, uncharacterized proteins at an increasing rate. However, structure alone is often insufficient to deduce the specific biochemical function of a protein. Here we determined the function for a protein using a strategy that integrates structural and bioinformatics data with parallel experimental screening for enzymatic activity. BioH is involved in biotin biosynthesis in Escherichia coli and had no previously known biochemical function. The crystal structure of BioH was determined at 1.7 Å resolution. An automated procedure was used to compare the structure of BioH with structural templates from a variety of different enzyme active sites. This screen identified a catalytic triad (Ser82, His235, and Asp207) with a configuration similar to that of the catalytic triad of hydrolases. Analysis of BioH with a panel of hydrolase assays revealed a carboxylesterase activity with a preference for short acyl chain substrates. The combined use of structural bioinformatics with experimental screens for detecting enzyme activity could greatly enhance the rate at which function is determined from structure.

The protein complement of both prokaryotes and eukaryotes remains largely uncharacterized. At least 30% of all proteins have no known biochemical function, and a larger percentage have sequence similarity to proteins of known biochemical activity (e.g. most predicted protein kinases) but for which the physiological role is unknown. The challenge in the post-genomic era is to define both the biochemical and physiological functions of all proteins as rapidly as possible. Structural proteomics, the large scale determination of protein structure, is expected to provide insight into the fundamental mechanisms by which a protein sequence adopts a defined three-dimensional structure. Most of the organized efforts in structural proteomics (Ref. 1; rcsb.org/pdb/strucgen.html) specifically target protein sequences for which there is no known structural homologue in the public data bases at a level of 30% sequence identity. One aim of this effort is to more fully define the universe of protein folds. Importantly, because protein structure is often conserved in the absence of detectable sequence homology, the comparison of new protein structures with those of known proteins will likely provide clues to biochemical function.

The discovery of biochemical function from a new protein structure begins with automated searches for structural homologues of known function. The results of these comparisons are provided as lists with significance scores. The methods of comparison are now used routinely in the structural community and have proved invaluable for detecting structural conservation and for providing the basis for hypotheses (2). However, the interpretation of the results from structural comparisons often consumes a significant amount of time and is influenced by the extent to which the investigator is able to scour the literature.

In an effort to improve the process by which function is derived from structure, we have combined two methods to facilitate functional studies. First, we have employed a data base of structural templates derived from the active sites of 189 different classes of enzymes. This exploits the fact that the chemistry of the reaction restricts the types and the topological arrangement of the catalytic amino acids and hence results in strong conservation of their spatial arrangement, even where the protein folds are very different (3). By focusing on the catalytic moieties, functional similarities can be detected in cases where there is no similarity in sequence, fold, or secondary structure. Second, we have created and used a panel of generic biochemical assays to test the functional hypotheses raised by the structural comparisons. These assays are based...
on simple, often nonphysiological, substrates; the experiment is designed to reveal the chemistry of the active site and not the cellular substrate.

Here we present the results of the combined structural, bioinformatic, and enzymatic analysis of *Escherichia coli* BioH, a target within the Midwest Center for Structural Genomics (www.mcsg.anl.gov). By comparing the crystal structure of BioH with other known enzymes, we found that BioH is a member of the protein hydrolase superfamily and contains a classical Ser-His-Asp catalytic triad. A screen with different hydrolase substrates revealed that BioH has significant carboxylesterase activity, with a preference for short acyl chain substrates, and weak thioesterase activity. The strategy used for BioH might facilitate analysis of novel, uncharacterized proteins and structures arising form structural proteomics projects.

**EXPERIMENTAL PROCEDURES**

**BioH Expression and Purification**—The open reading frame of *bioH* was amplified by PCR from *E. coli* DH5α genomic DNA. The gene was cloned as previously described (4) into the NdeI and BamHI sites of a modified form of pET15b (Novagen) in which a TEV protease cleavage site was placed under the same conditions as the native protein except for the addition of 5 mM 2-mercaptoethanol to the purification buffers.

**Crystallographic Data Collection**—BioH was crystallized by vapor diffusion in hanging drops (ratio of 2 μl of protein to 2 μl of precipitant) equilibrated against reservoir containing 1.2 M sodium citrate trihydrate and 0.1 M Tris-HCl (pH 8.0). X-ray quality crystals grow at 21 °C in 2–5 days. For diffraction studies, the crystals were stabilized with the crystallization buffer supplemented with 15% ethylene glycol as a cryoprotectant and flash frozen in liquid nitrogen.

**Mass Spectrometry**—All of the mass spectrometry data were acquired and analyzed using Masslynx 3.5 (Micromass, Manchester, UK). Electrospray ionization mass spectrometry (ESI-MS) was performed on a Micromass Q-Tof2 mass spectrometer. Positive ion mode ESI-MS analysis of BioH was performed on a Micromass Q-Tof2 mass spectrometer (Micromass) using an m/z range of 500–4000. ESI-MS and MS/MS analysis of the low pH tryptic digest were performed on a Micromass Q-Tof2 mass spectrometer using nano-LC with a C18 column (0.3 × 5 mm; LC Packings). Data-dependent acquisition parameters were set to select the doubly and triply charged unmodified and modified precursor ions corresponding to residues 78–100 of the protein. MS-MS spectra were processed by base-line subtraction and deconvoluted using the MaxEnt3 module of MassLynx 3.5. The peptide sequences were determined semi-automatically from the resulting singly charged, deisotoped spectra using PepSeq, version 3.3 supplied with MassLynx 3.5.

**Enzyme Assays**—Rapid screening for enzyme activities were performed using the following procedures: (a) fatty acid esterase activity was measured spectrophotometrically at 37 °C using p-nitrophenyl (pNP) acetate or pNP esters of other fatty acids (C3–C18) as substrates (5), (b) thioesterase activity was measured spectrophotometrically using CoA thioesters of fatty acids (acyetyl-CoA, malonyl-CoA, and palmitoyl-CoA) as described earlier (6), (c) lipase activity (with sonicated olive oil as substrate) was measured spectrophotometrically by the copper soap assay after extraction of released free fatty acids with chloroform:heptane:methanol mixture (7), (d) protease activity was measured using t-leucine p-nitroanilide (aminopeptidase activity) or Na-benzoyl-L-arginine p-nitroanilide (trypsin-like endopeptidase activity) as described (8, 9), (e) phosphatase activity was determined spectrophotometrically using 5 mM p-nitrophenyl phosphate in 50 mM HEPES-KCl (pH 7.5) buffer at 37 °C (10), and (f) bromoperoxidase activity was measured spectrophotometrically with phenol red or monochlorodimedon as described previously (11).

**Crystallographic Data Collection**—A two-wavelength multiple-wavelength anomalous dispersion experiment was carried out on the 19ID line of the Structural Biology Center at Advanced Photon Source (Argonne, IL). All of the crystallographic data were collected at 110 K on one crystal containing selenomethionine-substituted protein. The crystal belongs to the tetragonal space group P4₁ with unit cell dimensions a = b = 75.2 Å, c = 49.3 Å, α = β = γ = 90°. The multiple-wavelength anomalous dispersion data set was collected using inverse beam strategy at the selenium absorption peak energy (0.97947 Å) and at a remote wavelength (0.95373 Å). The absorption edge was determined from the x-ray fluorescence spectrum and the f’ and f” plots versus energy obtained with the program CHOOCH (12). High resolution data were collected from the unexposed part of the same crystal, which had been stored in liquid nitrogen. All of the data were measured with the CCD detector (13) 210 × 210 mm² sensitive area and fast duty cycle. Control of the experiment, data collection and visualization was done with d*TREK (14), and all of the data were integrated and scaled with the program package HKL2000 (15). Some of the basic statistics of data collection and processing are given in Table I.

**Structure Determination**—Multiple-wavelength anomalous dispersion phasing of BioH data was carried out with the program CNS (16). Experimental phases were extended from 2.5 to 2.0 Å resolution with density modification, using data collected at the f’’ peak wavelength. With these improved phases, the initial model was built with the program ARP/wARP (17). The high quality of the phases allowed 94% of the main chain to be built automatically and most of the side chains to be placed with a confidence level of 79%. The remainder of the model was built, and all of the side chains were corrected manually using the program O (18). This model was then refined against the 1.7 Å resolution data with several macro cycles of CNS, including simulated annealing, B-factor, and positional refinements. After each macro cycle, the model was inspected, and corrections and/or additions were made


**RESULTS AND DISCUSSION**

**The BioH Crystal Structure**—The final model of the BioH crystal structure consists of 256 residues, two molecules of ethylene glycol, and 240 water molecules. The last two residues of the model, Gly rarity and Ser rarity, were appended to the protein as a result of the cloning strategy. The first two residues of the native sequence, Met 1 and Asn 2, were not included because of the absence of the corresponding electron density. The two molecules of ethylene glycol from the cryoprotectant were bound to the protein molecule mostly because of hydrophobic interactions. Residue 100 was unambiguously identified as Arg, instead of Gln, in electron density maps and is likely a PCR-induced mutation. The side chains of residues Glu 116, Lys 121, Asp 183, Phe 186, Glu 187, and Lys 213 have incomplete electron density.

BioH is a two-domain protein (Fig. 1A). The α/β/α three layer sandwich of the large domain (residues 5–109 and 188–256; see below) consists of a twisted β-sheets formed by seven mostly parallel strands β1 (residues 5–9), β3 (residues 14–19), β4 (residues 110–105), β5 (residues 101–105), β6 (residues 189–203) and β7 (residues 225–230) and flanked on both sides by five α-helices α1 (residues 31–39), α2 (residues 60–70), α3 (residues 83–94), α8 (residues 215–222), and α9 (residues 237–252). Ile 82 and Pro 235 introduce ~90° kinks in the first and last helices, respectively. This domain resembles the Rossmann fold, which is commonly found in enzymes.

A small auxiliary domain is formed by the C-terminal segment of the polypeptide chain (Cys 110–Asp 187) and is inserted into the catalytic domain. The auxiliary domain contains four α-helices, residues 122–134 (α4), 136–145 (α5), 155–166 (α6), and 173–185 (α7), that create a bundle of two V-shaped bends (Fig. 1B). The two domains are connected by a hinge region near Cys 110 and Asp 187. The interface between domains is stabilized by multiple hydrophobic interactions including helices α6 and α7 that run across the surface of catalytic domain and intramolecular hydrogen bonds between the carbonyl of Pro 109 and the nitrogen of Leu 188 and two hydrogen bonds between Asp 187 and Arg 189.

**Automated Structural Bioinformatics Reveals a Ser-His-Asp Catalytic Triad**—One of the aims of structural proteomics is to perform more comprehensive automated analysis of protein structures to reduce the level of time-intensive human intervention. To screen new structures for potential catalytic functions, we have created a data base of ~189 three-dimensional enzyme active site structural templates. The BioH structure was scanned against this data base of using the TESS program (3). This automated search gave a close match of BioH to the Ser-His-Asp catalytic triad of lipases (21) (EC 3.1.1.3). The BioH residues involved (Ser 82, His 235, and Asp 239) matched the template with a root mean square deviation of 0.28 Å for the overlaid side chains (Fig. 2). This is well within the cut-off of 1.2 Å used for discriminating true from false matches for this template. The presence of the catalytic triad suggested that BioH might possess lipase, protease, or esterase activity. Furthermore, the serine nucleophile (Ser 82) is located within one of the two earlier identified Gly-Xaa-Ser-Xaa-Gly motifs (22), which is typical for acyltransferases and thioesterases.

The structure of BioH was also compared with all other known structures using conventional methods such as the DALI algorithm (23). The results from the DALI search revealed structural homology to a large number of proteins with a broad range of enzymatic functions. The closest matches with strong structural similarities include a bromoperoxidase (EC 1.11.1.10; Z score, 22.6; Protein Data Bank code lbr), an amionopeptidase (EC 3.4.11.5; Z score, 21.1; Protein Data Bank code 1qtr), two epoxide hydrolases (EC 3.3.2.3; Z scores, 20.5 and 18.2; Protein Data Bank codes 1ehy and lcr6, respectively), two halokane dehalogenases (EC 3.8.1.5; Z scores, 20.2 and 16.2; Protein Data Bank codes 1bn6 and 1bq6, respectively), and a lyase (EC 4.2.1.39; Z score, 17.2; Protein Data Bank code qj4).

**TABLE II**

| Phasing wavelength | Resolution | Number of reflections | Phasing power | FOM* |
|--------------------|------------|-----------------------|---------------|------|
| Peak               | 42–2.5     | 18206 (2027)          | 2.92 (2.28)   | 0.50 (0.44) |
| Remote             | 42–2.5     | 18210 (2024)          | 2.18 (1.67)   | 0.41 (0.35) |
| Overall            | 42–2.5     | 18414 (2047)          | 2.92 (2.28)   | 0.50 (0.44) |

**Refinement**

| Resolution (Å) | Number of reflections | R factor (%) | R free (%) | Correlation | Correlation free | Mean B factor |
|----------------|-----------------------|--------------|------------|-------------|-----------------|--------------|
| 75–1.70        | 27141 (3621)          | 14.7 (21.2)  | 18.9 (24.6) | 97.1        | 95.2            | 15.13        |
| 75–1.70        | 27141 (3621)          | 14.7 (21.2)  | 18.9 (24.6) | 97.1        | 95.2            | 15.13        |

**Deviations from ideal**

| Deviations from ideal | Refined | Target |
|-----------------------|---------|--------|
| Covalent bonds        | 0.022   | 0.021  |
| Bond angles           | 1.905   | 1.950  |
| Planarity             | 0.011   | 0.020  |
| Chiral centers        | 0.131   | 0.20   |
| Torsion angle 1       | 6.4     | 5.0    |
| Torsion angle 3       | 18.74   | 15.0   |
| VDW contacts          | 0.261   | 0.20   |

* Last resolution shell.
A comparison of BioH with a chloroperoxidase (EC 1.11.1.10) is shown in Fig. 3. The sequence identities between BioH and these proteins range from 15 to 25% and therefore do not suggest a specific catalytic function for BioH. Further manual analysis of these enzymes and literature review would have revealed to the expert that each contains a Ser-His-Asp catalytic triad in their active sites.

**Ser82 Is Covalently Modified by a Hydrolase Inhibitor**

The structuralinformatics provided initial evidence for the location of the BioH catalytic site. The experimental density maps also showed an unusual feature that extended from the side chain of Ser82 (Fig. 4). The shape of the density and its environment, insinuated that the corresponding compound was covalently linked to the enzyme.

**FIG. 2. Superposition of the Ser82, His235, and Asp207 residues onto the catalytic triad template.** The template side chains are depicted by the thicker, transparent bonds, whereas the BioH residues are represented by the thinner, solid bonds and include the main chain atoms. The root mean square deviation between equivalent atoms in the template and matched side chains is 0.28 Å.

**FIG. 3. Superposition of the E. coli BioH (magenta) and Streptomyces aureofaciens chloroperoxidase (yellow) structures; superposition of the catalytic domains, including the loops connecting secondary structure elements.** The overall architecture of the auxiliary domains is the same for both proteins, although the placement of the helices, especially of α5 and α6, relative to the catalytic domain differs.

A comparison of BioH with a chloroperoxidase (EC 1.11.1.10) is shown in Fig. 3. The sequence identities between BioH and these proteins range from 15 to 25% and therefore do not suggest a specific catalytic function for BioH. Further manual analysis of these enzymes and literature review would have revealed to the expert that each contains a Ser-His-Asp catalytic triad in their active sites.

**Ser82 Is Covalently Modified by a Hydrolase Inhibitor**—The structuralinformatics provided initial evidence for the location of the BioH catalytic site. The experimental density maps also showed an unusual feature that extended from the side chain of Ser82 (Fig. 4). The shape of the density and its environment, insinuated that the corresponding compound was covalently linked to the enzyme.
attached to the O\(^{-}\) atom of Ser\(^{82}\) and formed hydrogen bonds with the backbone nitrogens of Trp\(^{22}\) and Leu\(^{83}\) (oxyanion hole) are shown with magenta dashed lines and numbers.

For crystallographic experiments and initial MALDI and ESI-MS, the BioH protein was purified in the presence of PMSF (32) using the BioC and BioH proteins (33), whose exact biochemical roles have not been elucidated. The chemical properties compared with the other enzymes. Compared with BioH, YbaC and TesA exhibit higher affinities for the long chain fatty acids, pNP-ocanatoe (C8) and pNP-decanoate (C10). The specific activity of BioH for short C2 or C3 substrates was in the same range as for YbaC and at least 10–30 times lower as compared with TesA. Both BioH and TesA also displayed thioesterase activity with palmitoyl-CoA (however 13 times lower for BioH) but show no activity with acetyl-CoA as a substrate. The ratio of carboxylesterase/thioesterase activities (with pNP-palmitate/palmitoyl-CoA) was 0.3 for TesA and 1.3 for BioH.

The specificity for the short chain fatty acid esters likely arises from the fact that the catalytic site of BioH is buried between two domains (Fig. 1) and is not readily accessible for bulkier compounds. Substrates with acyl chain length of up to 6 carbons (C-2–C-6) could be accommodated within the hydrophobic crevice in the V-shaped cap domain of BioH where the invariant Phe\(^{143}\) (Fig. 1B) can act as a facilitator of binding. In fact, the walls of the active site are quite hydrophobic; therefore binding of acyl substrates to BioH is likely to be mediated mostly by hydrophobic interactions, and the active site is sufficiently large to accommodate short chain substrates with very similar affinities for the C-2–C-6 range. This is consistent with the observation that BioH shows essentially same \(K_m\) for C-2–C-6 substrates (Table III).

**A Possible Role for BioH in Biotin Biosynthesis**—In microorganisms and plants, biotin is synthesized from pimeloyl-CoA by the enzymes BioF, BioA, BioD, and BioB in a conserved four-step reaction (29–31). In the Gram-negative bacteria, such as *E. coli*, pimeloyl-CoA is produced from \(\alpha\)-alanine and/or acetate (32) using the BioC and BioH proteins (33), whose exact biochemical roles have not been elucidated. The *bioC* gene is

[![Fig. 4. Experimental electron density maps after density modification. Residues of the catalytic triad, Ser\(^{82}\), Asp\(^{207}\), and His\(^{335}\) of the refined model are labeled along with Trp\(^{22}\), Leu\(^{83}\), and several solvent molecules (labeled with numbers only). Note the additional electron density extending from the O\(^{-}\) atom of Ser\(^{82}\). According to biochemical data, this density was interpreted as PMSF, but only parts of it are visible because of disorder and partial occupancy. Several functionally important hydrogen bonds, including those between the sulfonate oxygen of the inhibitor and backbone nitrogens of Trp\(^{22}\) and Leu\(^{83}\) (oxyanion hole) are shown with magenta dashed lines and numbers.](http://www.jbc.org/)

---

_E. coli BioH_ 26043

In the absence of PMSF was subjected to several enzymatic assays that focused on hydrolase function including carboxylesterase, lipase, thioesterase, phosphatase, endopeptidase, aminopeptidase, and bromoperoxidase. BioH demonstrated significant carboxylesterase activity (Table III) (EC 3.1.1.1) and hydrolyzed \(\beta\)-nitrophenyl esters of fatty acids. The enzyme showed rather narrow pH optimum (8.0–8.5) and broad substrate specificity with a preference for short chain substrates (Fig. 5). The kinetic parameters of BioH were determined for several substrates (Table III). These results demonstrate that although BioH was most active with pNP-acetate, the \(K_m\) for all C-2–C-6 substrates was essentially the same. In agreement with the results of the mass spectrometry and crystallography, BioH was strongly inhibited by PMSF (10.5% of residual activity after 10 min of incubation with 2 mM PMSF). Purified BioH showed classical Michaelis-Menten kinetics, and linear double reciprocal plots were obtained for all of the pNP substrates tested (data not shown).

Purified BioH showed low enzymatic activities for thioesterase (using palmitoyl-CoA as a substrate; 186.5 ± 18.6 nmol/min/mg protein), lipase (using olive oil; 18.5 ± 1.3 nmol/min/mg protein), and aminopeptidase (using leucine-\(p\)-anilide as a substrate; 3.8 nmol/min/mg protein) and showed no detectable enzymatic activity for phosphatase (using \(\beta\)-nitrophenyl phosphate as a substrate), trypsin-like endopeptidase (using benzo-\(y\)-arginine-\(p\)-nitroanilide as a substrate), or bromoperoxidase (phenol red and monochlorodimedone as potential substrates).

Our data combined with results reported in the literature suggest that BioH represents a novel carboxylesterase in *E. coli*. *E. coli* is known to express at least three other proteins with carboxylesterase activity: carboxylesterase YbaC (25), thioesterase TesA (26, 27), and thioesterase TesB (28). BioH shows no significant sequence similarity with these enzymes (data not shown). BioH also possessed different enzymological properties compared with the other enzymes. Compared with BioH, YbaC and TesA exhibit higher affinities for the long chain fatty acids, pNP-ocanatoe (C8) and pNP-decanoate (C10). The specific activity of BioH for short C2 or C3 substrates was in the same range as for YbaC and at least 10–30 times lower as compared with TesA. Both BioH and TesA also displayed thioesterase activity with palmitoyl-CoA (however 13 times lower for BioH) but show no activity with acetyl-CoA as a substrate. The ratio of carboxylesterase/thioesterase activities (with pNP-palmitate/palmitoyl-CoA) was 0.3 for TesA and 1.3 for BioH.

The specificity for the short chain fatty acid esters likely arises from the fact that the catalytic site of BioH is buried between two domains (Fig. 1) and is not readily accessible for bulkier compounds. Substrates with acyl chain length of up to 6 carbons (C-2–C-6) could be accommodated within the hydrophobic crevice in the V-shaped cap domain of BioH where the invariant Phe\(^{143}\) (Fig. 1B) can act as a facilitator of binding. In fact, the walls of the active site are quite hydrophobic; therefore binding of acyl substrates to BioH is likely to be mediated mostly by hydrophobic interactions, and the active site is sufficiently large to accommodate short chain substrates with very similar affinities for the C-2–C-6 range. This is consistent with the observation that BioH shows essentially same \(K_m\) for C-2–C-6 substrates (Table III).
which is proposed to function as a specific pimeloyl-acyl-carrier protein in the synthesis of pimeloyl-CoA (22), may interact with BioH and facilitate the delivery of a pimeloyl unit to the BioH catalytic site.

**Perspective**—Three-dimensional structures are now being generated for many proteins of unknown function. In many cases, such as for BioH, the structural data combined with existing clues in the literature, or even the intuition of an experienced investigator, can point the experimentalist in the right direction to identify and confirm biochemical function. However, as structural proteomics efforts gain momentum, there will be an increase in the number of protein structures for which there is no existing body of literature. The annotation of these proteins will demand methods that do not depend on specialists who are experts in a specific area of biology. The three-dimensional structure of BioH was analyzed using several automated methods for structural comparison and also with a series of generic enzymatic assays. This approach enabled us to rapidly characterize BioH enzymatic activity and reveal a new enzyme in *E. coli*. The development and refinement of these combined methods will significantly increase the value of structural genomics/proteomics results in the future by assigning biochemical or enzymatic functions to complement the structural information.

**Acknowledgments**—We thank all members of the Structural Biology Center at Argonne National Laboratory and of the Ontario Centre for Structural Proteomics for help in conducting experiments and James J. De Voss for helpful discussions regarding BioH.

**REFERENCES**

1. Stevens, R. C., Yokoyama, S., and Wilson, I. A. (2001) Science 294, 89–92.

2. Zarembski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R., and Kim, S.-H. (1988) Proc. Natl. Acad. Sci. U. S. A. 95, 15189–15193.

3. Wallace, A. C., Borkakoti, N., and Thornton, J. M. (1997) Protein Sci. 6, 2308–2323.

4. Zhang, R.-G., Skarina, T., Kats, J. E., Beasley, S., Khachatryan, A., Vyas, S., Arrowsmith, C. H., Clarke, S., Edwards, A., Joachimiak, A., and Savchenko, A. (2001) Structure 9, 1055–1106.

5. Vorderwülbecke, T., Kieslich, K., and Erdmann, H. (1999) Enzyme Microb. Technol. 14, 631–639.

6. Berge, R. K., and Farstad, M. (1981) Methods Enzymol. 71, 234–242.

7. Nixon, M., and Chal, S. H. P. (1979) Anal. Biochem. 97, 403–409.

8. Bienvenue, D. L., Mathew, R. S., Ringe, D., and Holz, R. C. (2002) J. Biol. Inorg. Chem. 7, 129–135.

9. Gan, Z., Marquardt, R. B., and Xiao, H. (1999) Anal. Biochem. 268, 151–156.

10. Kuo, M.-H., and Blumenthal, H. J. (1961) J. Biol. Chem. 236, 1033–1039.

11. Pelletier, I., and Altenbuchner, J. (1995) Microbiology 141, 459–468.

12. Evans, G., and Penther, R. F. (2001) J. Appl. Crystallogr. 34, 82–86.

13. Westbrook, E. M., and Naday, I. (1997) Methods Enzymol. 276, 244–268.

14. Pfugrath, J. W. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1718–1725.

15. Ozonowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326.

16. Brunguer, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosser-Kunstein, R. W., Kuszewski, J., Nilges, M., Pannu, N., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921.

17. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Nat. Struct. Biol. 6, 458–463.

18. Jones, T. A., Zou, J.-Y., Cowan S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. A 47, 110–119.

19. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255.

20. Collaborative Computational Project 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763.

21. Schrag, J. D., Li, Y., Cygler, M., Lang, D., Burgdorf, T., Hecht, H. J., Schmid, R., Schomburg, D., Rydel, T. J., Oliver, J. D., Strickland, L. C., Dunaway, C. M., Larson, S. B., Day, J., and McPherson, A. (1997) Structure 5, 187–202.

22. Lemoine, Y., Wach, A., and Jeltsch, J. M. (1996) Mol. Microbiol. 19, 645–647.

**TABLE III**

| Substrate                  | $K_{cat}$ | $k_{cat}$ | $K_m$ | $k_{cat}/K_m$ |
|----------------------------|-----------|-----------|-------|---------------|
| pNP-acetate (C2)           | 0.29 ± 0.04 | 18.5 ± 1.7 | 63.8 × 10^3 |
| pNP-propionate (C3)       | 0.35 ± 0.08 | 13.1 ± 1.2 | 37.4 × 10^3 |
| pNP-butyrate (C4)         | 0.32 ± 0.06 | 6.1 ± 0.7  | 18.5 × 10^3 |
| pNP-caproate (C6)         | 0.25 ± 0.02 | 4.0 ± 0.1  | 16.0 × 10^3 |
| pNP-laurate (C12)         | 0.60 ± 0.13 | 1.5 ± 0.2  | 2.5 × 10^2  |
23. Holm, L., and Sander, C. (1994) Proteins 19, 165–173
24. Gold, A. M. (1965) Biochemistry 4, 897–901
25. Kanaya, S., Koyanagi, T., and Kanaya, E. (1998) Biochem. J. 332, 75–80
26. Bronner, W. M., and Bloch, K. (1972) J. Biol. Chem. 247, 3123–3133
27. Lee, Y.-L., Chen, J. C., and Shaw, J.-F. (1997) Biochem. Biophys. Res. Commun. 231, 452–456
28. Naggert, J., Narasimhan, M. L., DeVeaux, L., Cho, H., Randhawa, Z. L., Cronan, J. E., Jr., Green, B. N., and Smith, S. (1991) J. Biol. Chem. 266, 11044–11050
29. Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. (1988) J. Biol. Chem. 263, 6461–6464
30. Baldet, P., Alban, C., and Deuce, R. (1997) in Methods Enzymol. 279, 327–339
31. Demoll, E. (1996) in Escherichia coli and Salmonella: Cellular and Molecular Biology (Neidhardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., eds) pp. 704–709, ASM Press, Washington, D.C.
32. Ifuku, O., Miyaoka, H., Koga, N., Kishimoto, J., Haze, S., Washi, Y., and Kajiwara, M. (1994) Eur. J. Biochem. 220, 585–591
33. Barker, D. F., and Campbell, A. M. (1980) J. Bacteriol. 143, 789–800
34. Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2002) Genome Res. 12, 1507–1516
35. Gloeckler, R., Ohsawa, I., Speck, D., Ledoux, C., Bernard, S., Zinsius, M., Villeval, D., Kiseu, T., Kamegawa, K., and Lemoine, Y. (1996) Gene (Amst.) 87, 63–70
36. Bower, S., Perkins, J. B., Yocum, R. R., Howitt, C. L., Rahaim, P., and Pero, J. (1996) J. Bacteriol. 178, 4122–4130
37. Sanyal, I., Lee, S. L., and Flint, D. (1994) J. Am. Chem. Soc. 116, 2637–2638
38. Lezius, A., Ringelman, E., and Lynen, F. (1963) Biochem. Z. 336, 510–525
39. Tomczyk, N. H., Nettleship, J. E., Baxter, R. L., Crichton, H., Webster, S. P., and Campopiano, D. J. (2002) FEBS Lett. 513, 299–304
Integrating Structure, Bioinformatics, and Enzymology to Discover Function: BioH, A NEW CARBOXYLESTERASE FROM ESCHERICHIA COLI
Ruslan Sanishvili, Alexander F. Yakunin, Roman A. Laskowski, Tatiana Skarina, Elena Evdokimova, Amanda Doherty-Kirby, Gilles A. Lajoie, Janet M. Thornton, Cheryl H. Arrowsmith, Alexei Savchenko, Andrzej Joachimiak and Aled M. Edwards

J. Biol. Chem. 2003, 278:26039-26045.
doi: 10.1074/jbc.M303867200 originally published online May 5, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M303867200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 33 references, 7 of which can be accessed free at http://www.jbc.org/content/278/28/26039.full.html#ref-list-1