Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structural insights into the efficient CO_2-reducing activity of an NAD-dependent formate dehydrogenase from *Thiobacillus* sp. KNK65MA

Hyunjun Choe,[a]‡ Jung Min Ha,[b]‡ Jeong Chan Joo,[a]§ Hyunook Kim,[c] Hye-Jin Yoon,[d] Seonghoon Kim,[c] Sang Hyeon Son,[b] Robert M. Gengan,[e] Seung Taeg Jeon,[b] Rakwoo Chang,[c] Kwang Deog Jung,[f] Yong Hwan Kim[a]* and Hyung Ho Lee[d]*

[a]Department of Chemical Engineering, Kwangwoon University, Seoul 139-701, Republic of Korea, [b]Department of Bio and Nano Chemistry, Kookmin University, Seoul 136-702, Republic of Korea, [c]Department of Chemistry, Kwangwoon University, Seoul 139-701, Republic of Korea, [d]Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, Republic of Korea, [e]Department of Chemistry, Faculty of Applied Sciences, Durban University of Technology, Durban, South Africa, and [f]Energy Research Center, Korea Institute of Science and Technology, Seoul, Republic of Korea

‡ These authors contributed equally to this work.
§ Current address: Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada.

Correspondence e-mail: metalkim@kw.ac.kr, hyungholee@snu.ac.kr

Received 29 September 2014
Accepted 20 November 2014

PDB reference:
NAD-dependent formate dehydrogenase, 3wr5

CO_2 fixation is thought to be one of the key factors in mitigating global warming. Of the various methods for removing CO_2, the NAD-dependent formate dehydrogenase from *Candida boidinii* (CbFDH) has been widely used in various biological CO_2-reduction systems; however, practical applications of CbFDH have often been impeded owing to its low CO_2-reducing activity. It has recently been demonstrated that the NAD-dependent formate dehydrogenase from *Thiobacillus* sp. KNK65MA (TsFDH) has a higher CO_2-reducing activity compared with CbFDH. The crystal structure of TsFDH revealed that the biological unit in the asymmetric unit has two conformations, *i.e.* open (NAD^+-unbound) and closed (NAD^+-bound) forms. Three major differences are observed in the crystal structures of TsFDH and CbFDH. Firstly, hole 2 in TsFDH is blocked by helix $\alpha 20$, whereas it is not blocked in CbFDH. Secondly, the sizes of holes 1 and 2 are larger in TsFDH than in CbFDH. Thirdly, Lys287 in TsFDH, which is crucial for the capture of formate and its subsequent delivery to the active site, is an alanine in CbFDH. A computational simulation suggested that the higher CO_2-reducing activity of TsFDH is owing to its lower free-energy barrier to CO_2 reduction than in CbFDH.

1. Introduction

The removal of carbon dioxide (CO_2) from the atmosphere as a means of relieving global warming is a challenging endeavour. One approach for removing CO_2 is to convert it into useful chemicals such as formate by using enzymes such as formate dehydrogenase (FDH). Formate can be sequentially reduced to formaldehyde and methanol by coupling the aldehyde dehydrogenase and alcohol dehydrogenase reactions (El-Zahab *et al.*, 2008) or used to provide reducing power for chemical synthesis through the formate-oxidation reaction (Bommarius *et al.*, 1998).

Metal-dependent enzymes are known to exhibit high CO_2-reduction activities (Almendra *et al.*, 1999; Graentzdoerffer *et al.*, 2003; Reda *et al.*, 2008). Schuchmann and Müller reported that a hydrogen-dependent carbon dioxide reductase from *Acetobacterium woodii* can produce formate from H_2 and CO_2 or from syngas (Schuchmann & Müller, 2013). Alissandratos and coworkers reported that engineered *Escherichia coli* cells with oxygen-labile FDHs, including FDH from the anaerobic archaeon *Pyrococcus furiosus*, can convert H_2 and CO_2 into formate (Alissandratos *et al.*, 2014). However, practical applications of these enzymes are impeded by their high oxygen sensitivity. Although recently a few less oxygen-sensitive FDHs have been discovered, these enzymes still

doi:10.1107/S1399004714025474

require complicated catalytic components, *e.g.* a molybdenum cofactor, a selenocysteine or an iron–sulfur cluster, and thus no structure of these enzymes is available to date (Alissandratos *et al.*, 2013; Hartmann & Leimkühler, 2013). Compared with metal-dependent CO_2-reducing enzymes, NAD-dependent FDHs are oxygen-insensitive and do not require complicated catalytic components. In particular, *Candida boidinii* (CbFDH) has been widely used as a biocatalyst for CO_2 reduction (El-Zahab *et al.*, 2008; Yadav *et al.*, 2012) owing to its stability and its ease of expression in *E. coli* (Choe *et al.*, 2014); however, its low catalytic activity (Schirwitz *et al.*, 2007) hampers its viability. It has been challenging to find a practical enzyme for CO_2 reduction.

The crystal structures of several NAD-dependent FDHs have been determined. The first NAD^+–FDH structure was solved from *Pseudomonas* sp. 101 (PsFDH; Lamzin *et al.*, 1992). The apo and holo structures of *Moraxella* sp. C-1 FDH (MsFDH) have been solved, which helped in understanding the mechanism underlying interdomain movement: the C-terminal residues 374–399 are crucial for the transition from the open to the closed conformation (Shabalin *et al.*, 2009). In addition, the crystal structure of CbFDH was determined by mutating two residues on its surface to facilitate crystallization, and these mutations were subsequently shown to improve both the stability and the catalytic activity of the enzyme (Schirwitz *et al.*, 2007). Although these studies characterized the overall architecture of FDHs, until now no explanation of the CO_2-reducing activity has been presented based on the crystal structure of an FDH. We recently reported TsFDH as a CO_2-reducing biocatalyst (Choe *et al.*, 2014) that exhibits a 5.8-fold higher formate-production rate and an 84.2-fold higher ratio of CO_2 reduction to formate oxidation in catalytic efficiency ($k_{\rm cat}/K_{\rm B}$) compared with CbFDH. A structural understanding of TsFDH is crucial for protein engineering to further improve the CO_2-reduction efficiency and to provide insights into its high activity. However, a lack of structural information limited further investigation of TsFDH.

In this study, we determined the crystal structure of TsFDH both with and without NAD^+ and offer insight into the structural basis for the increased CO_2-reduction efficiency of TsFDH. Structural comparisons of TsFDH and CbFDH and a computer simulation of CO_2 reduction revealed that they have discernible differences in structure and in the free-energy barriers to the reactions.

2. Materials and methods

2.1. Activity assay of FDHs

CO_2-reducing activities were determined with recombinant FDHs prepared as described previously (Choe *et al.*, 2014). The absorbance change at 340 nm was monitored during the reaction. One unit of reduction activity was defined as the amount of enzyme required to consume 1 µmol NADH per minute under standard conditions. Sodium phosphate buffer (100 m*M*) was purged with CO_2 followed by readjusting the pH to 7.0 with NaOH. The CO_2-reduction reaction was performed using 1 ml of the CO_2-saturated buffer containing 1.21×10^{-8} mol FDH (0.50 mg CbFDH or 0.55 mg TsFDH) and 0.15 m*M* NADH. A screw-cap cuvette was used to prevent the escape of CO_2 gas from the solution during the reaction. Enzyme assays were performed in triplicate, and the activities were calculated by subtracting the decomposition rate of NADH in the absence of the enzyme.

2.2. Expression and purification of TsFDH for crystallization

The full-length *fdh* gene from *Thiobacillus* sp. strain KNK65MA was amplified by PCR using a codon-optimized synthetic gene as a template. The synthetic gene was synthesized by GenScript (Piscataway, New Jersey, USA) and the sequence of the synthetic gene was derived from BAC92737.1. The amplified gene was cleaved using BamHI and XhoI, and was then introduced into BamHI/XhoI-digested pGST-parallel2 vector (Sheffield *et al.*, 1999). TsFDH protein containing an N-terminal glutathione *S*-transferase (GST) tag was overexpressed in *E. coli* BL21 (λDE3) cells. The cells were grown in Luria–Bertani broth with 100 µg µl^{-1} ampicillin at 310 K until the culture reached an OD_{600} of 0.8–1.0. The cells were then induced with 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 298 K for 18 h. The cells were lysed with a microfluidizer (Microfluidics, USA) in lysis buffer (20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl) containing 1 m*M* phenylmethylsulfonyl fluoride (PMSF). The lysed cells were centrifuged at 12 000 rev min^{-1} (Vision V506CA rotor) for 30 min at 277 K to pellet the cell debris, and the supernatant was then applied onto a glutathione Sepharose column (GE Healthcare, USA) pre-equilibrated with lysis buffer. The protein was eluted with elution buffer (150 m*M* Tris–HCl pH 8.0, 15 m*M* reduced glutathione) and then cleaved using TEV protease to separate the GST tag from TsFDH. The eluate was further purified by gel filtration on a HiLoad 16/60 Superdex 200 column (GE Healthcare) pre-equilibrated with lysis buffer. The eluted TsFDH protein was reapplied onto the glutathione Sepharose column pre-equilibrated with lysis buffer to remove both the cleaved GST and the uncleaved GST-TsFDH protein. The flowthrough was concentrated to 12 mg ml^{-1} by ultrafiltration. The extinction coefficient of TsFDH was calculated with *ProtParam* (http://web.expasy.org/protparam/).

2.3. Crystallization and data collection

Prior to crystallization, the TsFDH protein was mixed with an NAD^+ solution in a 1:30 molar ratio and incubated for 1 h on ice. Crystals of the TsFDH–NAD^+ complex were grown at 298 K using the sitting-drop method by mixing 1 µl of a 12 mg ml^{-1} solution of TsFDH in 20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl, 1 m*M* dithiothreitol with 1 µl of a reservoir solution consisting of 0.1 *M* HEPES pH 7.0, 25%(*w*/*v*) polyethylene glycol 6000, 1.2 *M* lithium chloride. The crystals were then transferred to a solution comprised of the reservoir solution with 7% glycerol for cryoprotection. Data were collected at 100 K using a 1° oscillation range on beamline

Table 1
Statistics of data collection and refinement for partially NAD$^+$-bound TsFDH.

Values in parentheses are for the highest resolution shell.

Data-collection statistics	
X-ray source	BL-26B1, SPring-8
X-ray wavelength (Å)	0.8000
Space group	$P2_12_12_1$
a, b, c (Å)	55.9, 72.7, 386.8
α, β, γ (°)	90.00
Resolution range (Å)	50–2.14 (2.18–2.14)
Total/unique reflections	491705/88104
Completeness (%)	99.9 (100.0)
$\langle I/\sigma(I)\rangle$	21.3 (4.3)
$R_{\rm merge}$† (%)	11.4 (52.4)
$R_{\rm p.i.m.}$‡ (%)	5.4 (24.6)
Model-refinement statistics	
Resolution range (Å)	42.3–2.14
$R_{\rm work}/R_{\rm free}$§ (%)	17.2/20.9
No. of atoms	
Chain *A*	3098
Chain *B*	3055
Chain *C*	3063
Chain *D*	2931
Water O atoms	972
Average *B* factors (Å^2)	
Chain *A*	20.2
Chain *B*	23.5
Chain *C*	29.5
Chain *D*	37.5
Water O atoms	32.3
Ligand molecules	
NAD$^+$, chain *A*	16.0
NAD$^+$, chain *B*	17.7
NAD$^+$, chain *C*	22.1
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.009
Bond angles (°)	1.355
Ramachandran plot	
Favoured	99.2% [1548/1561]
Allowed	0.8% [13/1561]
Outliers	0.0% [0/1561]

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum_{hkl}$ is the sum over all reflections, and $\sum_i$ is the sum over i measurements of reflection hkl. ‡ $R_{\rm p.i.m.} = \sum_{hkl}\{1/[N(hkl) - 1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$. § $R = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl}|F_{\rm obs}|$, where $R_{\rm free}$ was calculated for a randomly chosen 10% of reflections, which were not used for structure refinement, and $R_{\rm work}$ was calculated for the remaining reflections.

26B1 at SPring-8, Hyogo, Japan. The TsFDH–NAD$^+$ crystals diffracted to a resolution of 2.14 Å and the diffraction data were processed and scaled using the *HKL*-2000 software package (Otwinowski & Minor, 1997). The crystal belonged to space group $P2_12_12_1$, with unit-cell parameters a = 55.9, b = 72.7, c = 386.8 Å.

2.4. Structure determination and refinement

The structure was determined by molecular replacement using PsFDH (PDB entry 2nad; Lamzin *et al.*, 1994) as a probe. A cross-rotational search and subsequent translational search were performed using *Phaser* (McCoy *et al.*, 2007). Subsequent manual model building was performed using *Coot* (Emsley & Cowtan, 2004), and restrained refinement was performed using *REFMAC*5 (Murshudov *et al.*, 2011) and *PHENIX* (Adams *et al.*, 2010). Several rounds of model building, simulated annealing, positional refinement and individual *B*-factor refinement were performed using *Coot* and *REFMAC*5. Table 1 lists the refinement statistics.

2.5. Analytical gel filtration

The purified TsFDH was subjected to analytical gel-filtration chromatography on a Superdex 200 10/300 GL column (GE Healthcare) with a running buffer consisting of 20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl at a constant flow rate of 0.5 ml min^{-1}. A standard curve was generated by using molecular-weight markers (Sigma). The Stokes radii of β-amylase, alcohol dehydrogenase, carbonic anhydrase and cytochrome *c* were calculated from the crystal structures of each protein (PDB entries 1fa2, 2hcy, 1v9e and 1hrc, respectively) using *HYDROPRO* (García de la Torre *et al.*, 2000).

2.6. PDB code

The atomic coordinates and structure factors were deposited in the PDB with accession code 3wr5.

2.7. Computer simulation

The hybrid quantum-mechanics/molecular-mechanics (QM/MM) method was used to study the conversion of CO_2 to formate catalyzed by the FDHs TsFDH and CbFDH.

2.7.1. Enzyme-model setup. The coordinates of heavy atoms for apo and holo forms of TsFDH were obtained from the X-ray structure of NAD$^+$-bound TsFDH with a resolution of 2.14 Å (PDB entry 3wr5). The initial position of a formate molecule was obtained from that of an azide molecule used as an inhibitor in the X-ray structure of FDH from *Pseudomonas* sp. (PDB entry 2nad; Lamzin *et al.*, 1994). The coordinates for CbFDH were obtained from the X-ray structure of its apo form with a resolution of 1.70 Å (PDB entry 2fss; Schirwitz *et al.*, 2007).

Since 2fss is a mutated form of the original CbFDH, some residues (Glu47, Asn51, Gln56, Ile79, Lys84, Val184, Asp202 and Gln325) in the crystal structure were replaced by the original residues using the information from the original CbFDH sequence (UniProtKB O13437; Magrane *et al.*, 2011). The coordinates of both NAD and formate from the TsFDH system were adopted.

H atoms were added to both enzymes using the *HBUILD* module in *CHARMM* with the protonation state at neutral pH. In addition, the hydrogen positions in the His333 residue of TsFDH and the His311 residue of CbFDH were determined from previous QM/MM simulations (Castillo *et al.*, 2008).

TIP3P water molecules were initially added in the space within 20 Å of the active site (or the centres of masses of the formate molecule), removing those within 2.8 Å of the enzyme–substrate system (Brooks & Karplus, 1983). The enzyme and cofactor in the MM region were modelled using the CHARMM force field for the protein (MacKerell *et al.*, 1998) and NAD(H) (Pavelites *et al.*, 1997). Following the prescription for water solvation from previous studies (Hou & Cui, 2012), each system was equilibrated using classical molecular-dynamics simulations for 20 ps with the entire protein, and substrate atoms were fixed at their original position followed by filling the empty space within 20 Å of the active site with additional water molecules again. This equilibration filling procedure was repeated until the number of water molecules in the spherical region reached a plateau. The resulting numbers of water molecules in the TsFDH and CbFDH systems were 82 and 191, respectively.

2.7.2. QM/MM simulation setup. Each enzyme–substrate system was partitioned into four regions (QM, MM, Langevin and continuum regions) following previous studies (Hou & Cui, 2012). The QM region consisted of key reacting molecules in the active site as shown in Supplementary Fig. S1: the nicotinamide group of NAD$^+$ and the formate molecule. To treat the QM/MM interface, we added a link atom between the nicotinamide and nucleotide moieties using the DIV scheme (König *et al.*, 2005).

The MM region spanned the spherical region at 18 Å from the active site encompassing the QM region. In this region, where all residues and molecules including water molecules (except for the QM part) were explicitly represented, classical molecular-dynamics simulations were employed using the CHARMM36 force field (Best *et al.*, 2012).

The Langevin region, which acts as a buffer region between the MM and continuum regions, spanned the spherical shell region at 18–20 Å from the active site, in which Langevin dynamics took into account the effects of water molecules in the form of friction and random forces. In this region, enzyme residues were harmonically restrained with force constants obtained from the crystallographic *B* factor (Hou & Cui, 2012).

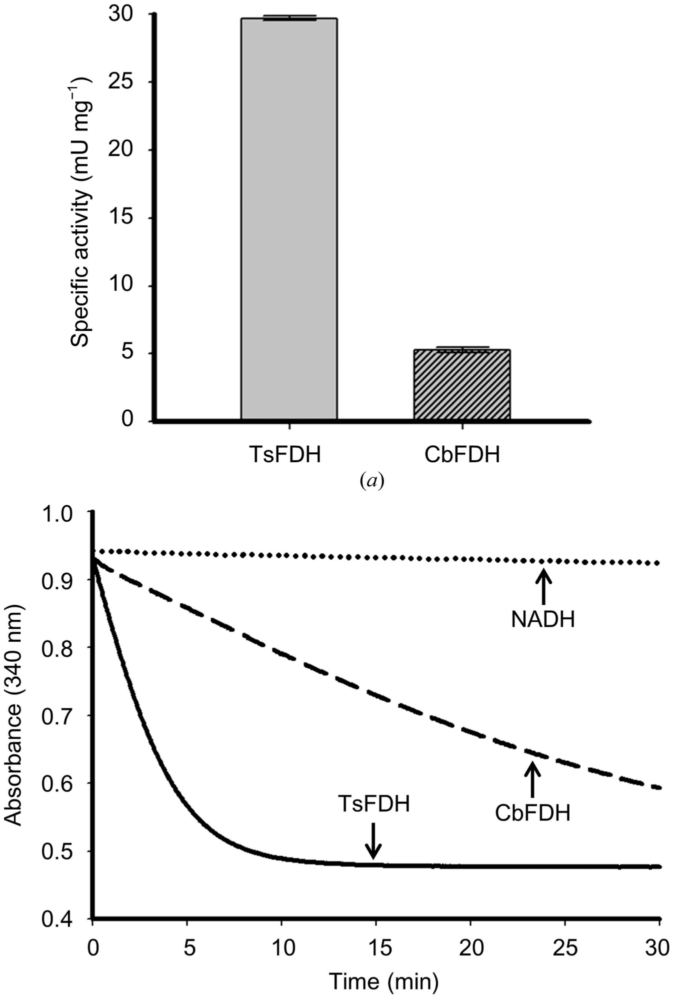
Finally, the continuum region consisted only of enzyme residues located beyond 20 Å from the active site, which were fixed during the simulation. In this region, the water molecules were implicitly represented as a continuum dielectric medium with a dielectric constant of 80. Further simulation details for the QM/MM simulation setup can be found elsewhere (Hou & Cui, 2012).

2.7.3. System equilibration and potential energy calculation. The two systems were minimized with the steepest-decent method and then gradually heated to 300 K for 100 ps with a 2.0 fs time step. After heating, the systems were equilibrated for 100 ps with a 1.0 fs time step. The final structures were used for potential energy calculation along the reaction coordinate defined as $\xi = r(\mathrm{NC4{-}H4}) - r(\mathrm{C1{-}H4})$ (see Supplementary Fig. S1 for atom labels) to study the catalytic reaction $CO_2 + NADH \rightarrow HCO_2^- + NAD^+$. The potential energy curve along the reaction coordinate was initially calculated by the QM/MM method with the self-consistent-charge density-functional tight-binding (SCC-DFTB) method developed by Elstner *et al.* (1998) for the QM calculation with a step size of 0.1 Å in the ξ range from −2.4 to 2.5 Å. At each reaction coordinate, the geometry was optimized using the Newton–Raphson method with a tolerance of 0.01 in the whole system. The optimized geometry at each ξ was used for single-point energy calculation by the *ab initio* B3LYP/6-311++G(2d,2p) method using *GAMESS* (Schmidt *et al.*, 1993) as implemented in the *CHARMM* package (c38b2; Brooks *et al.*, 2009). In addition, further geometry optimization and single-point energy calculation at each point were performed by the *ab initio* B3LYP/631+G(2d,2p) and B3LYP/6-311+G(2d,2p) methods to check the validity of the semi-empirical SCC-DFTB method in this study. The optimized geometries of the active site of both TsFDH and CbFDH in reactant, transition and product states are shown in Supplementary Figs. S2 and S3.

3. Results and discussion

3.1. CO_2-reducing activity of FDHs

The absorption at 340 nm was monitored to measure the CO_2-reducing activities of FDHs. TsFDH exhibited a 5.6-fold higher activity (29.71 mU mg^{-1}) than CbFDH (5.3 mU mg^{-1}) in CO_2-saturated sodium phosphate buffer pH 7.0 (Fig. 1*a*). This result is consistent with the rates of formate production by the two FDHs at pH 7.0 in our previous report: TsFDH showed a 5.8-fold higher rate of formate production than CbFDH (0.023 *versus* 0.004 s^{-1}; Choe *et al.*, 2014). The absorbance changes during the FDH-catalyzed CO_2-reduction



(*a*)

(*b*)

Figure 1
Enzyme activities of FDHs. (*a*) CO_2-reduction activities of TsFDH and CbFDH in CO_2-saturated conditions. (*b*) Absorbance change during the enzymatic CO_2 reduction.

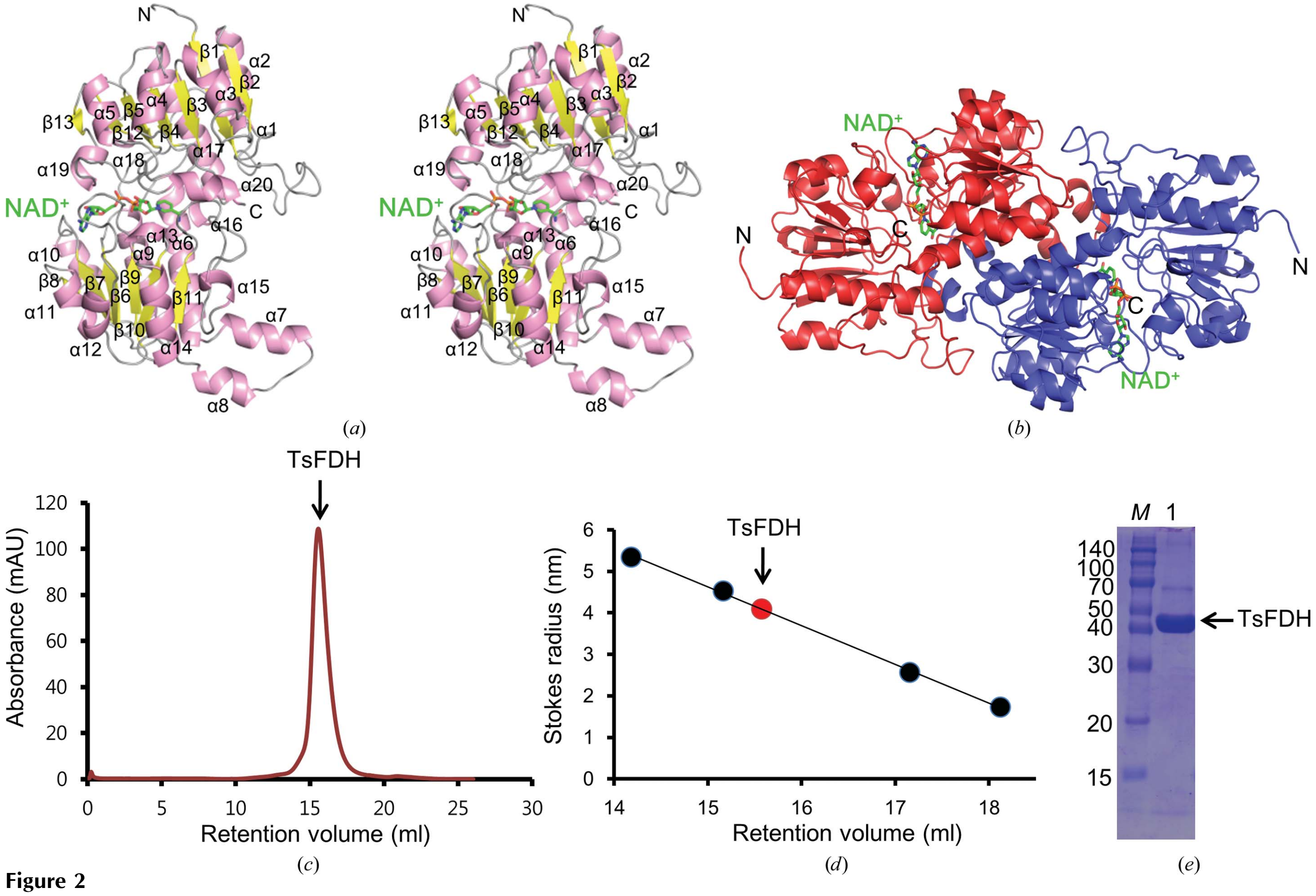
reactions are shown in Fig. 1(*b*). TsFDH rapidly produced formate by oxidizing NADH during 5 min incubation and reached an equilibrium between the reduction and oxidation reactions after 10 min incubation. In contrast, CbFDH slowly produced formate and did not reach equilibrium after 30 min incubation. The rate of NADH decomposition was very low (98 n*M* min^{-1}) in the CO_2-saturated buffer.

3.2. Crystal structure determination of the TsFDH–NAD$^+$ complex

To obtain structural insights into the CO_2-reduction activity of TsFDH, we analyzed its crystal structure. The 2.14 Å resolution crystal structure of TsFDH was solved by molecular replacement using PsFDH (PDB entry 2nad; Lamzin *et al.*, 1994) as a search model. The refined models showed $R_{\rm work}$ and $R_{\rm free}$ values of 17.2 and 20.9%, respectively, for the resolution range 42.3–2.14 Å (Table 1). The asymmetric unit contained four monomers of TsFDH, with three NAD$^+$ molecules (one each in monomers *A*, *B* and *C*) and 972 water molecules ($V_{\rm M}$ = 2.2 Å^3 Da^{-1}). The refined model of the TsFDH–NAD$^+$ complex accounts for residues 1–397 of monomer *A*, residues 1–392 of monomers *B* and *C* and residues 1–374 of monomer *D*. A Ramachandran plot of the refined model showed that 99.2, 0.8 and 0% of the nonglycine residues were in the most favoured, allowed and disallowed regions, respectively (Table 1).

3.3. Overall structure and structural comparisons

The crystal structure of the TsFDH–NAD$^+$ complex consists of 20 α-helices and 13 β-strands (Figs. 2 and 3), and contains two globular domains: an NAD$^+$-binding domain made up of Rossmann folds and a catalytic domain that contains the amino acids essential for catalysis (Fig. 2*a*). A structural similarity search with full-length TsFDH (chain *A*) was performed using the *DALI* server (Holm & Rosenström, 2010). The highest *Z*-score (67.6, with an r.m.s. deviation of 0.33 Å for the 384 C$^{\alpha}$ atoms) was obtained using the PsFDH–NAD$^+$ complex (PDB entry 2nad; Lamzin *et al.*, 1994). When TsFDH–NAD$^+$ (chain *A*) was superimposed onto the NAD$^+$-bound structures of FDH from *Arabidopsis thaliana* (AtFDH;



(*a*) (*b*) (*c*) (*d*) (*e*)

Figure 2
Overall architecture of TsFDH. (*a*) Stereo ribbon diagram of TsFDH. Helices are shown in pink and strands are shown in yellow. All of the figures, including the protein structures, were drawn using *PyMOL* (http://www.pymol.org). (*b*) Dimer structure of TsFDH. Monomers *A* and *B* are coloured red and blue, respectively. NAD$^+$ molecules bound to each monomer are shown as green sticks. (*c*) Analytical gel-filtration profile of TsFDH. (*d*) Standard curve of analytical gel filtration generated with molecular-weight markers. The positions of the molecular-weight markers (β-amylase, alcohol dehydrogenase, carbonic anhydrase and cytochrome *c*) are indicated by black dots. The position of TsFDH is marked with a red dot. (*e*) SDS–PAGE of the TsFDH preparation used for crystallization and analytical gel filtration. The gel was stained using Coomassie Brilliant Blue. Lane *M* contains molecular-weight markers (labelled in kDa).

PDB entry 3n7u) and *Moraxella* sp. (MsFDH; PDB entry 2gsd), the r.m.s. deviations were 1.14 and 0.45 Å for the 344 and 392 C^{α} atoms of monomer *A*, respectively. For CbFDH in the absence of NAD^+, the r.m.s. deviations were 2.22 Å for the 336 C^{α} atoms of monomer *A*. For a more detailed comparison, FDH dimers from other species were superimposed onto a dimer of TsFDH (monomers *A* and *B*; Fig. 2*b*). In the comparison of dimeric TsFDH (monomers *A* and *B*) with other NAD^+-bound FDHs, the r.m.s. deviations were 2.53 Å for the 760 C^{α} atoms of PsFDH (PDB entry 2nad; Lamzin *et al.*, 1994), 2.64 Å for the 416 C^{α} atoms of MsFDH (PDB entry 2gsd; Shabalin *et al.*, 2009) and 3.54 Å for the 688 C^{α} atoms of AtFDH (PDB entry 3n7u). These results suggest that the overall dimeric structure of TsFDH (NAD^+-bound form) is similar to those of other NAD^+-bound FDHs.

To define the quaternary structure of TsFDH in solution, analytical gel filtration was performed with a Superdex 200 (10/300 GL) column. The Stokes radius of TsFDH was estimated to be 3.72 nm, which is highly similar to the calculated Stokes radius of the TsFDH dimer structure (Figs. 2*c*, 2*d* and 2*e*). This result indicates that the TsFDH–NAD^+ complex exists as a homodimer in solution. Consistent with this observation, other structurally characterized FDHs from *Moraxella* sp. and *Pseudomonas* sp. 101 have been shown to exist as homodimers (Lamzin *et al.*, 1994; Shabalin *et al.*, 2009).

3.4. Open and closed conformations of TsFDH

NAD^+-dependent FDHs exist in either open (NAD^+-unbound) or closed (NAD^+-bound) conformations (Shabalin *et al.*, 2009). When we looked at the active sites of each monomer in the asymmetric unit of TsFDH, electron densities representing NAD^+ molecules were observed in monomers *A*, *B* and *C* (closed conformation), but were not seen in monomer *D* (open conformation) (Figs. 4*a* and 4*b*). To determine the differences between the NAD^+-bound and NAD^+-unbound forms of TsFDH, we calculated the C^{α}-atom deviation plots between monomer *A* and the other monomers in TsFDH. When monomer *A* was superimposed onto the other monomers in the asymmetric unit (monomers *B*, *C* and *D*), the r.m.s. deviations were 0.24, 0.25 and 1.44 Å for the 392, 392 and 368 C^{α} atoms of subunits *B*, *C* and *D*, respectively (Fig. 4*a*). One major structural difference in monomer *D* was that helices $\alpha18$ and $\alpha19$ were disordered (Fig. 3), whereas those in monomers *A*, *B* and *C* were ordered (Figs. 3 and 4*a*). The C^{α}-deviation plots also showed significant differences for all of the residues,

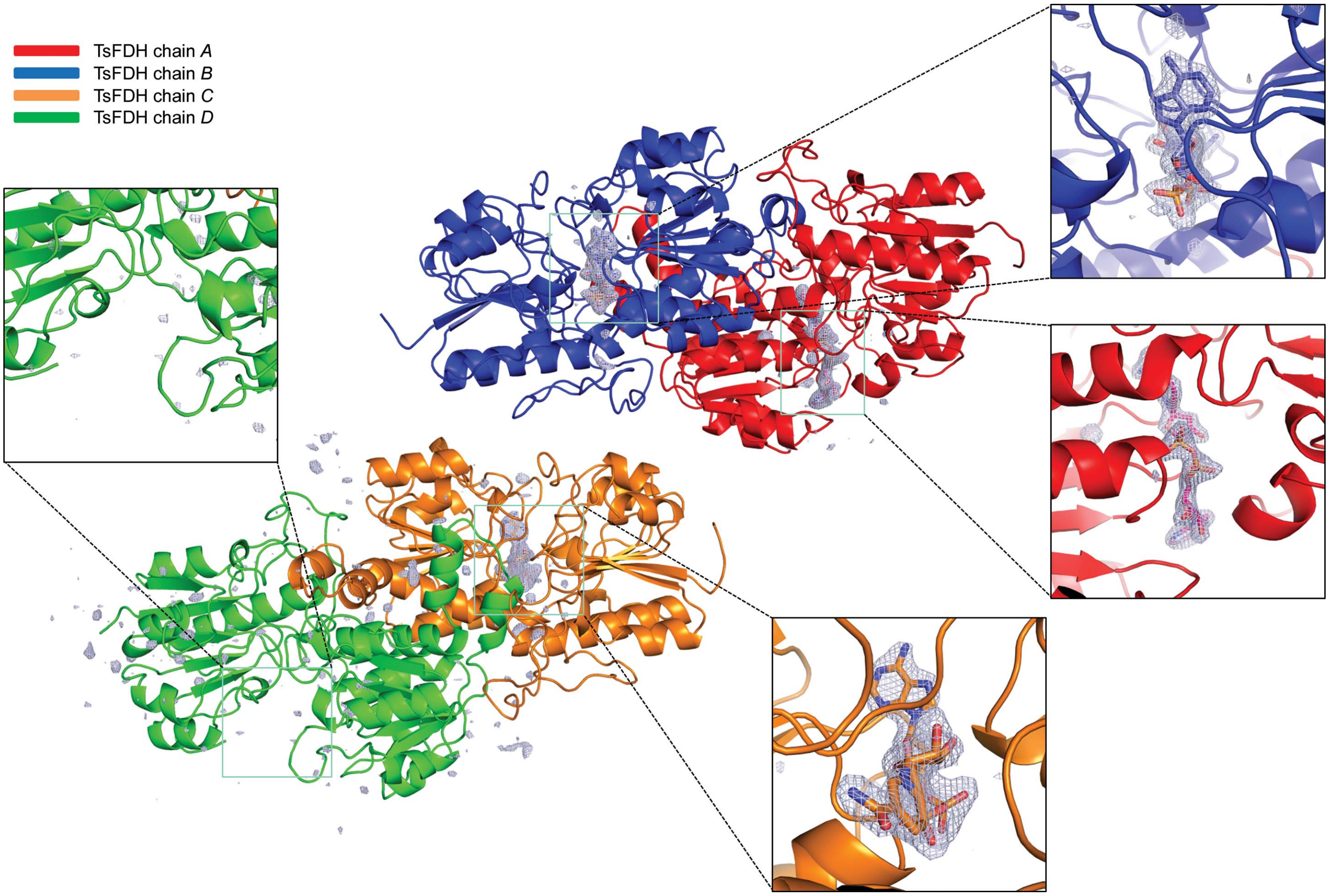
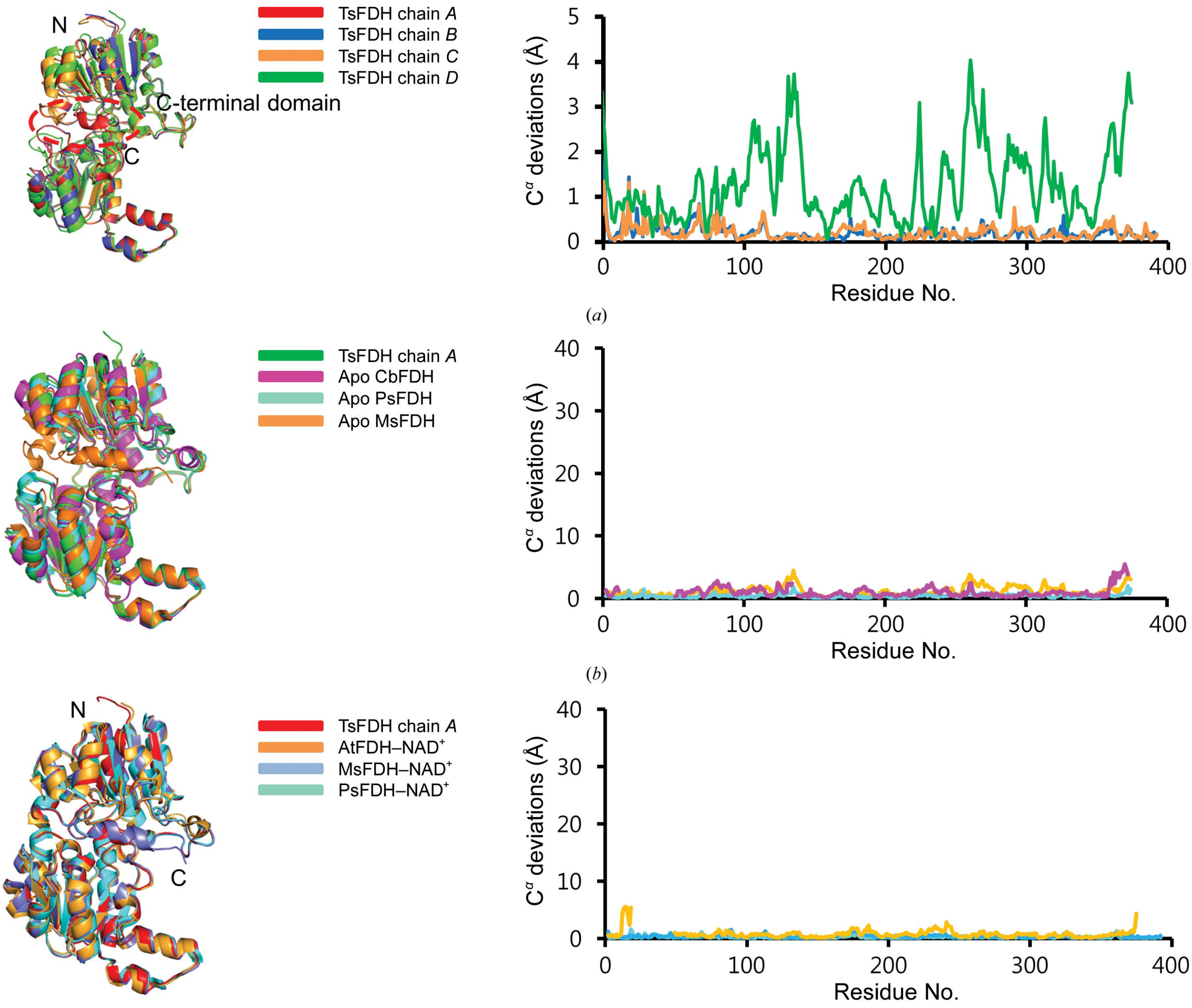


Figure 3
$F_o - F_c$ OMIT maps (3.0σ) of NAD^+ molecules in the asymmetric unit of the crystal. TsFDH chains *A*, *B*, *C* and *D* are coloured red, blue, orange and green, respectively. Chain *D* was moved by a symmetry operation to show two dimers of TsFDH. OMIT maps of chains *A*, *B* and *C* were drawn around NAD^+ molecules, while the entire molecule of chain *D* is shown.

indicating that the structure of monomer *D* was different from those of monomers *A*, *B* and *C* (Figs. 3 and 4*a*). To determine whether monomer *D* (TsFDH-*D*) in the TsFDH–NAD$^+$ complex was an apo form, TsFDH-*D* was superimposed onto other reported apo forms of FDHs from *Pseudomonas* sp. 101, *C. boidinii* and *Moraxella* sp. (PDB entries 2nac, 2j6i and 3fn4, respectively; Fig. 4*b*; Lamzin *et al.*, 1994; Schirwitz *et al.*, 2007; Shabalin *et al.*, 2009). The r.m.s. deviations were 0.48 Å for the 368 C$^\alpha$ atoms of *Pseudomonas* sp. 101, 1.49 Å for the 336 C$^\alpha$ atoms of *C. boidinii* and 1.34 Å for the 368 C$^\alpha$ atoms of *Moraxella* sp. The C$^\alpha$-deviation plots showed no significant differences for all of the residues (Fig. 4*b*), implying that the overall structure of TsFDH-*D* is similar to those of other apo FDH structures. The structure of the NAD-bound monomers in the TsFDH–NAD$^+$ complex (chains *A*, *B* and *C*) differs from the structure of apo TsFDH (chain *D*) at the C-terminus, as the C-terminal helix (helix α20) is disordered in the apo TsFDH structure. This observation suggests that the binding of NAD$^+$ to TsFDH increases the stability of the C-terminal cleft (Fig. 4*a*). On comparing monomer *A* of TsFDH with other NAD$^+$-bound FDHs, the r.m.s. deviations were observed to be 0.33 Å for the 384 C$^\alpha$ atoms of PsFDH (PDB entry 2nad; Lamzin *et al.*, 1994), 0.45 Å for the 392 atoms of MsFDH (PDB entry 2gsd; Shabalin *et al.*, 2009) and 1.14 Å for the 344



(*a*)

(*b*)

(*c*)

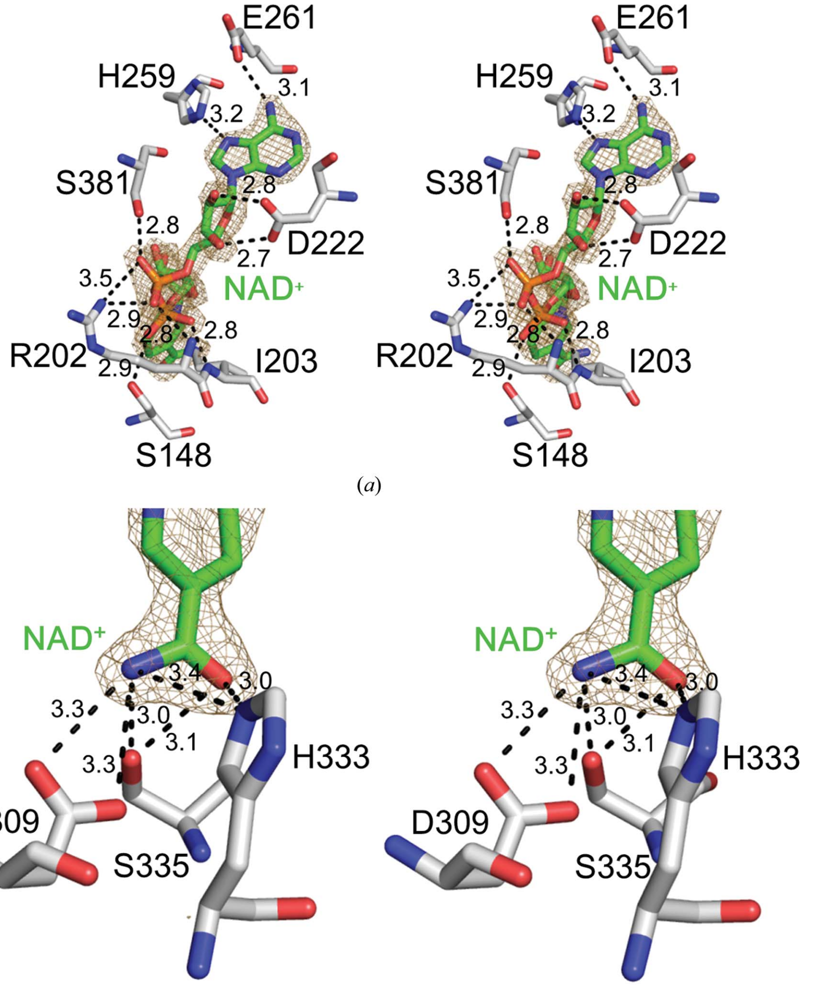
Figure 4

Structural comparisons of FDHs. (*a*) The left panel shows superposition of TsFDH chain *A* (PDB entry 3wr5, red) onto TsFDH chains *B*, *C* and *D* (blue, orange and green, respectively). The right panel shows a C$^\alpha$-deviation plot calculated by superimposing TsFDH chains *B*, *C* and *D* onto chain *A*. (*b*) The left panel shows superposition of apo FDHs. TsFDH (chain *D*, green) was superimposed onto CbFDH (PDB entry 2j6i, magenta), PsFDH (PDB entry 2nac, pink) and MsFDH (PDB entry 3jtm, orange). The right panel shows a C$^\alpha$-deviation plot calculated by superimposing TsFDH (apo form) onto CbFDH, PsFDH and MsFDH. (*c*) The left panel shows superposition of the TsFDH–NAD$^+$ complex (chain *A*, red) onto the NAD$^+$-bound forms of AtFDH (PDB entry 3n7u, orange), MsFDH (PDB entry 2gsd, light blue) and PsFDH (PDB entry 2nac, pink). The right panel shows a C$^\alpha$-deviation plot calculated by superimposing TsFDH (NAD$^+$-bound form; *A* chain) onto CbFDH, PsFDH and AtFDH.

C^{α} atoms of AtFDH (PDB entry 3n7u). These results suggest that the overall structure of NAD$^+$-bound TsFDH monomer *A* is similar to those of other NAD$^+$-bound FDHs (Fig. 4*c*). Rotation of apo TsFDH between the two domains (NAD$^+$-binding and catalytic domains) facilitates closure of the interdomain cleft.

Table 2
The size of the active-site openings (holes 1 and 2).

Source	Ligand	PDB code	Hole 1 (Å)	Hole 2 (Å)	Volume (Å^3)
Thiobacillus sp.	NAD$^+$ (chain *A*)	3wr5	7.4, 4.2	—	709.0
Thiobacillus sp.	Apo (chain *D*)	3wr5	8.0, 8.9	7.7, 6.3	2473.3
Candida boidinii	Apo	2j6i	7.4, 6.5	7.2, 5.2	832.5
Arabidopsis thaliana	Apo	3jtm	9.6, 10.5	10.7, 8.9	2741.1
Arabidopsis thaliana	NAD$^+$	3n7u	7.2, 6.1	—	892.1
Pseudomonas sp. 101	Apo	2nac	6.7, 5.9	—	2468.6
Pseudomonas sp. 101	Formate	2gug	4.4, 9.9	5.9, 3.7	1196.3
Pseudomonas sp. 101	NAD$^+$	2nad	6.5, 6.0	—	813.6
Pyrobaculum aerophilum	NADP$^+$	1qp8	8.2, 9.5	—	2933.8
Moraxella sp. C-1	Apo	3fn4	4.6, 5.5	—	854.2
Moraxella sp. C-1	NAD$^+$	2gsd	6.4, 7.4	—	868.4



(*a*)

(*b*)

Figure 5
NAD$^+$-binding mode and formate production *via* FDH-catalyzed CO_2 reduction. (*a*) Stereoview of the interactions between TsFDH and the nicotinamide ring of NAD$^+$. Seven TsFDH residues (Ser148, Arg202, Ile203, Asp222, His259, Glu261 and Ser381) and an NAD$^+$ molecule are shown as white and green sticks, respectively. An $F_o - F_c$ electron-density map is drawn around the NAD$^+$ molecule (4.0σ). Hydrogen bonds are indicated by dotted lines, with distances shown in Å. (*b*) Stereoview of the interactions between TsFDH and the adenine ring of NAD$^+$. An $F_o - F_c$ electron-density map is drawn around the NAD$^+$ molecule (4.0σ) and hydrogen bonds are indicated by dotted lines, with distances shown in Å.

3.5. NAD$^+$ coordination

In the TsFDH–NAD$^+$ complex, NAD$^+$ is bound inside a cavity located in the cofactor-binding domain (Fig. 5*a*). The nicotinamide-ring N7N atom of NAD$^+$ forms hydrogen bonds to the NE2 atom of His333 (3.4 Å), the OG atom of Ser335 (3.0 Å) and the OD1 and OD2 atoms of Asp309 (both 3.3 Å) (Fig. 5*b*). The O7N atom of NAD$^+$ forms hydrogen bonds to the NE2 atom of His333 (3.0 Å) and the OG atom of Ser335 (3.1 Å). The ribose hydroxyl group O2B and O3B atoms form hydrogen bonds to the OD1 (2.8 Å) and OD2 (2.7 Å) atoms of Asp222, respectively. The N6A atom of NAD$^+$ forms a hydrogen bond to the OE2 atom of Glu261 (3.1 Å), and the N7A atom of NAD$^+$ forms a hydrogen bond to the NE2 atom of His259 (3.2 Å). The O1A atom of NAD$^+$ interacts with the OG atom of Ser381 (2.8 Å), the O1A and O2A atoms interact with the NH1 atom of Arg202 (3.5 and 2.9 Å, respectively), and the O2A atom of NAD$^+$ forms a hydrogen bond to the backbone N atom of Arg202 (2.9 Å). The O1N atom interacts with the backbone N atom of Ile203 (2.8 Å), and the O2N atom interacts with the OG atom of Ser148 (2.9 Å).

3.6. Structural insights into the CO_2-reduction activity of TsFDH

We next investigated the structural reason for the higher CO_2-reduction activity of TsFDH compared with that of CbFDH. The open conformation of NAD$^+$-dependent FDHs permits the binding of NAD$^+$ and formate, whereas the closed conformation makes the enzyme more rigid, so that NAD$^+$ and formate are in the proper orientation to facilitate the reaction (Nilov *et al.*, 2012). Because NAD$^+$ binding to FDHs in the absence of formate changes the conformation of FDH from open to closed, we expected that holes would be formed around the active site for the recruitment of formate. When the surfaces of the structurally character-

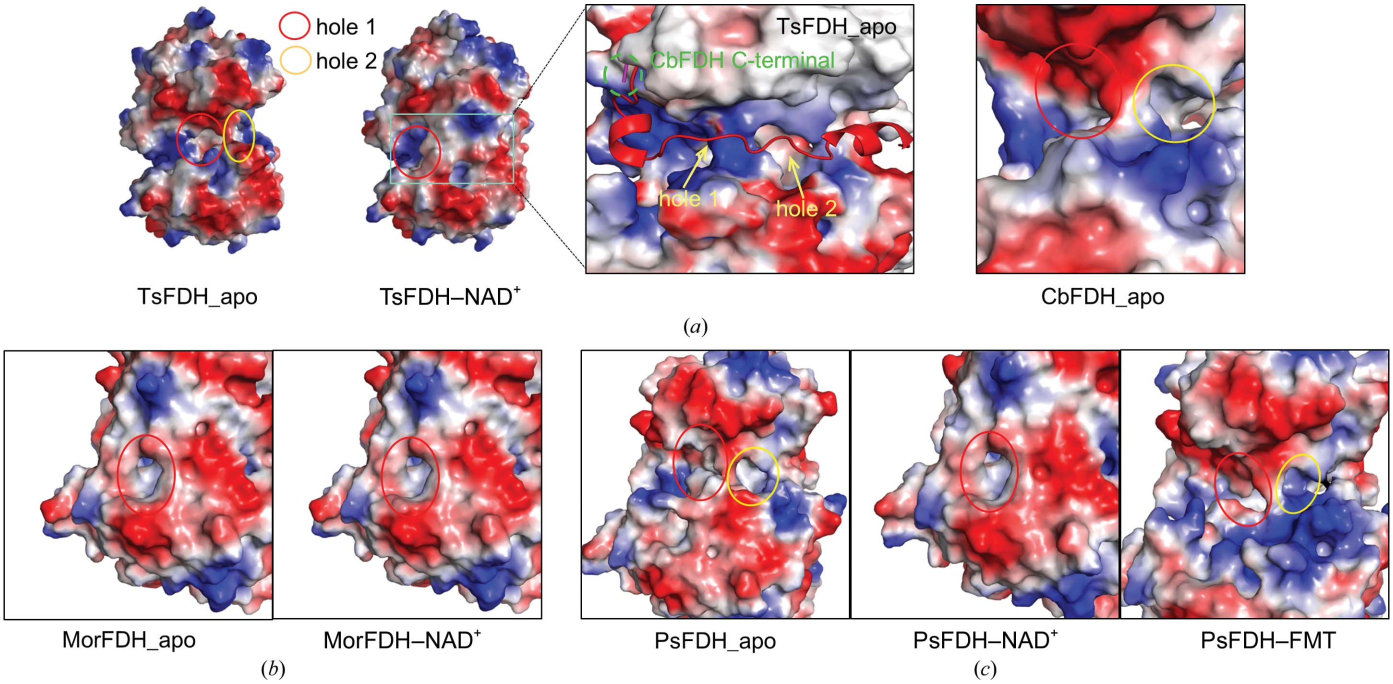


Figure 6
Electrostatic potentials and comparisons of hole 1 and hole 2 in various FDHs. (*a*) Electrostatic potentials (blue, positive; red, negative) at the molecular surfaces of TsFDH (apo and NAD^+-bound forms) and CbFDH (apo form). Red and yellow circles denote hole 1 and hole 2, respectively. Apo and NAD^+-bound structures of TsFDH were superimposed by overlapping the NAD^+-bound structure (red ribbons) on the surface of the TsFDH apo form. Hole 1 and hole 2 are indicated by yellow arrows. The CbFDH apo structure was also incorporated into the apo structure of TsFDH and is indicated by the pink ribbon and green dotted circle. (*b*) Electrostatic potentials at the molecular surfaces of MsFDH (apo and NAD^+-bound forms). (*c*) Electrostatic potentials at the molecular surfaces of PsFDH (apo, NAD^+-bound and formate-bound forms)

ized FDHs were examined, we found two holes, one near the active site of TsFDH and one near the C-terminus, and we designated these as ‘hole 1’ and ‘hole 2’, respectively (Fig. 6). Consistent with this observation, a molecular-dynamics simulation indicated that the formate molecule accesses the active site through the NAD^+-binding site, which is wide in apo FDH, or through the substrate channel, which corresponds to hole 2 in TsFDH (Nilov *et al.*, 2012). When we compared the sizes of hole 1 and hole 2 among other structurally characterized FDHs, hole 1 was present both with or without the NAD^+ molecule, whereas hole 2 was blocked by the closure of the interdomain cleft upon NAD^+ binding (Table 2). In the TsFDH–NAD^+ complex, hole 2 is blocked by helix α20 (Fig. 6*a*), and this blockage is thought to be essential for the exclusion of solvent molecules during catalysis.

When we checked TsFDH (apo form) and CbFDH (apo form; PDB entry 2j6i) for structural differences around hole 1 and hole 2, three structural discrepancies were observed. Firstly, hole 2 of TsFDH is blocked by helix α20. In contrast, CbFDH does not have helix α20, as it is shorter than TsFDH (Figs. 3 and 6*a*). Secondly, the sizes of hole 1 (an ellipse with axes of 8.0 and 8.9 Å) and hole 2 (an ellipse with axes of 7.7 and 6.3 Å) in TsFDH are larger than those of hole 1 (an ellipse with axes of 7.4 and 6.5 Å) and hole 2 (an ellipse with axes of 7.2 and 5.2 Å) in CbFDH (Fig. 6*a* and Table 2), respectively. The substrate channel (hole 2 in TsFDH) is crucial for substrate passage; therefore, the larger sizes of hole 1 and hole 2 might affect the reaction kinetics. Thirdly, Lys287 in TsFDH, which corresponds to Lys286 in PsFDH that is essential for the capture of the formate molecule and its subsequent delivery to the active site of FDH, is an alanine in CbFDH (Nilov *et al.*, 2012; Fig. 3). Arg285 has also been suggested to play a crucial role in substrate capture and delivery (Nilov *et al.*, 2012). Arg285 is conserved in both TsFDH and CbFDH (Fig. 3), suggesting that it may also play a crucial role in substrate

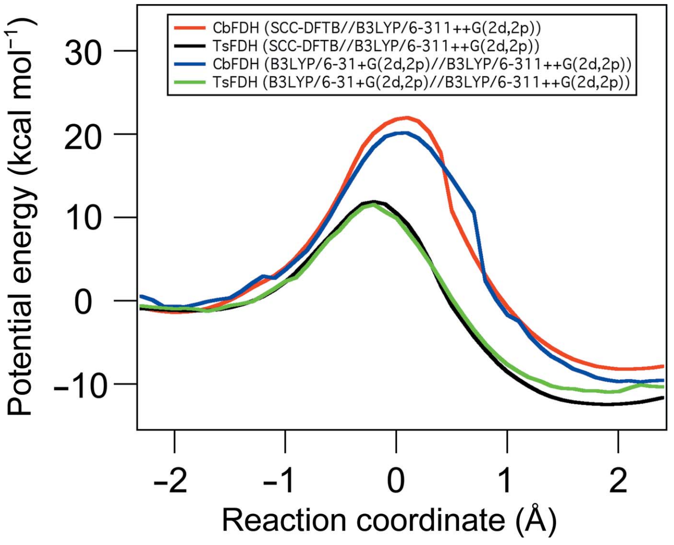


Figure 7
Potential energy curves for the $CO_2 \rightarrow$ formate conversion using different QM methods.

capture and delivery. The average temperature factors for Lys287 and Arg285 in TsFDH-*D* are 48 and 42 Å^2 (20 and 13 Å^2, respectively, in the *A* subunit, 19 and 15 Å^2, respectively, in the *B* subunit, and 28 and 20 Å^2, respectively, in the *C* subunit), respectively, which suggests their conformational flexibility in the open conformation. The volumes of each cavity were also calculated using the *CASTp* server (Dundas *et al.*, 2006; Table 2). Taken together, our results show that TsFDH has a higher rate of CO_2 reduction and a 55.4-fold higher CO_2 reduction to formate oxidation ratio when compared with CbFDH (Table 2). In addition, our results revealed several structural differences between TsFDH and CbFDH, which could be used to engineer FDHs that catalyze CO_2 reduction more efficiently on an industrial scale. Although much remains to be learned about the molecular mechanism underlying the enhanced CO_2-reduction activity of some FDHs, our experimental results and the model presented here provide a step towards such an understanding.

3.7. Computer simulation

Three types of FDH were investigated in the current QM/MM simulation studies: the holo and apo forms of TsFDH and the apo form of CbFDH. However, since the C-terminal region of apo TsFDH was not resolved in the X-ray structure, the residue Arg285, which played a role in stabilizing the formate ion by hydrogen bonding in the active site of holo TsFDH, moved away from the active site, resulting in destabilization of the product state. On the other hand, the corresponding Arg258 in CbFDH forms stable hydrogen bonding with the formate ion just as in holo TsFDH. Hence, we calculated the potential energy of the $CO_2 \rightarrow$formate conversion in the other two enzymes, holo TsFDH and apo CbFDH.

The potential energy curves for the $CO_2 \rightarrow$formate conversion are shown in Fig. 7. There are several interesting points that deserve mention. First of all, although there is some roughness in the *ab initio* method owing to loose convergence criteria, both semi-empirical and *ab initio* QM/MM methods give similar potential energy curves as well as optimized geometries (Supplementary Figs. S2 and S3) in the reactant, transition and product states. This is promising because the computational time for the semi-empirical SCC-DFTB method is faster than the corresponding *ab initio* density-functional theory (DFT) method by a factor of 2–3 orders of magnitude (Hou & Cui, 2012).

Secondly, the potential energy of formate is lower than that of CO_2 by about 10–12 kcal mol^{-1}. This is counterintuitive because these enzymes are FDHs. Hence, we additionally calculated the potential of mean force (PMF) along the reaction coordinate using umbrella sampling and weighted histogram analysis methods (Kumar *et al.*, 1992). The SCC-DFTB method was employed for the QM part in this PMF calculation to save computational time. Further simulation results and technical details of the PMF calculation will be given elsewhere. The resulting Gibbs energy difference [$\Delta G = G(CO_2) - G(\text{formate})$] between formate and CO_2 was indeed about 20 kcal mol^{-1}, which is also consistent with previous

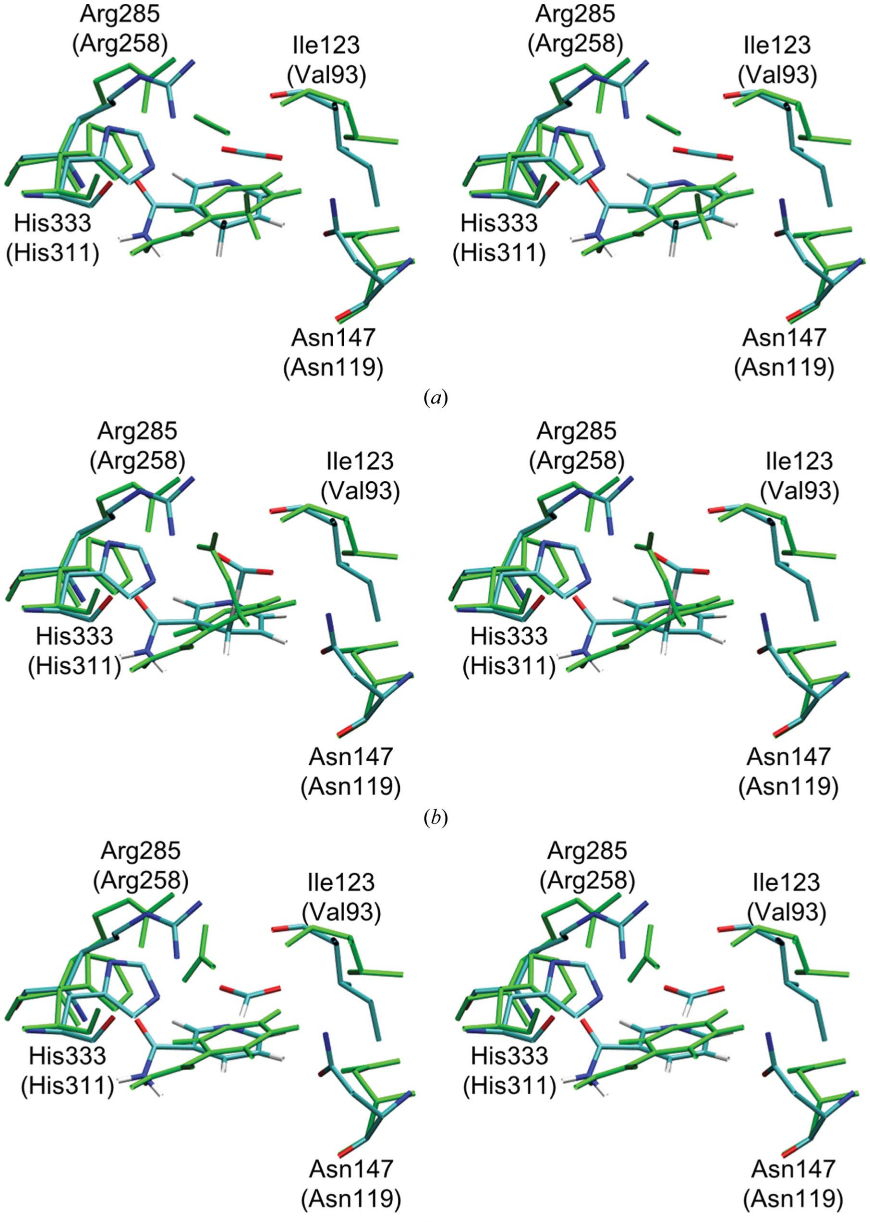


Figure 8
Structural comparison (stereo) of the active sites of TsFDH and CbFDH in the (*a*) reactant, (*b*) transition and (*c*) product states. Coloured stick models correspond to the optimized structures of TsFDH obtained using the B3LYP/6-31+G(2d,2p) method and green stick models correspond to those of CbFDH.

QM/MM studies (Castillo *et al.*, 2008; Vardi-Kilshtain *et al.*, 2012). This implies that the entropic contribution of CO_2 plays a significant role in the catalytic activity of the FDHs.

Thirdly, and most importantly, the activation energy of the $CO_2 \rightarrow$formate conversion was about 10 kcal mol^{-1} lower in TsFDH than in CbFDH, which is consistent with the experimental results. As shown in Fig. 8, the most noticeable difference in the geometry of the active site of the two FDHs was in the location of NADH. NADH in CbFDH (green colour) was shifted from that in TsFDH because the space near Asn119 became open on the rotation of the side chain of Asn119. Upon examining the enzyme structure near Asn119 in CbFDH, we observed that Tyr323 (Tyr345 in TsFDH) plays an essential role in the rotation of the side chain of Asn119. In the holo form of TsFDH, the tyrosine residue pushes the side chain of the asparagine residue, which again restrains the position of NADH in its active site. Hence, the difference in the activation energy shown in this computational modelling study may be owing to structural differences in the apo and holo states of the enzymes. We also admit that our calculation is based on a theoretical model, which should be confirmed by supporting experiments.

4. Related literature

The following references are cited in the Supporting Information for this article: Larkin *et al.* (2007) and Nicholas *et al.* (1997).

The authors thank the staff of beamline 26B1 at SPring-8, Hyogo, Japan and beamline 7A at the Pohang Light Source for their assistance during the X-ray experiments. This work was supported by a grant from the Korea CCS R&D Center (2014M1A8A1049296) and KETEP (20133030000300).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Alissandratos, A., Kim, H.-K. & Easton, C. J. (2014). *Bioresour. Technol.* **164**, 7–11.

Alissandratos, A., Kim, H.-K., Matthews, H., Hennessy, J. E., Philbrook, A. & Easton, C. J. (2013). *Appl. Environ. Microbiol.* **79**, 741–744.

Almendra, M. J., Brondino, C. D., Gavel, O., Pereira, A. S., Tavares, P., Bursakov, S., Duarte, R., Caldeira, J., Moura, J. J. & Moura, I. (1999). *Biochemistry*, **38**, 16366–16372.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M. & MacKerell, A. D. Jr (2012). *J. Chem. Theory Comput.* **8**, 3257–3273.

Bommarius, A. S., Schwarm, M. & Drauz, K. (1998). *J. Mol. Catal. B Enzym.* **5**, 1–11.

Brooks, B. R. *et al.* (2009). *J. Comput. Chem.* **30**, 1545–1614.

Brooks, C. L. III & Karplus, M. (1983). *J. Chem. Phys.* **79**, 6312–6325.

Castillo, R., Oliva, M., Martí, S. & Moliner, V. (2008). *J. Phys. Chem. B*, **112**, 10012–10022.

Choe, H., Joo, J. C., Cho, D. H., Kim, M. H., Lee, M. H., Jung, K. D. & Kim, Y. H. (2014). *PLoS One*, **9**, e103111.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.

Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S. & Seifert, G. (1998). *Phys. Rev. B*, **58**, 7260–7268.

El-Zahab, B., Donnelly, D. & Wang, P. (2008). *Biotechnol. Bioeng.* **99**, 508–514.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

García de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). *Biophys. J.* **78**, 719–730.

Graentzdoerffer, A., Rauh, D., Pich, A. & Andreesen, J. R. (2003). *Arch. Microbiol.* **179**, 116–130.

Hartmann, T. & Leimkühler, S. (2013). *FEBS J.* **280**, 6083–6096.

Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.

Hou, G. & Cui, Q. (2012). *J. Am. Chem. Soc.* **134**, 229–246.

König, P. H., Hoffmann, M., Frauenheim, T. & Cui, Q. (2005). *J. Phys. Chem. B*, **109**, 9082–9095.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. (1992). *J. Comput. Chem.* **13**, 1011–1021.

Lamzin, V. S., Aleshin, A. E., Strokopytov, B. V., Yukhnevich, M. G., Popov, V. O., Harutyunyan, E. H. & Wilson, K. S. (1992). *Eur. J. Biochem.* **206**, 441–452.

Lamzin, V. S., Dauter, Z., Popov, V. O., Harutyunyan, E. H. & Wilson, K. S. (1994). *J. Mol. Biol.* **236**, 759–785.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.

MacKerell, A. D. Jr *et al.* (1998). *J. Phys. Chem. B*, **102**, 3586–3616.

Magrane, M. & The UniProt Consortium (2011). *Database*, **2011**, bar009.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nicholas, K. B., Nicholas, H. B. Jr & Deerfield, D. W. II (1997). *EMBNEW.NEWS*, **4**, 14.

Nilov, D. K., Shabalin, I. G., Popov, V. O. & Švedas, V. K. (2012). *J. Biomol. Struct. Dyn.* **30**, 170–179.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pavelites, J. J., Gao, J., Bash, P. A. & MacKerell, A. D. Jr (1997). *J. Comput. Chem.* **18**, 221–239.

Reda, T., Plugge, C. M., Abram, N. J. & Hirst, J. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 10654–10658.

Schirwitz, K., Schmidt, A. & Lamzin, V. S. (2007). *Protein Sci.* **16**, 1146–1156.

Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M. & Montgomery, J. A. (1993). *J. Comput. Chem.* **14**, 1347–1363.

Schuchmann, K. & Müller, V. (2013). *Science*, **342**, 1382–1385.

Shabalin, I. G., Filippova, E. V., Polyakov, K. M., Sadykhov, E. G., Safonova, T. N., Tikhonova, T. V., Tishkov, V. I. & Popov, V. O. (2009). *Acta Cryst.* D**65**, 1315–1325.

Sheffield, P., Garrard, S. & Derewenda, Z. (1999). *Protein Expr. Purif.* **15**, 34–39.

Vardi-Kilshtain, A., Major, D. T., Kohen, A., Engel, H. & Doron, D. (2012). *J. Chem. Theory Comput.* **8**, 4786–4796.

Yadav, R. K., Baeg, J.-O., Oh, G. H., Park, N.-J., Kong, K.-J., Kim, J., Hwang, D. W. & Biswas, S. K. (2012). *J. Am. Chem. Soc.* **134**, 11455–11461.