FAST COMMUNICATIONS

Contributions intended for this section should be submitted to any of the Co-Editors of Acta Crystallographica *or* Journal of Applied Crystallography. *In the letter accompanying the submission authors should state why rapid publication is essential. The paper should not exceed two printed pages (about 2000 words or eight pages of double-spaced typescript including tables and figures) and figures should be clearly lettered. If the paper is available on 5.25" IBM PC compatible or 3.5" Apple /Macintosh diskettes it would be helpful if these could be sent with the manuscript together with details of the word-processing package used. Papers not judged suitable for this section will be considered for publication in the appropriate section of* Acta Crystallographica *or in* Journal of Applied Crystallography.

Acta Cryst. (1991). B47, 817–820

The structure of 6-phosphogluconate dehydrogenase refined at 2.5 Å resolution

BY MARGARET J. ADAMS,* SHEILA GOVER, RICHARD LEABACK, CHRISTOPHER PHILLIPS AND DONALD O'N. SOMERS†

Laboratory of Molecular Biophysics, Oxford University, Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

(*Received* 19 *August* 1991; *accepted* 23 *August* 1991)

Abstract. The three-dimensional structure of ovine 6-phosphogluconate dehydrogenase, refined at 2.5 Å resolution with a residual for all data of 18.5%, is reported. This model, based on improved diffraction data and a corrected sequence, supersedes that reported earlier. Each subunit of the dimer has three domains: a β-α-β domain binds NADP; an all α domain provides much of the dimer interface; the C-terminal tail burrows into the second subunit.

Introduction. The enzyme 6-phosphogluconate dehydrogenase (6-PGDH) (EC1.1.1.44) is dimeric and NADP dependent for almost all species (Rosemeyer, 1987). It catalyses an oxidative decarboxylation yielding ribulose 5-phosphate and is not metal-ion dependent. The subunit molecular weight is 52 000 daltons. This paper reports the molecular structure, refined at 2.5 Å resolution, of crystals of the sheep liver enzyme grown in the presence of 5′-ADP.

The low-resolution structure of the enzyme and the binding of coenzyme have been reported previously (Adams, Helliwell & Bugg, 1977; Abdallah, Adams, Archibald, Biellmann, Helliwell & Jenkins, 1979). A model based on data to 2.6 Å resolution (Adams *et al.*, 1983) proved difficult to refine with satisfactory geometry. Comparison of the amino-acid sequence (Carne & Walker, 1983) which had been used in defining the chain trace with that predicted for the *E. coli* enzyme from the *gnd* gene sequence (Nasoff, Baker & Wolf, 1984) suggests three cyanogen bromide fragments, corresponding to 97 amino acids, were wrongly positioned. The correct sequence for the sheep liver enzyme has now been established from its cDNA sequence (Somers, Medd, Adams & Walker, 1991) and has been used in the present structure determination.

* To whom correspondence should be addressed.

† Present address: The Wellcome Foundation Ltd., The Wellcome Research Laboratories, Langley Court, South Eden Park Rd., Beckenham, Kent BR3 3BS, England.

0108-7681/91/050817-04$03.00

Experimental. The enzyme was extracted by a new method (D. O'N. Somers, J. Hajdu & M. J. Adams, manuscript in preparation) using cellulose phosphate at an early stage. Crystals were grown from hanging drops containing 5–10 mg ml^{-1} enzyme, 30 m*M* 5′-ADP and 40% saturated ammonium sulfate at pH 6.5; wells contained 52–54% saturated ammonium sulfate at pH 6.5. The crystal form is that reported previously; the space group is $C222_1$ with cell dimensions a=72.74(7), b=148.40(11), c=102.35(8) Å.

Data for native crystals and two derivatives were collected on a Nicolet area detector with graphite-monochromated Cu $K\alpha$ radiation (Table 1). Anomalous differences were measured for the gold derivative; the overall ratio $\Delta_{anom}/\sigma_{anom}$ was 1.25 (2.18 to 4 Å). Initial data processing used the *XENGEN* package (Howard, 1988); subsequent scaling and merging of data used this or in-house programs.

In addition to the major heavy-atom sites found earlier, two minor sites were identified for the platinum derivative. Heavy-atom refinement used centric terms only. Occupancies and temperature factors behaved well for the gold derivative throughout but there was evidence for non-isomorphism in the platinum derivative. Sites 2 and 4 refined to significantly different occupancies in the two platinum crystals and sites 4 and 5 to positions differing by 3.5 Å. The coordinates given in Table 2 were used in phase determination. The figure of merit to 2.8 Å was 0.50.

The Wang (1985) solvent-flattening procedure was modified to allow direct input of a precalculated multiple isomorphous replacement/anomalous dispersion (MIR/AD) phase set. Density modification was cycled in five stages of increasing resolution between 6.0 and 2.8 Å with revision of the solvent envelope at each stage. The final figure of merit calculated by the program was 0.84, the residual was 0.155, the correlation coefficient was 0.984 and the mean accumulated phase change from the MIR/AD phase set was 36°.

The solvent-flattened map was interpreted using the *FRODO* graphics package (Jones, 1985). An experimental

Table 1. *Data-collection parameters*

N_{obs} = total number of observations. N_{ind} = number of independent observations. $R(I) = \sum_h \sum_{i=1,N} |I_{hi} - \langle I_h \rangle| / \sum_h N \langle I_h \rangle$, where h is the number of reflections measured more than once and N is the number of measurements of each reflection. mfid = mean fractional isomorphous difference for terms with $F > 3\sigma(F)$. % complete = % of terms included in mfid. Au derivative prepared by 18–24 h soak in 200 μM $KAu(CN)_2$. Pt derivative prepared by 24 h soak in 2.7 mM $K_2Pt(CN)_4$.

Crystal	Max. resolution (Å)	N_{obs}	N_{ind}	$R(I)$ (%)	Data set	Max. resolution (Å)	N_{ind}	% complete	$\langle I/\sigma I \rangle$	% $F > 5\sigma(F)$	mfid (%)
Native *A*											
1+2	2.79	30479	11962	3.87							
3	2.80	10485	7952	3.98	*A*	2.80	12537	89.2	184.4	91.9	
Native *B*											
4_1	2.18	49066	23329	4.93							
4_2	2.18	28265	19237	4.37	*B*	2.50	17510	89.5	116.0	84.2	
						(2.80	13026	92.9	145.4	89.9)	
					A+B	2.50	17543	89.7	180.0	88.4	
						(2.80	13063	93.2	231.0	95.6)	
Au											
1	2.78	15008	9835	5.28							
2	2.78	16362	8810	4.55							
3	2.77	12335	8793	4.90	Au	2.81	12279	88.4	119.2	90.0	0.146 (97.4)
Pt											
1	2.80	12357	8498	6.13	Pt1	2.82	7931	57.9	61.6	72.2	0.176 (82.7)
2	2.82	8084	6921	6.41	Pt2	2.82	5586	40.8	38.2	67.3	0.171 (83.5)

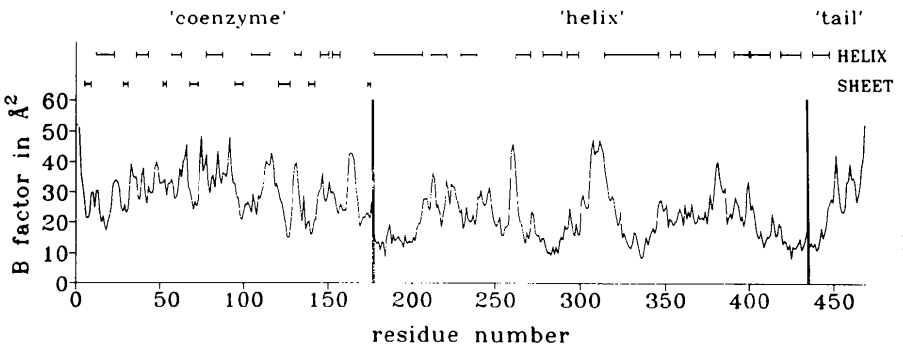

Fig. 1. Mean temperature factors for main-chain atoms of each residue. The overall means in the three domains are: coenzyme 29.9, helix 21.1 and tail 25.9 Å².

sequence for 415 residues (2930 atoms) with side chains never larger than the density and with 11 breaks where connectivity was uncertain was built into the map. Phases from the reliably built parts of the model (78% of the scattering matter) were combined with the MIR/AD phase set between 8 and 2.8 Å using probability coefficients (Hendrickson & Lattman, 1970). Terms in the combined electron density map were weighted individually (Stuart & Artymiuk, 1985). Two cycles of model building and recalculation of combined maps yielded 456 residues with three main-chain breaks.

A matrix was made of size and internal or external environment against residue number for 417 residues with guessed side chains which fitted the density well. A further parameter 'likeness' associated each residue with a series of plausible alternatives which would fit the density. The difference between this and a matrix prepared from the known sequence was minimized, beginning with regions in helices. One domain was aligned with 160 residues beginning at the N-terminus and a further 139 residues in α-helices were assigned. Residues 1–468 were built using four combined electron density maps, in the first of which model residues were replaced by alanine or glycine. Residues beyond 469 are poorly ordered in the crystal. The C-terminus of the original amino-acid sequence corresponded to tryptophan 468.

The model was refined by simulated annealing, using *XPLOR* (Brünger, Karplus & Petsko, 1989; Brünger,

Table 2. *Refined heavy-atom coordinates*

Occupancy (electrons)	x (fractional)	y (fractional)	z (fractional)	B (Å²)
Au*				
72.7 (1.5)	0.2323 (6)	0.0797 (2)	0.2264 (3)	27.9 (2.5)
30.6 (1.4)	0.4937 (20)	0.2220 (5)	0.1485 (8)	34.6 (6.4)
Pt1†				
50.3 (1.9)	0.3168 (23)	0.0874 (4)	0.4477 (5)	1.0 ‡
25.3 (2.4)	0.0224 §	0.1190 (8)	0.2339 (14)	50.0 ¶
30.0 (2.4)	0.1061 (41)	0.1922 (10)	0.5028 (16)	50.0 ¶
27.3 (2.6)	0.3101 (64)	0.2326 (11)	0.1414 (15)	50.0 ¶
35.8 (2.7)	0.1981 (47)	0.2847 (8)	0.0872 (12)	50.0 ¶
Pt2†				
48.7 (2.9)	0.3223 (20)	0.0892 (10)	0.4475 (7)	3.4 (8.6)
55.1 (4.5)	0.0224 §	0.1117 (17)	0.2328 (8)	50.0 ¶
28.3 (2.8)	0.1145 (32)	0.2028 (19)	0.4947 (15)	1.0 ‡
45.9 (4.0)	0.2683 (35)	0.2413 (18)	0.1426 (10)	50.0 ¶
38.2 (3.8)	0.2323 (37)	0.2924 (18)	0.0899 (14)	50.0 ¶

* $\langle F_H/E_{iso} \rangle$ to 2.8 Å is 1.83 (1.34 for worst range); F_H is the heavy-atom structure factor and E_{iso} the isomorphous lack of closure error.
† $\langle F_H/E_{iso} \rangle$ to 2.8 Å is 1.27 (1.06 for worst range).
‡ Reset from negative.
§ Coordinate not refined.
¶ Fixed.

1990) Version 2.1. Initially all observed data between 8 and 2.8 Å were used; temperature factors (B) were set at 20 Å^2 and not refined individually. The residual dropped from 40.0% for the initial structure to 24.8% with B=13.5 Å^2 after a single refinement cycle. The model was rebuilt using a weighted (Sim, 1959) $(2F_o - F_c)\alpha_c$ map which indicated the position of a well-bound sulfate ion and of 61 waters; there was no evidence for bound ADP. Parameters for a solvent mask were optimized at ρ=0.370 e Å^{-3} and B=95 Å^2 for 58.4% of the cell volume. Individual temperature factors were included in the positional refinement which was extended to all data between 20 and 2.5 Å. The resulting residual for all measured data is 18.5%; the standard deviation in bond lengths is 0.010 Å and in bond angles 2.63°. The residual corresponds to an average error of 0.25 Å in coordinates (Luzzatti, 1952). The mean temperature factor for main-chain atoms (Fig. 1) is 24.8 Å^2; that for side-chain atoms 28.1 Å^2. Only three residues have φ,ψ angles with unfavourable energy (Brant & Schimmel, 1967).* Extension to higher resolution is in progress; full details of the refinement will be published.

The structure. The 6-PGDH monomer, illustrated in Fig. 2(*a*), is labelled to correspond with the topology diagram (Fig. 2*b*). The monomer has three domains with 287 of its residues forming 19 helices and an eight-stranded sheet.

The N-terminal domain (1–176) has a β-α-β fold. Residues 1–128, comprising the parallel sheet strands βA to βF and helices αa to αe, are a dinucleotide-binding fold (Rossmann, Liljas, Brändèn & Banaszak, 1975). The coenzyme binds at the C-terminus of the sheet strands. A β-α-β unit anti-parallel to the Rossmann fold is formed by residues 129–176, βG, αg and βH (which is bulged and deviates from regular beta φ,ψ angles).

The second domain (177–434) is entirely α-helical with 11 helices. A motif of five helices αh–αl (Fig. 3) is repeated by αn–αr. The helices αh and αn are 30 and 33 residues long respectively; they form the centre of the domain and are anti-parallel. Helices αk, αl, αq and αr form much of the dimer interface. The helical domain is the core of the dimer and its mean temperature factor is lower than that of the 'coenzyme' domain (Fig. 1).

The 'tail' (435–468) has a single helix and a loop in which almost all side-chain contacts are with the other subunit of the dimer (Fig. 4). The sulfate ion and 20 water molecules lie in the interface between the tail and the twofold-related monomer.

Discussion. Two major factors contributed to building the correct chain trace for this molecule. The first was the realignment of peptides in the sequence. Conversely, the three-dimensional structure confirms the correctness of the cDNA sequence. The second factor was the improved quality of data collected on the area detector. Table 1

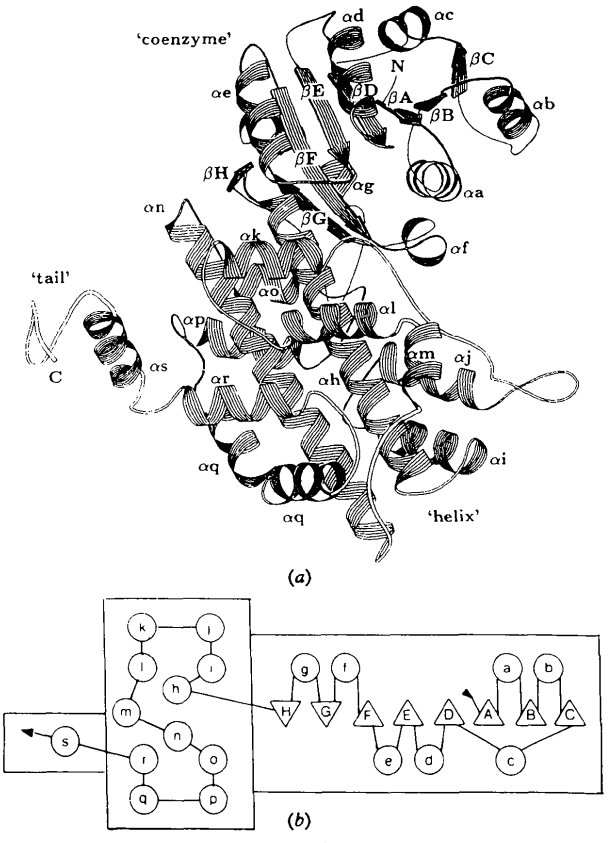

Fig. 2. The 6-PGDH monomer. (*a*) Ribbon diagram (after Richardson, 1985) with domains and secondary structural elements labelled sequentially. (*b*) Topology diagram with domains boxed: △ sheet strand viewed from N, ▽ sheet strand viewed from C, ○ helix.

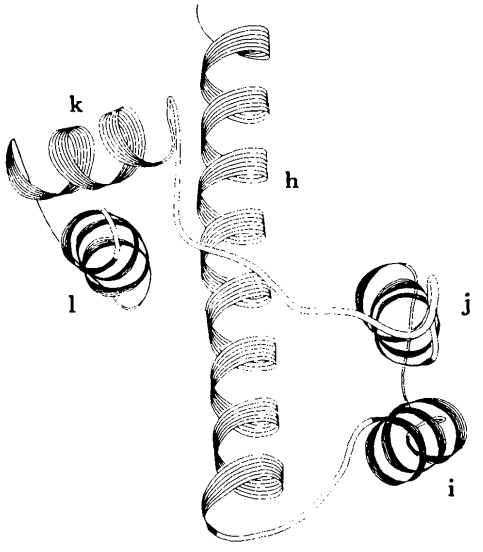

Fig. 3. Helix motif comprising the long helix αh, wrapped by two pairs of almost anti-parallel shorter helices αi, αj and αk, αl.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1PGD, R1PGDSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37054 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 1 January 1993.

shows values of R_{merge} for the gold derivative of ~5%; these contrast with R_{merge} of 14.1% for a similar fraction of the data used in the original structure determination.

The majority of secondary structural elements were recognized in the previous model but not their connectivity. The higher temperature factor of the coenzyme domain increased the risk of wrong connection. Helix αn follows an external extended chain (305–314) with few side-chain contacts and high temperature factors and was not visible above the noise level of early electron density maps. This helix was incorrectly assigned to residues near the N-terminus of the mis-assembled amino-acid sequence.

The 6-PGDH Rossmann fold is more like the lactate dehydrogenase (LDH) NAD domain (Abad-Zapatero, Griffith, Sussman & Rossmann, 1987) than that of any NADP enzyme of known structure. The mean difference in 118 equivalenced α-carbon positions is 2.29 Å; the closest similarity is in the central sheet strands βA, βB, βD and βE. The turn βA–αa has been recognized as the conserved fingerprint 'Gly-*X*-Gly-*X*-*X*-Gly' of the nucleotide-binding fold of NAD enzymes (Wierenga, Terpstra & Hol, 1986). An alternative fingerprint 'Gly-*X*-Gly-*X*-*X*-Ala' has been suggested for some NADP enzymes (Hanukoglu & Gutfinger, 1989; Scrutton, Berry & Perham, 1990). The corresponding residues in 6-PGDH are Gly-*X*-**Ala**-*X*-*X*-Gly (9–14). The central glycine of the fingerprint has been considered obligatory because of the proximity of the adenine ribose and this implies contacts with the coenzyme which are different in detail in 6-PGDH. In LDH aspartate 53 hydrogen bonds to the NAD ribose 2′-OH; it is replaced by arginine 33 in NADP-dependent 6-PGDH. The side chain of this residue is poorly ordered in the absence of NADP. The mobile loop of LDH which moves to enclose the substrate is not present in 6-PGDH.

The structural repeat of the five-helix motif (Fig. 3) is strict. 108 residues (83%) can be superposed with a mean Cα deviation of 1.62 Å. The nine amino-acid identities include an accurately superposed Gly-*X*-Pro turn which packs helices αk and αl (αq and αr) at 155° to each other.

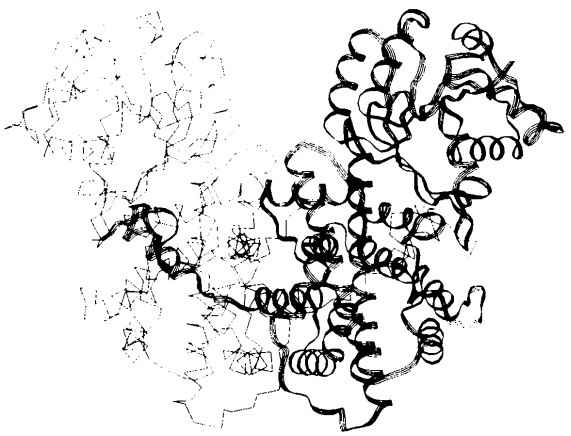

Fig. 4. 6-PGDH dimer demonstrating penetration of the tail of one subunit into the helical domain of the second. One subunit is drawn as a three-stranded ribbon, the other by an α-C trace. The dimer axis is the crystallographic diad parallel to b at $x = 1/2$ and $z = 1/4$.

The two motifs together give a compact bundle of helices. A search using an algorithm based on graph theory (P. Artymiuk, personal communication) indicates the motif is not represented in the Protein Data Bank.

The tail is buried in the second subunit and extends to the boundary between the helical and coenzyme domains. Arginine 446 and histidine 452 of the tail, tyrosine 191 of the second monomer and an ordered water are ligands for the sulfate. Residue 260, an 'essential' lysine (Minchiotti, Ronchi & Rippa, 1981) in the loop between αj and αk, approaches this ion. It is anticipated that the sulfate is bound in a part of the 6-phosphogluconate site. The active site of this dimeric enzyme is thus made up of residues from both larger domains of one subunit and from the tail of the second subunit. Binding studies are in progress.

We are grateful to Professor Sir David Phillips and to Professor Louise Johnson for support and to Dr David Stuart for helpful discussions. We acknowledge support of the EPA Cephalosporin fund and the Medical Research Council for studentships (to DO'NS and CP). MJA is the Dorothy Hodgkin–E. P. Abraham Fellow of Somerville College and a member of OCMS.

References

ABAD-ZAPATERO, C., GRIFFITH, J. P., SUSSMAN, J. L. & ROSSMANN, M. G. (1987). *J. Mol. Biol.* **198**, 445–467.

ABDALLAH, M. A., ADAMS, M. J., ARCHIBALD, I. G., BIELLMANN, J.-F., HELLIWELL, J. R. & JENKINS, S. E. (1979). *Eur. J. Biochem.* **98**, 121–130.

ADAMS, M. J., ARCHIBALD, I. G., BUGG, C. E., CARNE, A., GOVER, S., HELLIWELL, J. R., PICKERSGILL, R. W. & WHITE, S. W. (1983). *EMBO J.* **2**, 1009–1014.

ADAMS, M. J., HELLIWELL, J. R. & BUGG, C. E. (1977). *J. Mol. Biol.* **112**, 183–197.

BRANT, D. A. & SCHIMMEL, P. R. (1967). *Proc. Natl Acad. Sci. USA*, **58**, 428–435.

BRÜNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50–61.

BRÜNGER, A. T. (1990). *XPLOR Manual.* Version 2.1. Yale Univ., New Haven, CT 06511, USA.

CARNE, A. & WALKER, J. E. (1983). *J. Biol. Chem.* **258**, 12895–12906.

HANUKOGLU, I. & GUTFINGER, T. (1989). *Eur. J. Biochem.* **180**, 479–484.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

HOWARD, A. J. (1988). *A Guide to Macromolecular X-ray Data Reduction for the Nicolet Area Detector: The Xengen System.* Version 1.3. Protein Engineering Department, Genex Corporation, Maryland, USA.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

LUZZATTI, P. V. (1952). *Acta Cryst.* **5**, 802–810.

MINCHIOTTI, L., RONCHI, S. & RIPPA, M. (1981). *Biochem. Biophys. Acta*, **657**, 232–242.

NASOFF, M. S., BAKER, H. V. & WOLF, R. E. JR (1984). *Gene*, **27**, 253–264.

RICHARDSON, J. S. (1985). *Methods Enzymol.* **115**, 359–380.

ROSEMEYER, M. A. (1987). *Cell Biochem. Function*, **5**, 79–95.

ROSSMANN, M. G., LILJAS, A., BRÄNDÉN, C.-I. & BANASZAK, L. J. (1975). *The Enzymes*, 3rd ed., Vol. XIA, edited by P. D. BOYER, pp. 61–102. Orlando: Academic Press.

SCRUTTON, N. S., BERRY, A. & PERHAM, R. N. (1990). *Nature (London)*, **343**, 38–43.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

SOMERS, D. O'N., MEDD, S. M., ADAMS, M. J. & WALKER, J. E. (1991). *Biochem. J.* Submitted.

STUART, D. I. & ARTYMIUK, P. (1985). *Acta Cryst.* A**40**, 713–716.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

WIERENGA, R. K., TERPSTRA, P. & HOL, W. G. J. (1986). *J. Mol. Biol.* **187**, 101–107.