CROATICA CHEMICA ACTA CCACAA **73** (3) 887–899 (2000)

ISSN-0011-1643
CCA-2689

Conference Paper

The High Resolution Structure of GDP-4-keto-6-deoxy-D-mannose epimerase/reductase*

Camillo Rosano,[a] *Gaetano Izzo,*[a] *Laura Sturla,*[b] *Angela Bisso,*[b] *Michela Tonetti,*[b] *and Martino Bolognesi* [a,**]

[a] *Department of Physics-INFM and Advanced Biotechnology Center-IST, University of Genova, Largo Rosanna Benzi, 10; I-16132 Genova, Italy*

[b] *Department of Experimental Medicine, Section of Biochemistry, University of Genova, Viale Benedetto XV; 1. I-16132 Genova, Italy*

Received October 28, 1999; revised March 17, 2000; accepted June 28, 2000

GDP-4-keto-6-deoxy-D-mannose epimerase/reductase is a bifunctional enzyme involved in the biosynthesis of cell-surface structures, such as blood group antigens. Each subunit in the homodimeric enzyme consists of two domains. The N-terminal domain displays a Rossmann-fold topology and binds the $NADP^+$ coenzyme. The C-terminal domain is held to bind the substrate. The holo-enzyme structure has been refined at 1.45 Å resolution, based on synchrotron data, to a final *R*-factor of 0.127 (R_{free} = 0.167). The refined protein model highlights several residues involved in coenzyme recognition and binding and suggests that the enzyme belongs to the short-chain dehydrogenase protein homology family. Implications of the catalytic mechanism are discussed.

Key words: fucose metabolism, enzyme structure, short-chain dehydrogenase, $NADP^+$, epimerization, cell surface antigens, high resolution protein crystallography.

---

* Based upon the plenary lecture presented at the 8th Croatian-Slovenian Crystallographic Meeting, Rovinj, Croatia, June 17–19, 1999.

** Author to whom correspondence should be addressed. (E-mail: bolognes@fisica.unige.it)

INTRODUCTION

L-fucose is a fundamental component of glycoconjugates both in prokaryotes and eukaryotes. It can be found in polysaccharides of the bacterial cell wall and in lipopolysaccharides of some Gram-negative bacteria, as well as in animal cells where it is a constituent of ABH blood group antigens and of oligosaccharides belonging to the Lewis system (LeX).[1] These structures are the natural ligands for selectins, the cell-surface adhesion molecules. Metabolism of L-fucose and the subsequent fucosylation reaction constitute the basis of cellular recognition and communication processes. Fucose-related biorecognition of leukocyte LeX by selectins causes leukocyte rolling and extravasation into tissues at the onset of the inflammatory response. Moreover, a genetically inherited defect in L-fucose metabolism is held to be responsible for the Leukocyte Adhesion Deficiency type II (LAD II) syndrome, a disease leading to severe immunodeficiency.[2] Alterations of the fucose content in glycoconjugates have been observed in cancer patients; increased fucosylation of cell membrane glycoconjugates has in fact been related to enhanced cell metastatic potential.

GDP-L-fucose is the substrate of fucosyltransferases, which insert L-fucose into glycoconjugates. The nucleotide-sugar is formed in the cytoplasm of animal cells, mainly *via* a *de novo* pathway that involves GDP-D-mannose, GDP-D-mannose-4,6-dehydratase and the bifunctional enzyme GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GMER), which catalyses both epimerization and a NADPH–dependent reduction, leading to GDP-L-fucose production.

In order to shed more light on the *de novo* pathway of GDP-L-fucose synthesis, and hence to better understand the cell to cell communication mechanisms, a high resolution refinement of recombinant GMER, crystallized in a trigonal crystal form, was carried out at 1.45 Å resolution. Details of the enzyme structure and functional mechanism are presented.

EXPERIMENTAL

Crystallization

Crystals of apo-GMER were grown by the hanging drop vapour diffusion methods from 1.5 M lithium sulfate, 0.1 M Hepes buffer, pH = 7.0, at 21 °C, with a protein concentration of 23 mg/ml. In order to bind the coenzyme, the apo-enzyme crystals were soaked in their mother liquor solution, saturated with $NADP^+$. Both apo- and holo- GMER crystals belong to the trigonal space group $P3_221$ with unit cell constants: $a = b = 102.8$ Å, $c = 74.9$ Å, $\gamma = 120°$, and contain one enzyme molecule (35 kDa) per asymmetric unit.

Data Collection and Structure Determination

X-ray diffraction data were collected using synchrotron radiation (λ = 0.855 Å, at the ELETTRA facility, Trieste, Italy), at 100 K, using the original mother liquor solution modified by addition of 20% glycerol, as cryoprotectant. Diffracted intensities were integrated and scaled to yield 80, 994 unique reflections using the HKL program suite.[3] Table I provides a summary of the data collection statistics. Preliminary analysis of the GMER structure was carried out using difference Fourier methods, and a previoulsy determined GMER structure based on a data set at lower resolution (PDB code 1BWS).[4]

TABLE I

Data collection statistics

Completeness /%	99.1
Unique reflections collected	80994
Redundancy	4.5
Reflection used for R_{free} calculation (5%)	4027
R_{merge} /%	5.7
$I/\sigma(I)$ overall	14.8
$I/\sigma(I)$ outer shell (1.48–1.45 Å)	5.5
Space group	$P3_221$
Unit cell constants/Å, deg	a = 102.8
	b = 102.8
	c = 74.9
	γ = 120

Crystallographic Refinement of the Complex

Crystallographic refinement of the GMER structure was performed using the REFMAC program (CCP4 suite).[5] Rigid body refinement of the available model (determined at 300 K) was preliminarily performed in the 30.0–3.0 Å resolution range, yielding a R-factor value of 38.7%. Subsequently, alternate cycles of positional refinement and inspection of the model structure brought the R-factor (and R_{free}) to 28.8% (and 29.1%) at 2.0 Å resolution. At this point, the $NADP^+$ molecule was modeled in its electron density within the enzyme structure. Additional individual B-factor and atomic coordinates refinement cycles were performed, 379 water molecules were added until convergence was achieved at 1.45 Å resolution. Eventually, individual anisotropic B-factors were refined. Final statistics for the refined structure is reported in Table II.

TABLE II

Crystallographic refinement statistics

Resolution range/Å	10.0–1.45
Total number of non-hydrogen atoms	2945
number of protein atoms	2502
number of water molecules	379
number of $NADP^+$ atoms	48
number of atoms in SO_4^{2-} groups	20
R-factor	0.127
R_{free}	0.167
rmsd from ideal geometry:	
bond length/Å	0.015
bond angle/Å	0.035
planes (1–4)/Å	0.033
Averaged *B*-factors/Å^2:	
main chain	23.4
side chain	28.5
water	52.7
$NADP^+$	41.9
Cruickshank DPI/Å	0.047

RESULTS AND DISCUSSION

Quality of the Refined Structure

The GMER refined model consists of 2502 protein atoms, 379 water molecules, 4 SO_4^{2-} groups and one $NADP^+$ molecule (48 atoms). The electron density is clearly defined for all residues, except for 1 and 317–321, which are disordered. All residues are in the allowed regions of the Ramachandran plot. Residue 118 is a *cis*-proline.

GMER is bound in a dimeric structure in the crystalline state, in agreement with previous solution gel filtration studies. Although the crystallographic asymmetric unit contains only one subunit, a compact dimer is obtained by rotation around a crystallographic two-fold axis in the $P3_221$ space group, yielding a dimeric species that matches those observed in homologous dimeric or tetrameric short chain dehydrogenase enzymes.[6] The subunit interface contact area covers *ca.* 1500 Å^2 in each GMER subunit and it is based on interactions provided by four (two per subunit) extended α-helices, arranged in an α-helical bundle motif.

Each GMER subunit is composed of two domains, which can be defined as mostly N-terminal (*ca*. 190 residues) and mostly C-terminal (*ca*. 100 residues; see Figure 1). The two domains define a main interdomain cleft, which is occupied by the $NADP^+$ coenzyme molecule, present in an extended conformation. Moreover, a wide empty surface cleft is observed in the C-terminal domain region. Inspection of the secondary structure elements and of their topological assembly shows that the N-terminal domain can be properly described as a modified Rossmann fold,[7] consisting of six parallel β-strands and of six α-helices. This fold, which is also known as the nucleotide binding domain, shows the topological switch point between β-strands 1 and 3 (βA and βD) it and effectively hosts the $NADP^+$ molecule, as expected by analogy with other nucleotide binding domains. Helices αD and αE, within the N-terminal domain (the longest helices in the whole subunit structure),

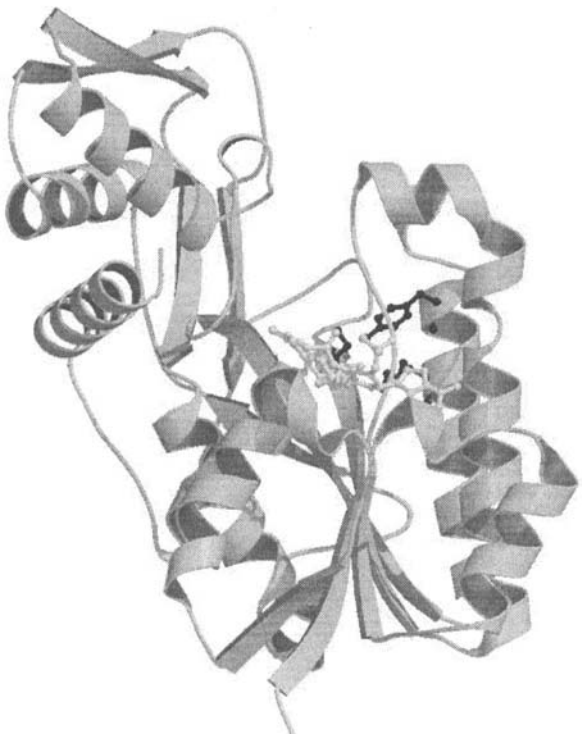

Figure 1. A RIBBON representation of the GMER subunit structure, displaying the N- and C-terminal domains (lower and upper in the figure) and the active site cleft (central). The bound $NADP^+$ coenzyme (in white color) and the active site residues of the proposed catalytic triad, Ser107, Tyr136 and Lys140 (in black) are also displayed in their full structures.

provide the subunit association interface. Remarkably, the interaction between GMER subunits is supported, among others, by a hydrophobic contact involving two pairs of symmetry-related Met (Met84 and Met85) residues at the center of the αD and αE helices, and by two (symmetry-related) salt bridges (residues Glu134 and Lys145), tying together the ends of the interacting helices.

The GMER N-terminal domain fold is clearly related to the molecular fold observed in other short chain dehydrogenases, particularly in UDP-galactose epimerase (UGE),[8] which however displays a Rossmann fold domain based on seven parallel β-strands, containing an additional full α/β unit in the first half of the nucleotide binding domain (the rmsd calculated over 286 Cα atoms common to GMER and UGE is 2.3 Å; see Figure 2). Analysis of the Protein Data Bank archives of three-dimensional structures[9] allows spotting additional, and unexpected, structural relationships. Among them,

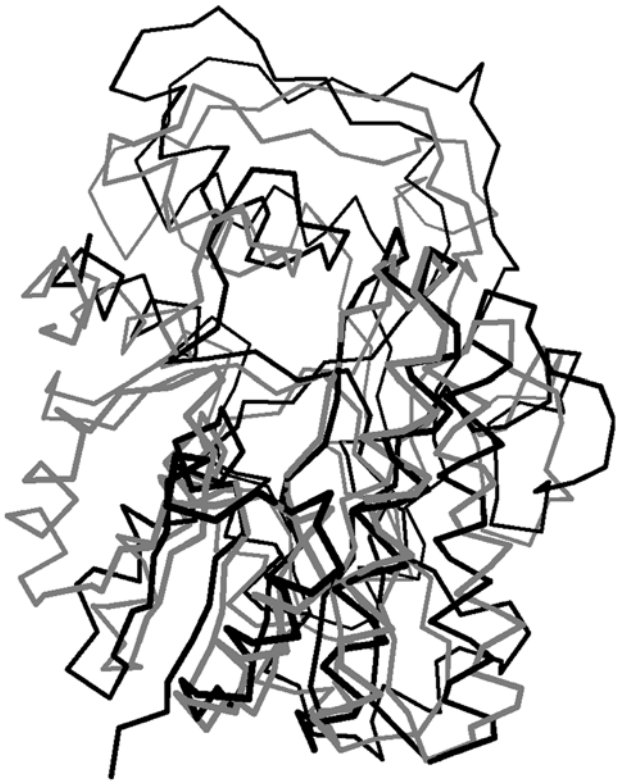

Figure 2. A structural overlay of the Cα backbones of GMER (black line) and of UGE (grey line).

GMER shows a significant structural similarity to the periplasmic glucose/galactose receptor protein from *Salmonella typhimurium*[10] and to the sugar binding domains of Lac repressor.[11] Such results poin to a possible common evolutionary origin of these proteins from an ancestor (mono)saccharide-binding protein displaying a prototype Rossmann fold topology.

The GMER C-terminal domain shows a less regular topological scheme, being based on 3 α-helices and 4 β-strands. The secondary structure elements are connected by elongated loop regions, which may play a role in structural rearrangements of this region during substrate binding (see below). Again, structural comparisons show that this domain is related to the C-terminal domain of UGE, although large local deviations (*ca.* 10 Å; see Figure 2) can be observed in the loop regions of the two proteins. Such differences may be related to the different substrate binding properties that have to be achieved in the two enzymes.

Binding of the nucleotide by GMER is achieved according to the known $NADP^+$ molecular recognition principles, at the Rossmann fold topological switch point. Upon binding, the adenine moiety is buried in a sort of surface cleft, which is sealed by the side chain of residue Arg36, which is a disordered side chain in the apo-enzyme structure.[4] Hydrogen bonds to the adenine N6 atom are provided by residue Asn40 and a water molecule, hydrogen bonded to Gln82 OE1. Moreover, a hydrophobic environment is provided by the side chains of Met14, Val15, Val41, Leu42, Val66, Ile86 and Leu166. The adenine-ribose phosphate group is salt linked to residue Arg36; the phosphate negative charge is also compensated by the side chain of the neighbouring Arg12. Residue 36 is properly positioned to discriminate between NAD^+ and $NADP^+$; in fact, the presence of a glutamyl residue at this site in homologous enzymes provides a strong selection against $NADP^+$ binding.

The $NADP^+$ diphosphate bridge runs antiparallel to a Gly-rich motif, as typically observed in many dehydrogenases, where it provides conformational flexibility for the protein–coenzyme interaction. In contrast to many dehydrogenases displaying the Gly-X-X-X-Gly-X-Gly motif, the Gly-rich stretch in GMER has the sequence Gly(10)-X-X-Gly(13)-X-X-Gly(16), a motif specifically present and conserved in the short chain dehydrogenase family.[4,12,15] No direct hydrogen bonds occur between the diphosphate oxygen atoms and residues Arg12 or Arg36, the binding of diphosphate in this region being stabilized by an α-helix dipole. Moreover, diphosphate oxygen atoms are hydrogen bonded to peptidic N of Met14, to a water molecule, bound to peptidic N of Gly16 and to the O atom of Ala62.

The ribose-nicotinamide end of the nucleotide provides interactions with the expected centre of the enzyme active site, in keeping with the catalytic

nature of the nicotinamide extremity of $NADP^+$. In particular, it is found that the ribose moiety is strongly connected to the protein through hydrogen bonds at the OH(2) centre (with Leu105 O, 2.85 Å; Gly106 O, 2.35 Å; Lys140 NZ, 2.87 Å) and at the OH(3) centre (with Lys140 NZ, 2.96 Å; Tyr136 OH, 2.80 Å) (see Figure 3a). The electron density of the nicotinamide ring is not completely defined, even at 1.45 Å resolution, presumably due to the rotational disorder of the aromatic ring in the absence of a substrate or inhibitor molecule. Indeed, despite the presence of a well defined and structured enzyme cavity (residues Met14, Val15, Ser107, Tyr136, Pro163, Asn165, His179), potentially able to host the nicotinamide ring and to provide hydrophobic contacts, enough room is observed in the present structure to allow free rotation of the reducing end of the coenzyme.

The location of the nicotinamide end, which can nevertheless be identified next to residues Ser107, Tyr136 and Lys140, is in keeping with previous observations on short chain dehydrogenase enzymes, for which such a catalytic triad is conserved in more than 150 known amino acid sequences.[12–14] The high resolution structure shows that no direct hydrogen bonding occurs between the side chains of Tyr136 and Lys140, the 136OH···140 NZ atoms distance being 4.27 Å (see Figure 3b). Nevertheless, also due to the low dielectric constant of the surrounding region, the Lys140 positive electrostatic field is expected to have a profound influence on the Tyr136 pK_a, with a shift to rather low values. In fact, a parallel study on UGE, which displays a very similar active site structure, has shown that Tyr136 displays a pK_a of about 6.1.[15] It is expected, therefore, that Tyr136 is present as a tyrosinate anion in the resting enzyme. Ser107, on the other hand, is very loosely hydrogen bonded to the NO7 N-atom of the $NADP^+$ nicotinamide tail.

Despite several different experimental approaches, by us and by others, it has not been possible so far to bind a substrate analogue to GMER. This difficulty is reflected, on the one side, in unfavourable packing contacts limiting access to the substrate binding site in the trigonal crystal form here described, but also in potential conformational readjustements that may occur during substrate binding. In fact, as mentioned above, a structural comparison with the C-terminal domain of UGE (the latter enzyme crystallized in the presence of the inhibitor UDP-glucose) shows that the substrate binding site can be properly adapted to the shape and polarity of the bound species, accounting for the rather evident conformational differences observed between the two enzymes in this region.

The crystal structure of UGE in the presence of an inhibitor has been used previously to analyse the potential substrate binding mode in GMER, through structual superposition of the homologous active site regions.[4]

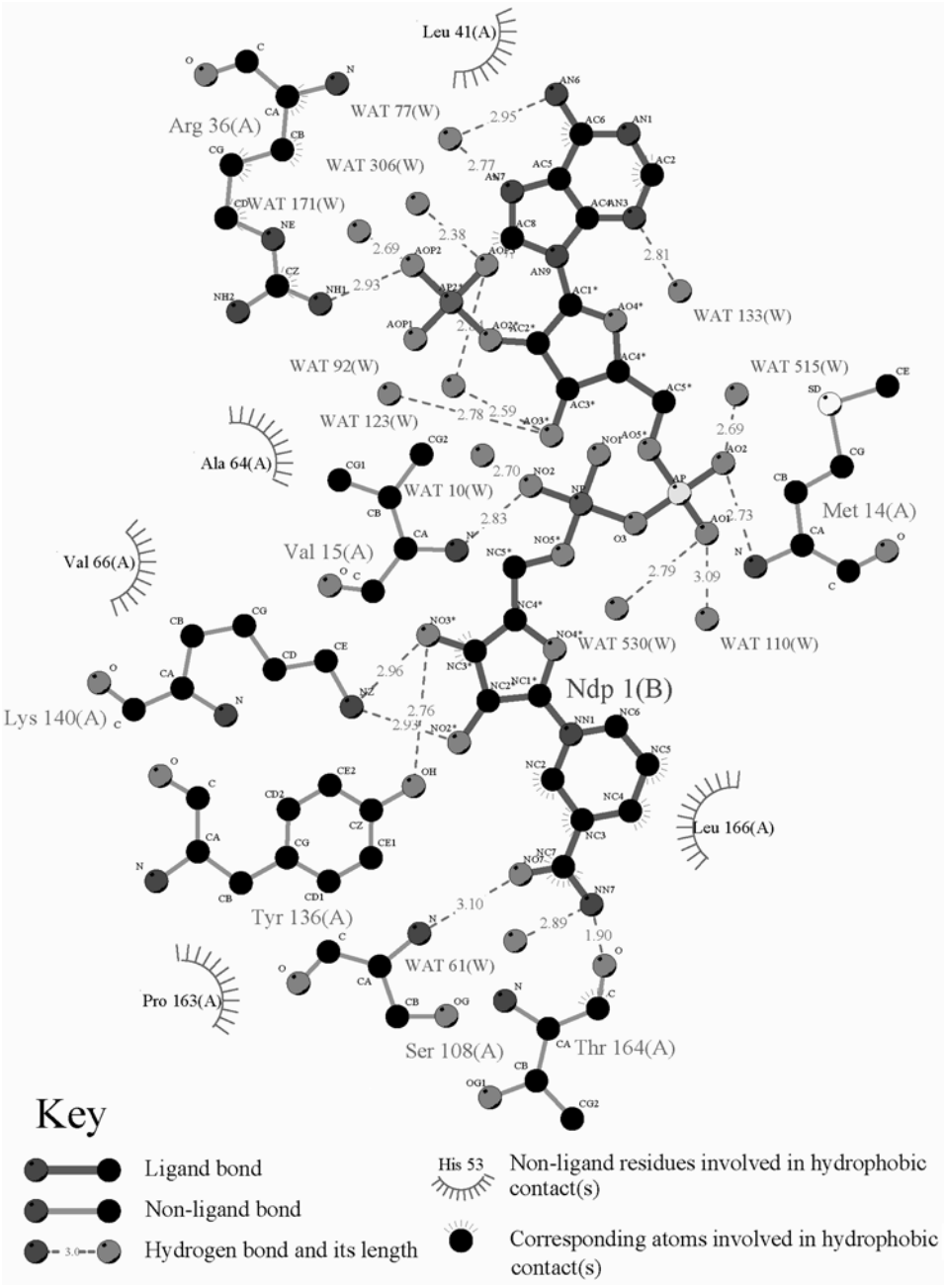

Figure 3a. Structural surroundings of the NADP+ molecule, rendered in a LIGPLOT[16] representation.

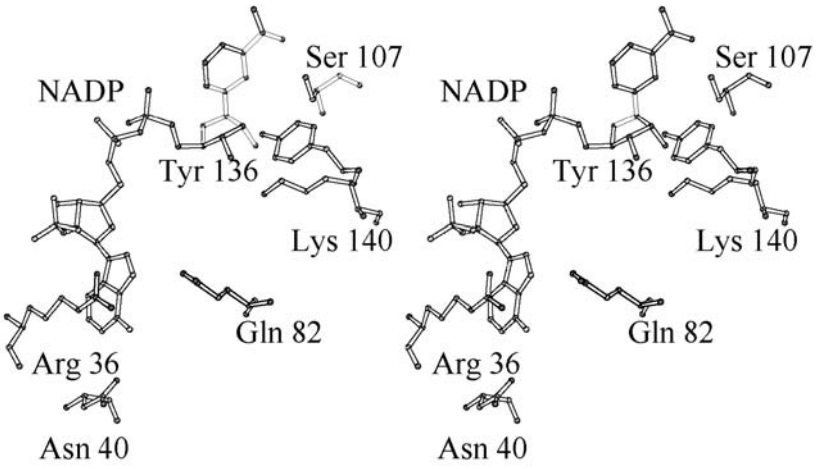

Figure 3b. The structural environment of the active site catalytic triad in GMER. The orientation of nicotinamide ring reflects the best interpretation of the electron density, which is defined only for part of the aromatic ring. Modeling experiments suggest that the 4-keto-6-deoxymannose ring of the substrate should fall next to the nicotinamide end of $NADP^+$. (Drawn with MOLSCRIPT.)[17]

Within the limits of a homology modelling study, this investigation has shown that indeed the mannose ring of the GMER substrate (the 4-keto-6-deoxy mannose moiety) falls next to the nicotinamide end of $NADP^+$, and that several potential substrate recognition interactions can be achieved in the GMER C-terminal domain, which is expected to adapt to the bound substrate, particularly in the guanosine region, after binding. The modelling shows also that in the neighbourhood of the 4-keto-6-deoxy mannose moiety a number of hydrogen bond donors/acceptors from the N-terminal domain are already positioned in order to provide stabilizing interactions to the sugar OH side groups. Potential candidates suggested by the modelling are Ser107, Ser108, Cys109, Asn165 and His179. The extent of these interactions, and the nature of the enzyme residues involved, may however vary remarkably as a function of the precise position occupied by the substrate in the active site cleft.

CONCLUSION

The above sructural data can be merged in a tentative model of the enzyme catalytic cycle. It should be noted that GMER is a bifunctional enzyme, and that the two catalytic events are believed to take place in a sequential order. In fact, since reduced NADPH is not regenerated on site at

the end of each catalytic cycle, removal of the oxidized coenzyme and its substitution with a reduced species are necessary events before an enzymatic cycle is carried over to a new substrate molecule. Since epimerization cannot occur on the reduced substrate, it is believed that reduction of the (properly epimerized) substrate is the last event in the catalytic cycle of GMER, after which the product and coenzyme diffusion from the active site takes place.

Epimerization of the C3 and C5 centres of 4-keto-6-deoxy mannose can be achieved through tautomeric equilibria, which are based on the C4 keto group, and involve, in two independent events, both C3 and C5. The role of the active site catalytic triad (Ser107, Tyr136 and Lys140) in this context may be that of a general acid/base catalyst (particularly Tyr136), allowing the exchange of the protons required for the epimerization reactions. However, stabilization of the correct keto/enol form may also be based on additional neighbouring hydrogen bond donors/acceptors, provided by residues that have been revealed by the substrate binding modelling experiment. It should be noted that, in contrast to UGE, where a single epimerization event occurs (at C4 of the galactose substrate ring), epimerization at C3 and C5 in GMER does not appear to require rotation of the substrate saccharide ring by 180° within the active site cleft.

Once the correct epimeric form of the substrate is achieved, yielding GDP-4-keto-6-deoxy-L-galactose, proper occupation of the active site cleft by the sugar ring may bring the 4-keto centre of the substrate into close proximity of the nicotinamide C4 atom, from which hydride transfer may take place. Again, the catalytic triad residues may play a role in the proton shuttling, but also other residues surrounding the substrate saccharide ring are expected to play a key role in the stabilization of the reaction intermediates.

A final consideration concerns the absence of quaternary structure modifications observed in a comparison of the holo-enzyme (this study) and the apo-enzyme[4] structures. Although a final word on this topic, which relates to possible cooperativity in substrate binding in a dimeric enzyme, must await a comparison with a substrate-bound complex structure, the absence of any tertiary and quaternary structure adaptation to the presence/absence of the coenzyme may reflect the total lack of communication between the two active sites, which are facing different directions and are separated by a distance of about 50 Å. Indeed, activity measurements in solution do not indicate the presence of any cooperativity in GMER.

The high resolution structure of GMER is an unusual achievement, since proteins of this size do not generally lead to highly ordered crystals. In this case, however, the compact nature of the dimeric enzyme, and the application of a high brilliance synchrotron source have provided a precise

view of the enzyme's active site. The information gained lends itself to verification through the design of site-specific GMER mutants, in which rational substitution of active site residues will shed light onto details of the catalytic mechanism and of the substrate recognition process, which are presently still uncharacterized.

The coordinates of the high resolution structure of GMER in a complex with $NADP^+$ will be deposited with the Protein Data Bank.[9]

Acknowledgments. – This work was partly supported by a grant from the Ministry of University and Scientific Research, MURST-PRIN 1998 and by the ASI grant ARS-98-174, to M.B.

REFERENCES

1. A. Varki, *Proc. Nat. Acad. Sci. USA* **91** (1994) 7390–7397.
2. A. Etzioni and R. Gershoni-Baruch, *N. Engl. J. Med.* **327** (1992) 1789–1792.
3. Z. Otwinowski and W. Minor, *Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology, 276,* in: C. W. Carter Jr. and R. M. Sweet (Eds.), *Macromolecular Crystallography*, Part A, Academic Press, Cambridge, 1997, pp. 307–326.
4. M. Rizzi, M. Tonetti, P. Vigevani, L. Sturla, A. Bisso, A. De Flora, D. Bordo, and M. Bolognesi, *Structure* **6** (1998) 1453–1465.
5. G. N. Murshudov, A. A. Vagin, and E. J. Dodson *Acta Crystallogr.*, *Sect D* **53** (1997) 240–257.
6. D. Gosh, V. Z. Pletnev, D-W. Zhu, Z. Wawrzak, W. L. Duax, W. Pangborn, F. Labrie, and S-X. Lin, *Structure* **3** (1995) 503–513.
7. C. Branden, and J. Tooze, *Introduction to Protein Structure,* Garland Publishing, New York, 1991.
8. A. J. Bauer I. Rayment, P. A. Frey, and H. M. Holden, *Proteins* **12** (1992) 372– 381.
9. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Jr. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol*. **112** (1977) 535–542.
10. J-Y. Zou, M. M. Flocco, and S. L. Mowbray, *J. Mol. Biol*. **223** (1993), 739–752.
11. M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H.C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, *Science* **271** (1996) 1247–1254.
12. H. Jornvall, B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffrey, and D. Gosh, *Biochemistry,* **34** (1995) 6004–6012.
13. M. E. Baker, and R. Blasco, *FEBS Lett.,* **301** (1992) 89–93.
14. G. Labesse, A. Vidal-Cros, J. Chomilier, M. Gaudry, and J.-P. Mornon, *Biochem. J*. **304** (1994) 95–99.
15. Y. Liu, J. B. Thoden, J. Kim, E. Berger, A. M. Gulik, F. J. Ruzicka, H. M. Holden, and P. A. Frey, *Biochemistry* **36** (1997) 10675–10684.
16. A. C. Wallace, R. A. Laskowski, and J. M. Thornton, *Protein Eng*. **8** (1995) 127–134.
17. P. J. Kraulis, *J. Appl. Crystallogr.* **24** (1991) 946–950.

SAŽETAK

Struktura GDP-4-keto-6-deoksi-D-manoza-empimeraze/reduktaze određena s visokim razlučivanjem

Camillo Rosano, Gaetano Izzo, Laura Sturla, Angela Bisso, Michaela Tonetti i Martino Bolognesi

GDP-4-keto-6-deoksi-D-manoza-epimeraza/reduktoza bifunkcionalan je enzim uključen u biosintezu struktura statične površine, takvih kao što su antigeni krvnih grupa. Svaka podjedinica u homodimernom enzimu sastoji se od dvije domene. N-terminalna domena pokazuje topologiju Rossmanova pregiba i veže koenzim $NADP^+$. Za C-terminalnu domenu smatra se da veže supstrat. Struktura holoenzima utočnjena je s rezolucijom 1,45 Å, iz sinkrotronskih podataka, do konačnog faktora R = 0,127 (R_{slobodni} = 0,167). Utočnjeni proteinski model jasno pokazuje nekoliko ostataka uključenih u prepoznavanje i vezanje koenzima i daje naslutiti da enzim pripada proteinskoj homolognoj porodici kratkolančanih dehidrogenaza. Diskutira se o implikacijama za katalitički mehanizam.