

A structural model for the N-terminal N1 module of *E. coli* glycogen branching enzyme

Leila LO LEGGIO^{1*}, Heidi A. ERNST¹, Ida HILDEN² & Sine LARSEN¹

¹Centre for Crystallographic Studies, Chemistry Institute, Universitetsparken 5, DK-2100 Copenhagen N, Denmark; tel.: ++ 45 35320295, fax: ++ 45 35320299, e-mail: leila@ccs.ki.ku.dk

²Danisco Cultor Innovation, Langebrogade 1, DK-1001 Copenhagen K, Denmark (current address: M&E Biotech, Kogle Allé 6, DK-2970 Hørsholm, Denmark)

LO LEGGIO, L., ERNST, H. A., HILDEN, I. & LARSEN, S., A structural model for the N-terminal N1 module of *E. coli* glycogen branching enzyme. *Biologia, Bratislava*, 57/Suppl. 11: 109–118, 2002; ISSN 0006-3088.

Prokaryotic branching enzymes can be divided into Group 1, with long N-domains, and Group 2, with short N-domains. Sequence analysis and fold-recognition approaches suggest that in Group 1 enzymes the N-domain has originated from duplication of a module with an immunoglobulin-type fold, similar in fold to the N-domain found in some other family 13 enzymes. These modules are here referred to as N1 and N2. Group 2 enzymes only have one such module. A three-dimensional model of the N1 module of *E. coli* GBE, a Group 1 branching enzyme, has been constructed based on secondary structure alignment to the N-domains of *Pseudomonas amyloclavata* isoamylase and *Sulfolobus solfataricus* glycosyltrehalose trehalohydrolase.

Key words: glycogen branching enzyme, N-domain, α -amylase family, domain duplication, homology model, structure.

Abbreviations: 3-D, three-dimensional; BE, branching enzyme; FAD, flavin adenine dinucleotide; GBE, glycogen branching enzyme; GTT, *Sulfolobus solfataricus* glycosyltrehalose trehalohydrolase; nGBE, N-terminally-truncated glycogen branching enzyme; PAI, *Pseudomonas amyloclavata* isoamylase; PDB, protein data bank; rmsd, root mean square deviation; SBE, starch branching enzyme; TMA, *Thermus maltogenicus* amylase; TVII, *Thermoactinomyces vulgaris* R-47 amylase II.

Introduction

Glycogen branching enzymes (GBE) (EC 2.4.1.18) catalyse the cleavage of an α -1,4 linkage in a glycogen chain to form an α -1,6-linked branch. This is an essential step in glycogen biosynthesis. GBEs belong to family 13 in the glycoside hydrolase classification (COUTINHO & HENRISSAT, 1999) which in turn belongs to the α -amylase superfamily of

enzymes (JESPERSEN et al., 1991, 1993). A variety of enzyme activities are represented in the family, as most recently reviewed by MACGREGOR et al. (2001).

All enzymes in the superfamily possess a catalytic domain with 8-fold $\beta\alpha$ -barrel architecture (commonly referred to as domain A), but vary in the presence or absence of extra domains. In particular, some members of family 13 possess

* Corresponding author

Table 1. List of branching enzyme sequences used in this study.

	Name	Accession no.	Reference
Group 1	<i>Agrobacterium tumefaciens</i> GBE	P52979	UGALDE et al. (1998)
	<i>Escherichia coli</i> GBE	P07762	BAECKER et al. (1986)
	<i>Haemophilus influenzae</i> GBE	P45177	FLEISCHMANN et al. (1995)
	<i>Mycobacterium tuberculosis</i> GBE	Q10625	COLE et al. (1998)
	<i>Neisseria denitrificans</i> BE	AAF_04747	BUETTCHER et al. (1999)
	<i>Pasteurella multocida</i> GBE	AAK02625	MAY et al. (2001)
	<i>Streptomyces aureofaciens</i> GBE	P52980	HOMEROVA & KORMANEC (1994)
	<i>Streptomyces coelicolor</i> BE I	S70079	BRUTON et al. (1995)
	<i>Streptomyces coelicolor</i> BE II	S34218	BRUTON et al. (1995)
	<i>Synechococcus</i> sp. BE	JQ0550	KIEL et al. (1990)
	<i>Synechocystis</i> sp. GBE	P52981	KANEKO et al. (1995)
Group 2	<i>Arabidopsis thaliana</i> SBE II	NP_181180	LIN et al. (1999)
	<i>Bacillus caldolyticus</i> GBE	P30537	KIEL et al. (1992)
	<i>Bacillus stearothermophilus</i> GBE	P30538	KIEL et al. (1991)
	<i>Bacillus subtilis</i> GBE	P39118	KIEL et al. (1994)
	<i>Butyrivibrio fibrisolvens</i> GBE	P30539	RUMBAK et al. (1991)
	<i>Drosophila melanogaster</i> CG4023 Gene product	AAF58416	ADAMS et al. (2000)
	<i>Gracilaria gracilis</i> SBE	AAB97471	LLUISMA & RAGAN (1998)
	<i>Homo sapiens</i> GBE	NP_000149	THON et al. (1993)
	<i>Saccharomyces cerevisiae</i> GBE	P32775	THON et al. (1992)
	<i>Solanum tuberosum</i> SBE II	CAB40749	JOBLING et al. (1999)

an N-terminal extension, which has been generally referred to as the N-domain (JESPERSEN et al., 1991). Structure determination of several family 13 enzymes has shown that this domain is not always structurally conserved. In *Pseudomonas amyloclavata* isoamylase (PAI) (KATSUYA et al. 1998), *Thermoactinomyces vulgaris* R-47 amylase II (TVII) (KAMITORI et al., 1999), *Sulfolobus solfataricus* glycosyltrehalose trehalohydrolase (GTT) (FEESE et al., 2000), and *Thermus* maltogenic amylase (TMA) (KIM et al., 1999), the N-domain has an immunoglobulin-type fold, while in amylosucrase (SKOV et al., 2001) it displays an all- α architecture.

An N-terminally truncated form of *E. coli* glycogen branching enzyme (nGBE, HILDEN et al., 2000) missing the first 107 residues has been produced and found to retain half of the activity of full length GBE. Similar findings have been independently reported by others (BINDERUP et al., 2000). Both groups reported crystallisation of the truncated forms of GBE. More recently a preliminary report of the structure has been given (PREISS et al., 2001). However, to our knowledge, nobody has succeeded in obtaining crystals of intact GBE, suitable for structure determination. The structure and function of the N-terminal domain in GBE remains thus unclear.

Prokaryotic branching enzymes have been

previously divided into two groups (HILDEN et al., 2000) depending on the length of their N-domain. *E. coli* GBE belongs to Group 1 branching enzymes, having a long N-domain. Here we present further sequence analysis, which suggests domain duplication within the N-domain of Group 1 enzymes. A structural model for the N-terminal domain of GBE has been constructed according to this hypothesis.

Material and methods

Sequence analysis

Initial homology searches were carried using PSI-BLAST (ALTSCHUL et al., 1997) through the NCBI server at <http://www.ncbi.nlm.nih.gov:80/BLAST/>, using the BLOSUM62 matrix and a gap cost of 11 for existence and 1 for extension.

Sequence alignments and the distance tree were generated using multalign and visualized using ALSRIPT (BARTON, 1993). Secondary structure predictions shown were made using the consensus method of the J-Pred server at <http://circinus.ebi.ac.uk:8888/submit.html> (CUFF et al., 1998). The sequences used are listed in Table 1.

Three fold-recognition methods were used:

- 3D-PSSM at <http://www.sbg.bio.ic.ac.uk/~3dpssm/> (KELLEY et al., 2000);
- the Bioingbu server at <http://www.cs.bgu.ac.il/~bioingbu/> (FISCHER et al., 2000); and
- mGenThreader at <http://bioinf.cs.ucl.ac.uk/psipred/> (JONES, 1999).

Construction, validation and analysis of a three-dimensional model for N1

An initial model for the N1 module of *E. coli* GBE was constructed manually in O (JONES et al., 1991) based on the backbone conformation of the N-terminal domain of PAI (PDB code 1bf2) supplemented, where appropriate, with the backbone of the equivalent domain of GTT (PDB code 1eh9) with few adjustments. The sequences were aligned based on predicted secondary structure elements or strongly-conserved short sequence stretches. Side chains were then introduced according to the *E. coli* N1 sequence. The model comprised residues 24-103. Residues 5-15 were predicted by J-Pred to assume an α -helical conformation, and could fold as an amphipathic helix, which might pack against the N1 module. However, given the lack of such a helical region in the available templates, residues 1-23 were not included in the model for N1.

The geometry of the model was optimised both visually and using CNS (BRÜNGER et al., 1998) and validated using PROCHECK (LASKOWSKI et al., 1993), WHAT_CHECK (HOOFT et al., 1996) and O. The structure-based sequence alignment between model and templates was output by O and further edited manually. The molecular surface was displayed in GRASP (NICHOLLS et al., 1991), and cartoon representations were displayed using Molscript (KRAULIS, 1991).

Results

Sequence duplication at the N-terminus of Group 1 glycogen branching enzymes

A PSI-BLAST search (3 iterations), using the N-terminal portion of the GBE sequence (res 1-107), identified several hits, most of which were annotated in the database as branching enzymes. The identity region identified by BLAST was always at the N-terminus of the sequence, and often two hits in consecutive portions of the N-terminus were identified. This leads us to the hypothesis that gene duplication has occurred at the N-terminus of some BEs, creating two structurally-similar modules, each approximately 100 residues long. We refer to the two N-terminal modules as N1 and N2 (Fig. 1).

After three PSI-BLAST iterations searching with *E. coli* N1 (residues 1-106), four sequences showed the presence of two N-terminal modules, namely *E. coli* GBE, *M. tuberculosis* GBE, *Synechocystis* sp. GBE and *P. multocida* GBE. The first three sequences had previously been classified as belonging to Group 1. Further PSI-BLAST iterations using GBE N1 as search sequence found evidence for N2 modules in many other Group 1 sequences (listed in Figure 2). Similar results could be obtained by searching with *E. coli* N2 (residues

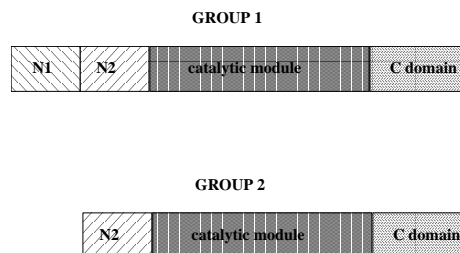


Fig. 1. Schematic representation of domain structures in branching enzymes.

108-208) for Group 1 enzymes. On the other hand, only one N-terminal module could be identified in Group 2 enzymes, searching either with *E. coli* N1 or N2. Searches using *E. coli* N2 identified plant and mammalian BEs as well as microbial enzymes, so that in this sense they can also be assigned to Group 2. We propose the general domain structure in Figure 1 for enzymes from the two groups. Thus what previously has been referred to as the N-domain comprises the N1 and N2 modules in Group 1 enzymes and only the N2 module in Group 2.

Microbial sequences identified by PSI-BLAST for N1 and N2 modules for Group 1 and selected Group 2 sequences (including selected eukaryotic ones) were aligned. A distance tree (Fig. 2) was generated, showing that N1 modules are more similar to each other than to the N2 modules. Among the N2 modules, prokaryotic and eukaryotic sequences clearly form distinct groups. The highest identity between an N1 and an N2 sequence was found between *N. dinitrificans* BE N2 and *S. coelicolor* BE II N1, with 31 identical residues over the lengths of the two sequences (102/108 residues) (Fig. 3). Alignment of N1 and N2 sequences as two separate groups showed five areas with high sequence conservation in N1 and six in N2 modules. These are shown boxed in the alignment in Figure 3, and correspond to areas of high sequence conservation between N1 and N2, except for one of the blocks that is missing in N1s.

Secondary structure prediction

As a first step towards a structural model for the N1 module, the secondary structure for *E. coli* N1 was predicted using the J-Pred server. The consensus predicted secondary structure is shown in Figure 3 for the N1 module, and for comparison also for the N2 module. Except for the initial por-

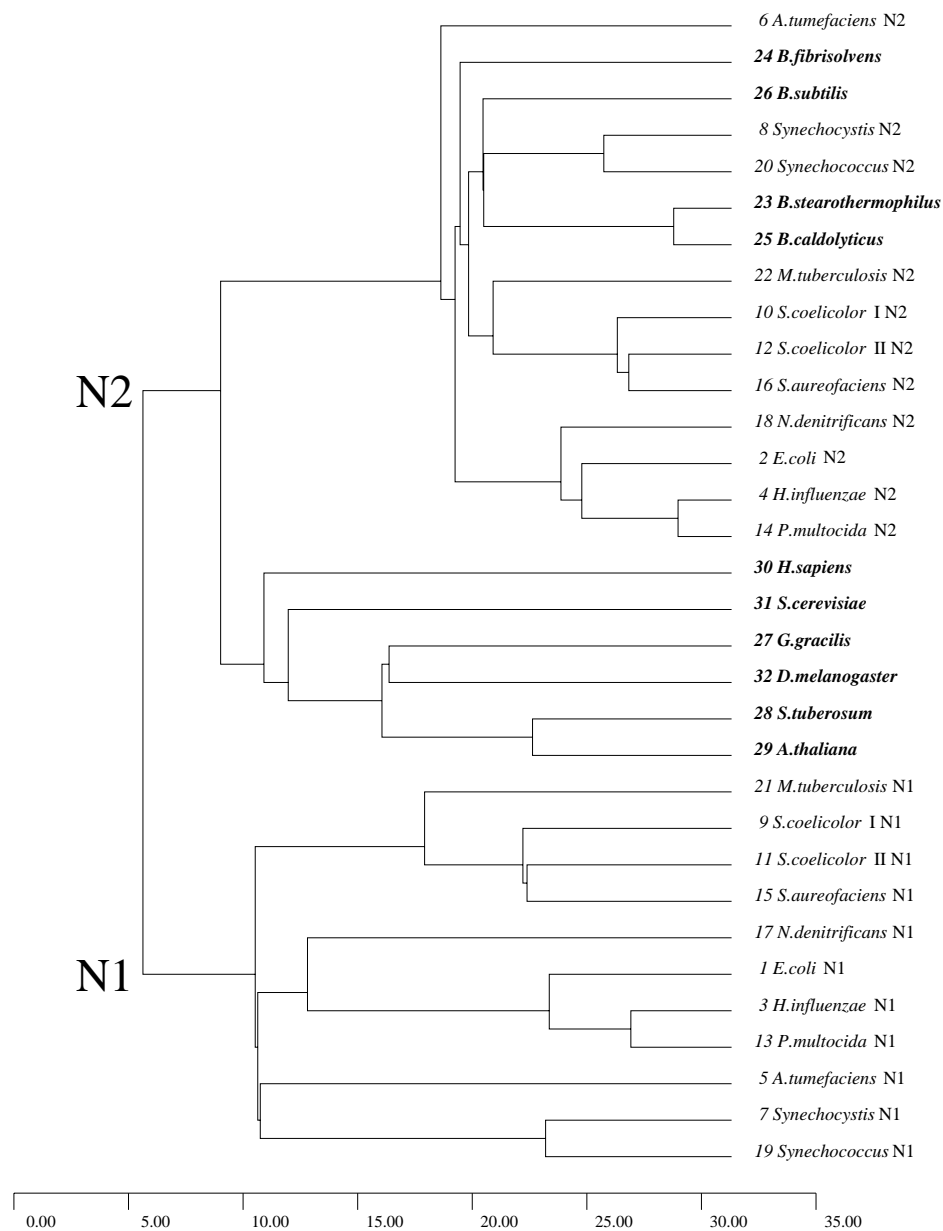


Fig. 2. Distance tree of N1 and N2 module sequences output by multalign (based on RSCORE). Group 2 enzymes are in bold.

tion of the sequence, which is predicted to be a helix, the sequence of N1 is predicted to consist mostly of β -strands. N2 is similarly predicted to consist mostly of β -strands. The predicted secondary structure elements in N1 and N2 are generally in good agreement except for the initial α -helix in N1 and the initial β -strand in N2. The secondary structure elements are found mostly in the

highly conserved blocks, and gaps in the alignment are not found in the predicted secondary structure elements. Similarity between the N2 region of *E. coli* GBE and the N-domain of *Pseudomonas amylocleramosa* isoamylase has been previously suggested by BINDERUP *et al.*, (2000); to our knowledge, however, no fold assignment has been made for the N1 portion.

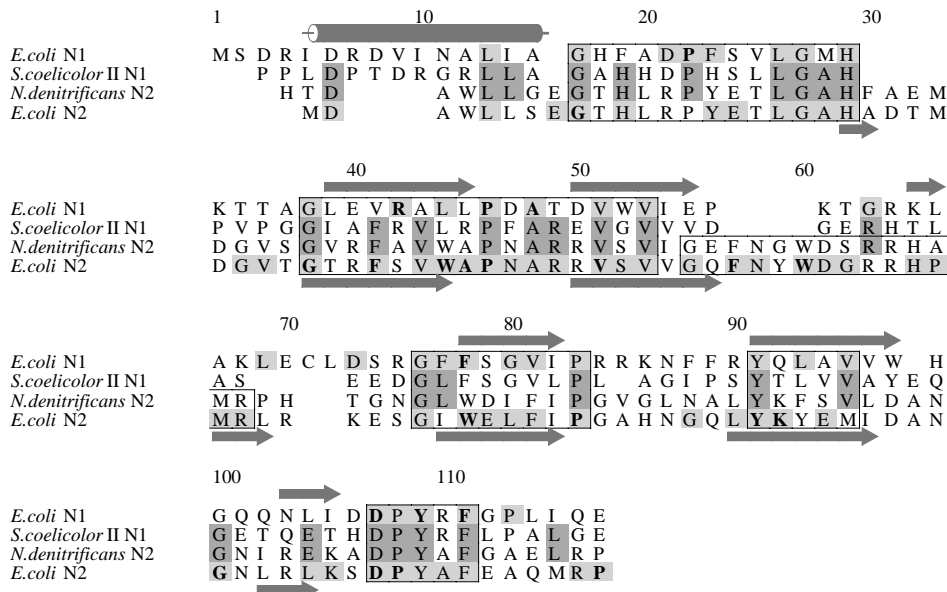


Fig. 3. Alignment of N1 and N2 module sequences. The two most similar N1 and N2 sequences are aligned (*S. coelicolor* BE II N1 and *N. denitrificans* N2) with conserved residues shadowed in dark grey. *E. coli* N1 and N2 sequences are also included in the alignment. Residues shadowed in light grey in the *E. coli* N1 and N2 sequences are conserved in more than 50% of sequences in multiple alignments (not shown) of N1 modules or N2 modules, respectively. Residues in bold are fully conserved in the multiple alignments of N1 modules or N2 modules, respectively. Six conserved blocks are found in the branching enzyme N1 and N2 modules (only five in N1 sequences), and these are boxed in the figure.

Fold prediction

The eleven N1 sequences in Group 1 which had been identified through the BLAST search were submitted to three fold recognition servers. A similar analysis (not shown) using N2 sequences gave very high probabilities that this module folds as the immunoglobulin-like N-domains of PAI, GTT, TVII and TMA, as previously suggested, so the sequence similarity between N1 and N2 would suggest in itself that N1 would also have an immunoglobulin-type fold. However, given the low sequence identity between N1 and N2 sequences, the compatibility of N1 sequences with this fold was investigated independently.

The results of the fold recognition searches for N1 are summarised in the following:

a) Bioingbu: the *E. coli* N1 sequence identified three N-domains (GTT, TVII, and TMA) as the top hits with good discrimination against the next best score using the GONP method. Similar results were obtained with this method for all search sequences, although the N-domain of PAI was also often among the top scores, and sometimes not all the immunoglobulin-like N-domains were represented. 43 out of 55 top scores and 14 out of 15

significant top scores obtained with the five individual scoring methods, as well as 10 out of 11 top scores obtained with the consensus method were domains with an immunoglobulin-type fold.

b) 3D-PSSM: the N-domains of PAI and of TVII came up as top hits for search sequences from *S. coelicolor* BE II and *P. multocida* GBE, respectively, with reasonably high probabilities (38% and 37%, respectively). PAI also came up as top hit for the *Synechocystis* sp. search sequence, but with low probability. The only other significantly high score was obtained with the *Synechococcus* sp search sequence, which gave as top hit the C-terminal jelly-roll domain of family 13 oligo-1,6-glucosidase (WATANABE et al., 1997) with 61% probability. Three search sequences came up with the FAD-binding central domain of flavohemoglobin (ERMLER et al., 1995) as the top hit (all with very low probabilities); otherwise the remaining search sequences returned as top hits only domains with immunoglobulin-type folds.

c) mGenTHREADER: for eight out of eleven search sequences, the top scoring hit was an immunoglobulin-like N-domain. The N-domain of

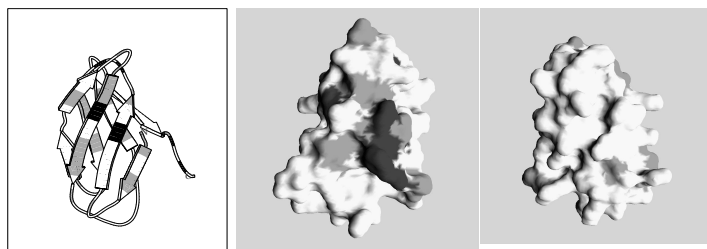


Fig. 4. (a) Molscript cartoon of the model for the N1 module, showing residues which are fully conserved (dark grey) or more than 50% conserved (light grey) in a multiple sequence alignment of N1 modules. (b) and (c) are GRASP representations of the model for the N1 module, coloured according to the same colour scheme as (a). (c) is rotated 200° around a vertical rotation axis with respect to (b).

GTT was the top scoring alignment for two search sequences (*Synechococcus* and *M. tuberculosis* N1) with high confidence (99%). The N-domain of PAI was the top scoring alignment for four search sequences, in the case of *S. coelicolor* BE II N1 with high confidence (99%) and in the three other cases with marginal confidence level (40%). The N-domain of TVII was the top scoring alignment for two search sequences with marginal to low confidence (40% to 70%). The remaining three top hits were unrelated to immunoglobulin-like domains.

Taken together, the fold-recognition results strongly suggest that if the N1 module assumes a known fold, this is the same as the immunoglobulin-type fold found in other family 13 enzymes' N-domains, and strongly predicted for the N2 module.

Construction and validation of a 3-D model for N1
A model for N1 was constructed as described in 'Materials and Methods' and subjected to several validation tests. The model comprises residues 24-103 of *E. coli* N1 and therefore does not include the initial predicted helix, or residues 16-23 which are part of one of the conserved blocks in N1-N2.

62% of all non-Ala, non-Gly residues could be modelled in the most frequent rotamer according to the rotamer database in O without steric clashes. 80.6% of all residues were in the core regions of the Ramachandran plot, 17.9% in the allowed regions, with only one residue (His 66) in the generously allowed regions. According to PROCHECK, main-chain and side-chain parameters were better than average compared to structures determined to at least 2.5 Å resolution, with rmsds of 0.005 Å for bond lengths and 2.36° for bond angles, respectively, indicating that the geometry has been properly restrained during refinement. No bad contacts were detected by

PROCHECK. The WHAT.CHECK RMS Z-score for the inside/outside distribution is 1.132, close to the ideal value of 1, and the 2nd generation packing quality Z-score is -2.949, within the normal range. All these quality indicators show that it is possible to fold the sequence of N1 as an immunoglobulin-like N-domain, thus obtaining a model that has the characteristics of normal protein structure in terms of packing and geometry. Figure 4a shows a ribbon diagram of the final model, which is deposited in the PDB with code 1gty.

Discussion and conclusions

As shown by our results, and assuming that this module has a known fold, the most probable fold of GBE N1 is the immunoglobulin-type fold already found in non-catalytic N-domains of other family 13 enzymes. Figure 5a shows a structure-based alignment between the N1 model and the other immunoglobulin-like N-domains, including the ones used as templates. Although strictly speaking there is little or no sequence conservation, and the N-domains are not all oriented in the same way with respect to the catalytic domain, colouring by residue type highlights the similarities between the five sequences, especially in second, third and fourth conserved blocks in N1 modules, suggesting that they might have originated by divergent evolution.

Figure 4 shows the pattern of conservation of residues within an N1 module mapped on a cartoon representation of the *E. coli* N1 model (Fig. 4a) and on the surface (Figs 4b,c). Alternate residues on the β -strands are sometimes conserved (Fig. 4a), and some of these are the residues contributing to hydrophobic packing between β -strands. A patch of conserved residues is formed on one side of the N1 module (Fig. 4b), while little

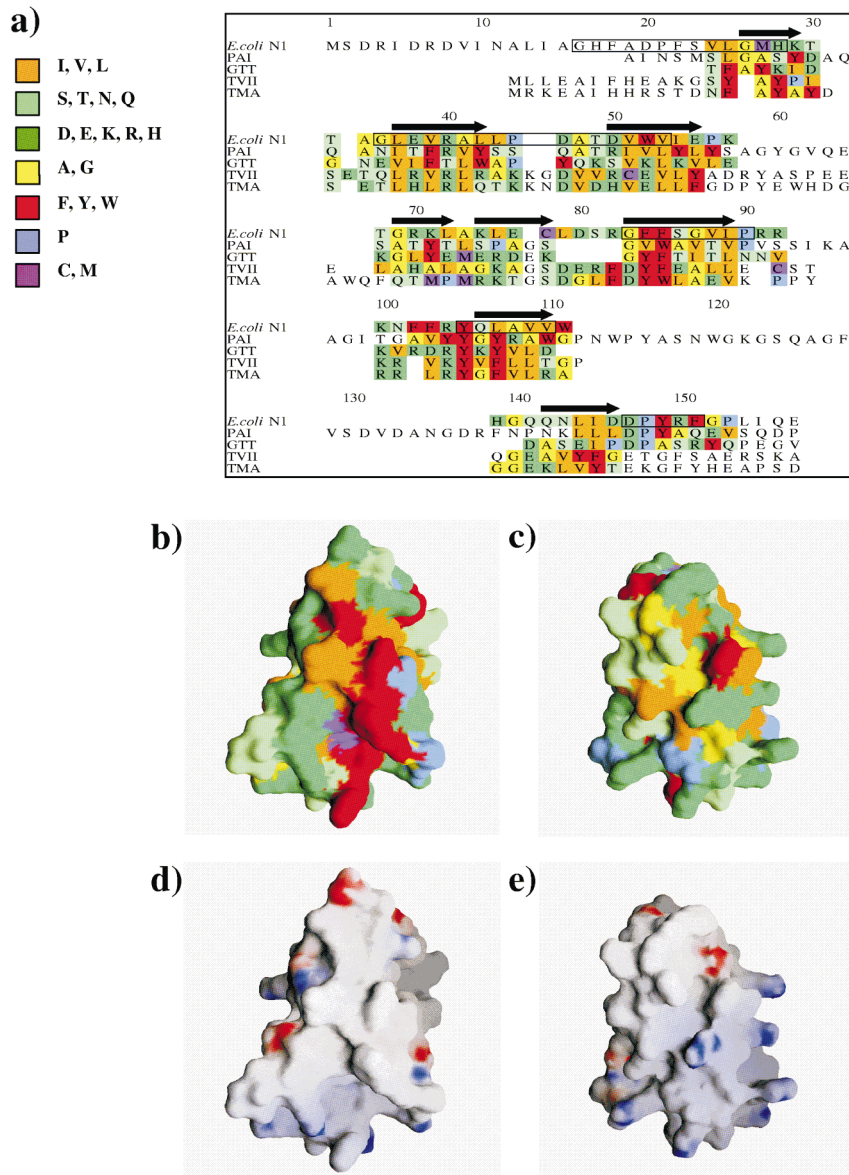


Fig. 5. Structure-based alignment of the N1 model with the immunoglobulin-type N-domains. Only residues coloured in the N1 sequence have been modelled, and only residues coloured in the other sequences have structural equivalents in the N1 model. Residues are coloured according to residue type. The conserved blocks in N1 sequences are boxed. (b) and (c) show GRASP representations for the N1 model coloured according to the colour scheme in (a) and in the same orientations as in Figures 4b,c, (d) and (e) also show GRASP surface representations in the same orientations, but coloured according to charge (blue for positive and red for negative).

conservation is found on the other side (Fig. 4c). This patch is formed by non-adjacent residues in the sequence, further validating the constructed model.

The patch consists mostly of aromatic and aliphatic residues (Figs 5b,d), while the non-conserved side is more polar, and mostly positively charged (Figs 5c,e). Several roles are possible for

this patch: it could be complementary to the amphipathic helix which has not been modelled but constitutes a large part of the extreme N-terminus of N1; it could be involved in packing of the N1 module against the catalytic domain or the N2 module, as this portion of the structure is involved either in dimer formation or packing against the catalytic domain for the known immunoglobulin-like N-domains; lastly, it could be involved in interaction with the glucan chain. As the patch is relatively extended, it is also possible that it might serve more than one role.

Removal of N1 causes a reduction in specific activity to between 38% and 62% of the activity for intact *E. coli* GBE (BINDERUP et al., 2000; HILDEN et al 2000). BINDERUP et al. (2000) also reported that this does not result from an effect on K_M , but suggested nonetheless that the truncated portion might affect the branching pattern of the product. In the dimeric TMA (KIM et al., 1999) the N-domain of one monomer participates in formation of the substrate-binding groove in another monomer. YOKOTA et al. (2001) have shown that in TVII deletion of the N-domain results in loss of dimerization, but also drastic loss of activity. Loss of the N1 module in GBE does not have such a drastic effect, perhaps because of the domain duplication, as N2 might partly compensate for the loss of the N1 module.

We have presented a structural model for the N1 module in *E. coli* GBE based on remote sequence homologies. While such a model cannot provide reliable atomic positions, it is our hope it will provide a basis and inspiration for further studies on the role of the duplicated N-domain in Group 1 branching enzymes.

Note added in proof

Since the acceptance of this manuscript, the structure of the N-terminally truncated *E. coli* glycogen branching enzyme has been reported (ABAD et al., 2002). Although coordinates are not yet available, the structure confirms that the N2-module folds as the N-domain of isoamylase. Our predicted assignment of secondary structure in N2 is highly consistent with the secondary structure reported by ABAD et al. (2002).

Acknowledgements

This work was supported by the Danish National Research Foundation and a long term EMBO Fellowship to LL.

References

- ABAD, M. C., BINDERUP, K., RIOS-STEINER, J., ARNI, R. K., PREISS, J., & GEIGER, J. H. 2002. The X-ray crystallographic structure of *Escherichia coli* branching enzyme. *J. Biol. Chem.* **277**: 42164–42170.
- ADAMS, M. D., CELNIKER, S. E., HOLT, R. A., EVANS, C. A., GOCAYNE, J. D., AMANATIDES, P. G., et al. & VENTER, J. C. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J. H., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- BAECKER, P. A., GREENBERG, E. & PREISS, J. 1986. Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* 1,4- α -D-glucan:1,4- α -D-glucan 6- α -D-(1,4- α -D-glucano)-transferase as deduced from the nucleotide sequence of the *glg B* gene. *J. Biol. Chem.* **261**: 8738–8743.
- BARTON, G. J. 1993. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**: 37–40.
- BINDERUP, K., MIKKELSEN, R. & PREISS, J. 2000. Limited proteolysis of branching enzyme from *Escherichia coli*. *Arch. Biochem. Biophys.* **377**: 366–371.
- BRÜNGER, A. T., ADAMS, P. D., CLORE, G. M., DELANO, W. L., GROS, P., GROSSE-KUNSTLEVE, R. W., JIANG, J. S., KUSZEWSKI, J., NILGES, M., PANNU, N. S., READ, R. J., RICE, L. M., SIMONSON, T. & WARREN, G. L. 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**: 905–921.
- BRUTON, C. J., PLASKITT, K. A. & CHATER, K. F. 1995. Tissue-specific glycogen branching isoenzymes in a multicellular prokaryote, *Streptomyces coelicolor* A3(2). *Mol. Microbiol.* **18**: 89–99.
- BUETTCHER, V., QUANZ, M., & WILLMITZER, L. 1999. Molecular cloning, functional expression and purification of a glucan branching enzyme from *Neisseria denitrificans*. *Biochim. Biophys. Acta* **1432**: 406–412.
- COLE, S. T., BROSCHE, R., PARKHILL, J., GARNIER, T., CHURCHER, C., HARRIS, D., et al. & BARRELL, B. G. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537–544.
- COUTINHO, P. M. & HENRISSAT, B. 1999. Carbohydrate-active enzymes: an integrated database approach, pp. 3–12. In: GILBERT, H. J., DAVIES, G. J., HENRISSAT, B. & SVENSSON, B. (eds) Recent Advances in Carbohydrate Bioengineering, The Royal Society of Chemistry, Cambridge.
- CUFF, J. A., CLAMP, M. E., SIDDIQUI, A. S., FINLAY, M. & BARTON, G. J. 1998. JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**: 892–893.

- ERMLER, U., SIDDIQUI, R. A., CRAMM, R. & FRIEDRICH, B. 1995. Crystal structure of the flavohemoglobin from *Alcaligenes eutrophus* at 1.75 Å resolution. *EMBO J.* **14**: 6067–6077.
- FEESE, M. D., KATO, Y., TAMADA, T., KATO, M., KOMEDA, T., MIURA, Y., HIROSE, M., HONDO, K., KOBAYASHI, K. & KUROKI, R. 2000. Crystal structure of glycosyltrehalose trehalohydrolase from the hyperthermophilic archaeum *Sulfolobus solfataricus*. *J. Mol. Biol.* **301**: 451–464.
- FISCHER, D. 2000. Hybrid fold recognition: combining sequence derived properties with evolutionary information, pp. 119–130. In: ALTMAN, R. B., LAUDERDALE, K., DUNKER, A. K., HUNTER, L. & KLEIN, T. E. (eds) *Biocomputing 2000*. Proceedings of the Pacific Symposium, World Scientific Publishing Co., Hawaii.
- FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., et al. & VENTER, J. C. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- HILDEN, I., LO LEGGIO, L., LARSEN, S. & POULSEN, P. 2000. Characterization and crystallization of an active N-terminally truncated form of the *Escherichia coli* glycogen branching enzyme. *Eur. J. Biochem.* **267**: 2150–2155.
- HOMEROVA, D. & KORMANEC, J. 1994. Cloning of the putative glucogen branching enzyme gene, glgB, from *Streptomyces aureofaciens*. *Biochim. Biophys. Acta* **1200**: 334–336.
- HOOFT, R. W. W., VRIEND, G. SANDER, C. & ABOLA, E. E. 1996. Errors in protein structure. *Nature* **381**: 272.
- JESPERSEN, H. M., MACGREGOR, E. A., HENRISSAT, B., SIERKS, M. R. & SVENSSON, B. 1993. Starch- and glycogen-debranching and branching enzymes: prediction of structural features of the catalytic (β/α)₈-barrel domain and evolutionary relationship to other amylolytic enzymes. *J. Protein Chem.* **12**: 791–805.
- JESPERSEN, H. M., MACGREGOR, E. A., SIERKS, M. R. & SVENSSON, B. 1991. Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem. J.* **280**: 51–55.
- JOBLING, S. A., SCHWALL, G. P., WESTCOTT, R. J., SIDEBOTTOM, C. M., DEBET, M., GIDLEY, M. J., JEFFCOAT, R. & SAFFORD, R. 1999. A minor form of starch branching enzyme in potato (*Solanum tuberosum* L.) tubers has a major effect on starch structure: cloning and characterisation of multiple forms of SBE A. *Plant J.* **18**: 163–171.
- JONES, D. T. 1999. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**: 797–815.
- JONES, T. A., ZOU, J. Y., COWAN, S. W. & KJELDGAARD, M. 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**: 110–119.
- KANEKO, T., TANAKA, A., SATO, S., KOTANI, H., SAZUKA, T., MIYAJIMA, N., SUGURA, M. & TABATA, S. 1995. Sequence analysis of the genome of the unicellular cyano-bacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res.* **2**: 153–166.
- KAMITORI, S., KONDO, S., OKUYAMA, K., YOKOTA, T., SHIMURA, Y., TONOZUKA, T. & SAKANO, Y. 1999. Crystal structure of *Thermoactinomyces vulgaris* R-47 α -amylase II (TVaII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. *J. Mol. Biol.* **287**: 907–921.
- KATSUYA, Y., MEZAKI, Y., KUBOTA, M. & MATSUURA, Y. 1998. Three-dimensional structure of *Pseudomonas* isoamylase at 2.2 Å resolution. *J. Mol. Biol.* **281**: 885–897.
- KELLEY, L. A., MACCALLUM, R. M. & STERNBERG, M. J. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**: 499–520.
- KIEL, J. A., BOELS, J. M., BELDMAN, G. & VENEMA, G. 1990. Nucleotide sequence of the *Synechococcus* sp. PCC7942 branching enzyme gene (glgB): expression in *Bacillus subtilis*. *Gene* **89**: 77–84.
- KIEL, J. A., BOELS, J. M., BELDMAN, G. & VENEMA, G. 1991. Molecular cloning and nucleotide sequence of the glycogen branching enzyme gene (glgB) from *Bacillus stearothermophilus* and expression in *Escherichia coli* and *Bacillus subtilis*. *Mol. Gen. Genet.* **230**: 136–144.
- KIEL, J. A., BOELS, J. M., BELDMAN, G. & VENEMA, G. 1992. The glgB gene from the thermophile *Bacillus caldolyticus* encodes a thermolabile branching enzyme. *DNA Seq.* **3**: 221–232.
- KIEL, J. A., BOELS, J. M., BELDMAN, G. & VENEMA, G. 1994. Glycogen in *Bacillus subtilis*: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol. Microbiol.* **11**: 203–218.
- KIM, J. S., CHA, S. S., KIM, H. J., KIM, T. J., HA, N. C., OH, S. T., CHO, H. S., CHO, M. J., KIM, M. J., LEE, H. S., KIM, J. W., CHOI, K. Y., PARK, K. H. & OH, B. H. 1999. Crystal structure of a maltogenic amylase provides insights into a catalytic versatility. *J. Biol. Chem.* **274**: 26279–26286.
- KRAULIS, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**: 946–950.
- LASKOWSKI, R. A., MOSS, D. S. & THORNTON, J. M. 1993. PROCHECK – a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**: 283–291.
- MACGREGOR, E. A., JANECEK, S. & SVENSSON, B. 2001. Relationship of sequence and structure to specificity in the α -amylase family of enzymes. *Biochim. Biophys. Acta* **1546**: 1–20.
- MAY, B. J., ZHANG, Q., LI, L. L., PAUSTIAN, M. L., WHITTAM, T. S. & KAPUR, V. 2001. Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. Natl. Acad. Sci. USA* **98**: 3460–3465.

- LIN, X., KAUL, S., ROUNSLEY, S., SHEA, T. P., BENITO, M. I., TOWN, C. D. et al. & VENTER, J. C. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**: 761–768.
- LLUISMA, A. O. & RAGAN, M. A. 1998. Cloning and characterization of a nuclear gene encoding a starch branching enzyme from the marine red alga *Gracilaria gracilis*. *Curr. Genet.* **34**: 105–111.
- MAY, B. J., ZHANG, Q., LI, L. L., PAUSTIAN, M. L., WHITTAM, T. S. & KAPUR, V. 2001. Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. Natl. Acad. Sci. USA* **98**: 3460–3465.
- NICHOLLS, A., SHARP, K. A. & HONIG, B. 1991. Protein folding and association – insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281–296.
- PREISS, J., BALLICORA, M. A. & BINDERUP, K. 2001. Designing ADP-glucose pyrophosphorylase and branching enzyme to alter starch levels and structure. 27th FEBS Meeting, Lisbon, Portugal, Abstract number 20174.
- RUMBAK, E., RAWLINGS, D. E., LINDSEY, G. G. & WOODS, D. R. 1991. Characterization of the *Butyrivibrio fibrisolvens* glgB gene, which encodes a glycogen-branching enzyme with starch-clearing activity. *J. Bacteriol.* **173**: 6732–6741.
- SKOV, L. K., MIRZA, O., HENRIKSEN, A., DE MONTALK, G. P., REMAUD-SIMEON, M., SARCABAL, P., WILLEMOT, R. M., MONSAN, P. & GAJHEDE, M. 2001. Amylosucrase, a glucan-synthesizing enzyme from the α -amylase family. *J. Biol. Chem.* **276**: 25273–25278.
- THON, V. J., VIGNERON-LESENS, C., MARIANNEPEPIN, T., MONTREUIL, J., DECO, A., RACHEZ, C., BALL, S. G. & CANNON, J. F. 1992. Coordinate regulation of glycogen metabolism in the yeast *Saccharomyces cerevisiae*. Induction of glycogen branching enzyme. *J. Biol. Chem.* **267**: 15224–15228.
- THON, V. J., KHALIL, M. & CANNON, J. F. 1993. Isolation of human glycogen branching enzyme cDNAs by screening complementation in yeast. *J. Biol. Chem.* **268**: 7509–7513.
- UGALDE, J. E., LEPEK, V., UTTARO, A., ESTRELLA, J., IGLESIAS, A. & UGALDE, R. A. 1998. Gene organization and transcription analysis of the *Agrobacterium tumefaciens* glycogen (glg) operon: two transcripts for the single phosphoglucomutase gene. *J. Bacteriol.* **180**: 6557–6564.
- YOKOTA, T., TONOZUKA, T., KAMITORI, S. & SAKANO, Y. 2001. The deletion of amino-terminal domain in *Thermoactinomyces vulgaris* R-47 α -amylases: effects of domain N on activity, specificity, stability and dimerization. *Biosci. Biotechnol. Biochem.* **65**: 401–408.
- WATANABE, K., HATA, Y., KIZAKI, H., KATSUBE, Y. & SUZUKI, Y. 1997. The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. *J. Mol. Biol.* **269**: 142–153.

Received November 2, 2001
Accepted February 12, 2002