Sequence and structural aspects of functional diversification in class I α-mannosidase evolution

I. King Jordan[1], G. Reid Bishop[2,*] and Daniel S. Gonzalez[3]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA, [2]Department of Chemistry, Millsaps College, Jackson, MS 39210, USA and [3]United States Department of Agriculture, Aquatic Animal Health Research Unit, Auburn, AL 36831, USA

ABSTRACT

Motivation: Class I α-mannosidases comprise a homologous and functionally diverse family of glycoside hydrolases. Phylogenetic analysis based on an amino acid sequence alignment of the catalytic domain of class I α-mannosidases reveals four well-supported phylogenetic groups within this family. These groups include a number of paralogous members generated by gene duplications that occurred as far back as the initial divergence of the crown-group of eukaryotes. Three of the four phylogenetic groups consist of enzymes that have group-specific biochemical specificity and/or sites of activity. An attempt has been made to uncover the role that natural selection played in the sequence and structural divergence between the phylogenetically and functionally distinct Endoplasmic Reticulum (ER) and Golgi apparatus groups.

Results: Comparison of site-specific amino acid variability profiles for the ER and Golgi groups revealed statistically significant evidence for functional diversification at the sequence level and indicated a number of residues that are most likely to have played a role in the functional divergence between the two groups. The majority of these sites appear to contain residues that have been fixed within one organelle-specific group by positive selection. Somewhat surprisingly these selected residues map to the periphery of the α-mannosidase catalytic domain tertiary structure. Changes in these peripherally located residues would not seem to have a gross effect on protein function. Thus diversifying selection between the two groups may have acted in a gradual manner consistent with the Darwinian model of natural selection.

Contact: bishogr@millsaps.edu

INTRODUCTION

The asparagine (*N*)-linked oligosaccharide biosynthetic pathway involves numerous steps and the activity of an array of specific glyosyl transferases and glycosyl hydrolases (Kornfeld and Kornfeld, 1985). Glycosyl hydrolases perform trimming functions that provide essential substrates for glycosylation (Herscovics, 1999). Processing α-mannosidases comprise a diverse set of glycosyl hydrolases. There are three well-defined phylogenetic groups (Gonzalez and Jordan, 2000) of processing α-mannosidases, each with characteristic phyletic distribution and biochemical specificity (Herscovics, 1999; Moremen *et al.*, 1994). Class I α-mannosidases—gylcoside hydrolase family 47 (Henrissat, 1991)—are eukaryote specific and show a relatively narrow phyletic distribution (Gonzalez and Jordan, 2000). Enzymes of this class also posses specific biochemical activity limited to the cleavage of α-1,2 mannose residues (Herscovics, 1999; Lal *et al.*, 1998). The class II and III α-mannosidase groups show more diversity than class I in their phyletic distribution (Gonzalez and Jordan, 2000) and less substrate specificity in their biochemical activity (Beccari *et al.*, 1997; Eades *et al.*, 1998; Hiramoto *et al.*, 1997; Howard *et al.*, 1997; Rivera-Marrero *et al.*, 2001).

The trimming of mannose residues by processing class I α-mannosidases begins in the Endoplasmic Reticulum (ER) and continues in the Golgi apparatus. Class I α-mannosidases include enzymes with both ER and Golgi specific activity. Class I ER α-mannosidases trim an α-1,2 mannose residue from $Man_9GlcNAc_2$ to form a specific isomer of $Man_8GlcNAc_2$ (Gonzalez *et al.*, 1999; Lipari *et al.*, 1995; Tremblay and Herscovics, 1999). Class I Golgi α-mannosidases then process $Man_8GlcNAc_2$ to produce $Man_5GlcNAc_2$ (Lal *et al.*, 1998). Each of these organelle-specific groups shows instances of multiple members within a species. For example, there are at least two specific forms of class I Golgi α-mannosidases (IA and IB) found in mammals (Bause *et al.*, 1992; Herscovics *et al.*, 1994; Lal *et al.*, 1994, 1998). Each form is encoded by a separate gene (Campbell Dyke *et al.*, 1997; Herscovics, 1999; Tremblay *et al.*, 1998), produces unique isomers (Lal *et al.*, 1994; Tabas and Kornfeld, 1979; Tulsiani and Touster, 1988) and has a

---

*To whom correspondence should be addressed.

distinct expression pattern (Herscovics *et al.*, 1994; Lal *et al.*, 1994; Tremblay *et al.*, 1998). Thus, class I processing α-mannosidases exhibit functional diversification both within and between organelle-specific groups.

ER and Golgi specific α-mannosidases fall into two well-supported and distinct phylogenetic groups within class I (Gonzalez and Jordan, 2000). The phylogenetic relationship and species distribution among these groups indicates that they were generated by a gene duplication event that likely preceded the divergence of metazoans. Gene duplication has long been recognized as a significant force in the evolution of novel biochemical functions. It has been postulated that, due to the conservative nature of natural selection, gene duplication is essential for the functional diversification of genes (Kimura and Ohta, 1974; Ohno, 1970). While gene duplication may not necessarily be a pre-requisite for the evolution of novel function (Hughes, 1999) it is probably the most important mechanism for generating new gene functions (Li, 1997).

The number of functionally well-characterized sequences available, and the robust phylogenetic relationships among them, render the class I α-mannosidases an attractive and tractable system for studying the interplay among gene duplication, natural selection and functional diversification. Evaluation of these sequences, together with predicted sequences from several genome projects, in an explicitly evolutionary context may allow for an assessment of the sequence and structural determinants of the diversification between functionally distinct groups. An attempt has been made here to evaluate the ways in which natural selection has shaped protein sequence and structure differences between functionally distinct ER and Golgi class I α-mannosidases. The approach used here relies on the comparison of functional (selective) constraints between phylogenetically and functionally distinct groups of sequences. The functional constraints on a related group of proteins can be represented via the construction of site-specific amino acid variation profiles (Naylor and Gerstein, 2000). As two paralogous (related by gene duplication) groups of proteins diverge from a common ancestor, any change in function between groups will be reflected by a change between the variation profiles of each group. Thus, changes in variation profiles can be used to examine functional divergence between groups of proteins (Gu, 1999).

Amino acid sequence variation profiles were determined for the ER and Golgi groups of class I α-mannosidases. These profiles were compared using the method of Gu (1999) to assess the effects of natural selection and functional divergence at the sequence level and to predict specific amino acid residues that are likely to have played a role in the functional diversification between the two organelle-specific groups. These selected residues were then mapped on to the x-ray structure of the catalytic domains of *Saccharomyces cerevisiae* and *Homo sapiens* ER class I α-mannosidases to evaluate their relative positions within the structure of the enzyme.

SYSTEMS AND METHODS

Sequence analysis

Processing α-mannosidase amino acid sequences were retrieved as described previously (Gonzalez and Jordan, 2000). Additional sequences were retrieved from GenBank (Benson *et al.*, 2000) release 116 using the BLAST program (Altschul *et al.*, 1997). For each sequence, the identity of the catalytic domain that characterizes class I α-mannosidases—Pfam (Bateman *et al.*, 2000) glycosyl hydrolase family 47 (Henrissat, 1991)—was determined using the SMART server (Schultz *et al.*, 2000). The mannosidase catalytic domains were aligned using the Clustal X program (Thompson *et al.*, 1997). Pairwise distances were determined for the mannosidase catalytic domain alignment using the PROTDIST program, with the PAM distance matrix, from the PHYLIP package (Felsenstein, 1996). The pairwise distances were used with the neighbor-joining algorithm (Saitou and Nei, 1987) implemented in the NEIGHBOR program of PHYLIP to reconstruct the class I mannosidase phylogeny. One hundred bootstrap alignments were generated using the SEQBOOT program from PHYLIP and each alignment was analyzed as described above. A consensus bootstrap tree was generated from the resulting trees using the CONSENSE program from PHYLIP.

The GZ97 and GZf2 programs (Gu, 1999) were used to test for evidence of functional divergence between Golgi and ER class I α-mannosidase sequences. Golgi and ER sequences were re-aligned using Clustal X, gap-stripped, and separated into group specific alignments for analysis. For each alignment, the GZ97 program was used to determine the expected number of amino acid substitutions at each site. The resulting site-specific variation profiles were compared using the GZf2 program to generate estimates of functional divergence (θ) and the posterior probability (p) that each site is likely to play a role in functional divergence between groups. The posterior probability ratio (pr) was calculated as follows: $pr = p/(1 - p)$.

Structure analysis

The structures of ER class I α-mannosidases from *S.cerevisiae* (PDB entry 1DL2) and *H.sapiens* (PDB entry 1FMI) were visualized using the program Visual Molecular Dynamics (VMD) v1.5a2 (Humphrey *et al.*, 1996). Conserved and selected sites were mapped onto the homologous residues in this structure by comparison with the ER–Golgi group multiple sequence alignment described in the previous section. Distances between conserved and selected sites and the α-carbon of T525 were

calculated using the VMD program. Solvent accessible surfaces (Thompson and Goldstein, 1996) were generated using the MOLMOL program (Koradi *et al.*, 1996) assuming a van der Waals radius of 1.4 Å for the solvent. Electrostatic potentials (Nicholls and Honig, 1990) were painted on to the solvent accessible surface and were calculated assuming a solvent radius of 1.4 Å of normal dielectric constant, a salt concentration of 150 mM with an average salt radius of 2.0 Å. The structure was isolated within a solvent box that is 15 Å larger on each axis than the molecule using Debye–Hueckel screening.

IMPLEMENTATION AND DISCUSSION

Mannosidase phylogenetic and functional relationships

Class I processing α-mannosidase protein sequences (Table 1) were retrieved from GenBank and aligned as described in Section **Systems and methods**. A phylogenetic reconstruction based on this sequence alignment shows four distinct and well-supported groups within class I (Figure 1). The presence of a phyletically diverse set of paralogous members, including plant, fungal and animal representatives, indicates that the gene duplications that helped form the family occurred as early as the initial divergence of the crown-group of eukaryotes. Three of the four class I phylogenetic groups are comprised of enzymes with unique functions and localizations. For instance, there is a fungal group made up of entirely secreted proteins. Also, all of the enzymes shown to have Golgi specific activity fall into one phylogenetic group as do all of the enzymes with ER specific activity. In addition to the functionally characterized members of these organelle-specific groups, there are a number of members that have been uncovered by genome projects. It is predicted that these proteins will show biochemical activities consistent with their phylogenetic affinities (Gonzalez and Jordan, 2000). Finally, there is an entire group of class I sequences of which none have been functionally characterized. This group contains the most phyletically diverse assemblage of sequences in class I and seems likely to represent a unique and essential (selectively conserved) mannosidase function. There is evidence that class I α-mannosidases are included in quality control of glycoproteins via the targeting of proteins with trimmed gycosyl groups towards degradation (Ayalon-Soffer *et al.*, 1999; Cabral *et al.*, 2000; Chillaron *et al.*, 2000). Perhaps this as yet functionally uncharacterized and relatively ancient class I group plays some role in this process. Results reported while this paper was in press demonstrate that a member of the functionally uncharacterized group does, in fact, play a role in protein degradation (Hosokawa *et al.* 2001).
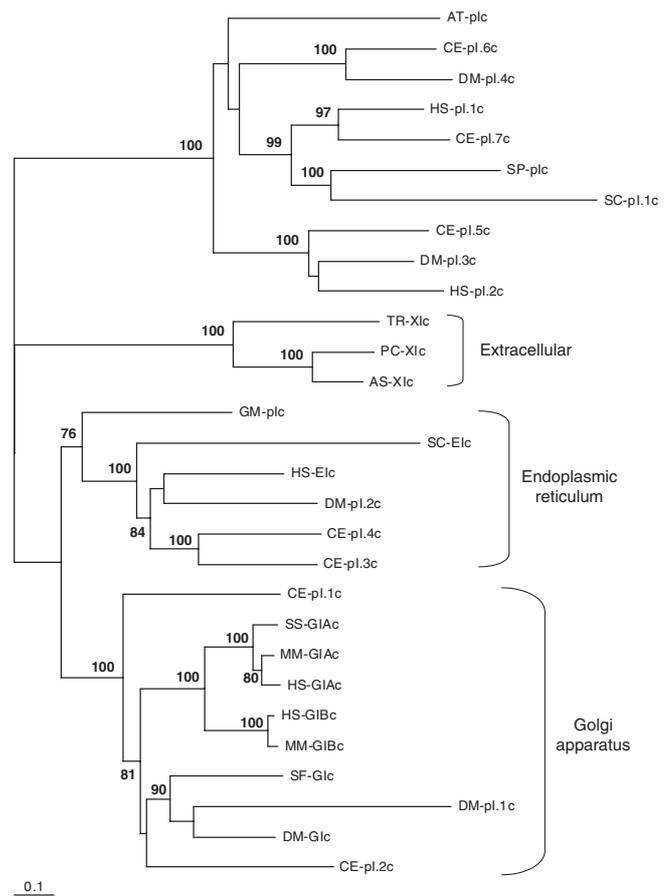
Fig. 1. Phylogeny of class I α-mannosidases. The phylogeny was reconstructed from an amino acid alignment of the class I α-mannosidase catalytic domains as described in Section **Systems and methods**. The taxa (Table 1) nomenclature scheme is as described previously (Gonzalez and Jordan, 2000). Bootstrap values (100 replicates) are shown adjacent to the nodes that they support. The phylogeny is arbitrarily rooted with the functionally uncharacterized (see text) group.

Gene duplication, selection and functional diversification

The phylogenetic structure and taxonomic distribution of class I α-mannosidases show unequivocal evidence of gene duplication followed by functional diversification. As such class I α-mannosidases exemplify an archetypal mechanism for the evolution of novel gene functions (Ohno, 1970). The evolution of new functions subsequent to gene duplication is thought to involve a relaxation of functional constraints (Ohno, 1973) or an acceleration of amino acid substitution due to positive (diversifying) selection (Zhang *et al.*, 1998). In either case, the resulting functional diversification between groups should be manifest in a change in the nature of the functional (selective)

Table 1. Class I α-mannosidase protein sequences analyzed here

Gi[a]	Name[b]	Medline[c]	Description
987 684	PC-XI	95 367 592	*Penicillium citrinum* Extracellular
1171 477	AS-XI	96 106 423	*Aspergillus saitoi (phoenicis)* Extracellular
6979 644	TR-XI	20 146 865	*Trichoderma reesei (Hypocrea jecorina)*
6754 620	MM-GIB	94 193 678 97 288 511	*Mus musculus* Golgi IB
3127 047	HS-GIB	98 256 187	*H.sapiens* Golgi IB
6678 788	MM-GIA	94 193 679	*M.musculus* Golgi IA
5174 521	HS-GIA	94 039 087	*H.sapiens* Golgi IA
2154 997	SS-GIA	97 363 201	*Sus scrofa* Golgi IA
840 752	DM-GI	95 246 933	*Drosophila melanogaster* Golgi
2245 570	SF-GI	97 292 542	*Spodoptera frugiperda* Golgi
7301 866	DM-pI.1	20 196 006	*D.melanogaster* predicted from genome project
3875 192	CE-pI.1	99 069 613	*Caenorhabditis elegans* predicted from genome project
3875 394	CE-pI.2	99 069 613	*C.elegans* predicted from genome project
3881 381	CE-pI.3	99 069 613	*C.elegans* predicted from genome project
1086 860	CE-pI.4	99 069 613	*C.elegans* predicted from genome project
7301 742	DM-pI.2	95 246 933	*D.melanogaster* predicted from genome project
5579 331	HS-EI	99 340 081	*H.sapiens* ER
417 305	SC-EI	91 332 031 20 141 183	*S.cerevisiae* ER
6552 504	GM-pI	20 063 322	*Glycine max* Putative
2832 777	DM-pI.3	95 246 933	*D.melanogaster* predicted from genome project
7023 026	HS-pI.2	Na	*H.sapiens* from Helix Research Institute cDNA project
3875 740	CE-pI.5	99 069 613	*C.elegans* predicted from genome project
7298 014	DM-pI.4	95 246 933	*D.melanogaster* predicted from genome project
3881 448	CE-pI.6	99 069 613	*C.elegans* predicted from genome project
5668 763	AT-pI	Na	*Arabadopsis thaliana* predicted from genome project
3875 111	CE-pI.7	99 069 613	*C.elegans* predicted from genome project
1504 008	HS-pI.1	97 191 544	*H.sapiens* predicted from cDNA
5777 718	SP-pI	Na	*Schizosaccharomyces pombe* predicted from genome project
458 945	SC-pI.1	94 378 003	*S.cerevisiae* predicted from genome project

[a]GenBank unique sequence identifier number.
[b]Nomenclature scheme is the same as described previously (Gonzalez and Jordan, 2000).
[c]Medline identifier number representing the reference where the sequence was reported.

Table 2. Results of Gu analysis

	$\theta \pm SE^a$	r_X^b	r_{MAX}^c	z^d	P^e
Golgi versus ER	0.21 ± 0.08	0.42	0.53	2.91	3.6×10^{-3}

[a]Functional divergence parameter (θ) and standard error as described in the text and Figure 2.
[b]Coefficient of correlation for the number of changes between the ER and Golgi groups.
[c]The expected value of r_X when the evolutionary rate between groups is completely correlated.
[d]Value of the test statistic determined when evaluating the null hypothesis (H_O) that $\theta = 0$ (H_O is equivalent to $r_X = r_{MAX}$).
[e]Probability that $\theta > 0$ due to chance.

constraint on sequences between the different groups. This prediction is the basis of a test that evaluates the nature of functional diversification at the amino acid level (Gu, 1999).

This test was applied to the two organelle-specific groups of class I α-mannosidases: ER and Golgi. The phylogenetic relationships within and between these two groups are shown together with the mathematical essence (Gu, 1999) of the test used (Figure 2). Levels of amino acid variation within and between groups were surveyed in order to determine a measure of functional divergence (θ) between groups. Amino acid variation for each of the two groups was represented by a variability profile where the expected amino acid variation was determined for each site (data not shown). These site-specific profiles can be considered to represent group specific rates of evolution (ν_E and ν_G in Figure 2). The formula to determine the coefficient of rate correlation (r) is shown (Figure 2). If the two groups have experienced no functional divergence, then the groups will not have different functional (selective) constraints and the expected levels of amino acid variation should be the same or proportional (across sites) between groups. If this is the case then r will be equal to 1. Functional divergence between groups on the other hand will reduce r ($0 \leqslant r \leqslant 1$). A measure of functional divergence therefore is $1 - r$ or θ. The higher the value of θ, the greater the degree of functional divergence revealed by the sequences. The value of θ for the ER versus Golgi group comparison (Table 2) is statistically significant, which indicates, not surprisingly when the functional identities of the proteins are considered, that the amino acid sequences show evidence of functional diversification.

Perhaps more interestingly, the method of Gu can be employed to identify which amino acid residues are most likely to have contributed to the functional divergence between the two groups. For each site, a posterior probability ratio (see Section **Systems and methods**) can be calculated (Figure 3). This per-site
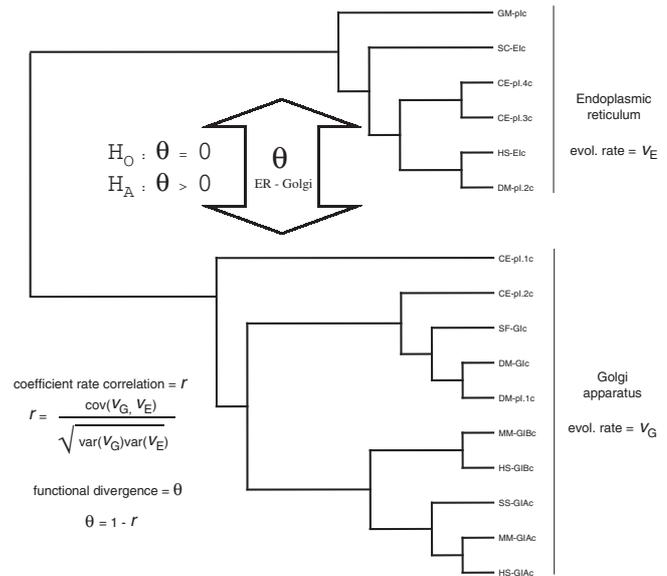
Fig. 2. Scheme for assessing the functional divergence between the ER and Golgi groups of class I α-mannosidases. For each of the two functionally and phylogenetically distinct groups, the rate of evolution (ν_E or ν_G) is represented by a site-specific profile of expected amino acid variation. The functional divergence parameter (θ) is determined from the group specific rates of evolution as shown. The null hypothesis (H_O) of no functional divergence between groups is tested against the alternative (H_A) of functional divergence significantly greater than zero.

ratio is based on the difference in the expected levels of variation between groups. Sites that have the most pronounced differences between groups in the expected level of amino acid variation show the most compelling evidence for altered functional constraints. A highly conserved (low variability) site in one group suggests that it is selectively conserved for some functional utility. If the same site in the other group shows a high level of variability, then it is most likely that this site has experienced a change in the function between groups. While the choice of a posterior probability ratio cut-off is somewhat arbitrary (Gu, 1999), a conservative minimum of 2.0 is employed here. There are seven sites that meet this criterion (Figure 3 and Table 3).

Each of these selected sites shows low (or no) expected variation in one group and relatively high variation in the other group (Table 3 and Figure 4). Such a bimodal pattern of between group variation can be explained in two ways. Firstly, it is possible that such a site was functionally important in the ancestral sequence and thus subject to a high degree of selective constraint. Subsequent to duplication, a relaxation of functional constraints could have occurred in one group leading to a high level of variation at that site in that group. Alternatively, it is
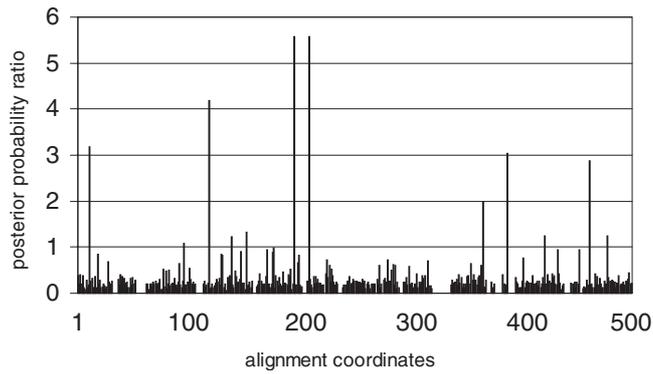
Fig. 3. Posterior probability ratio for the likelihood that an amino acid site played a role in the functional diversification between the ER and Golgi groups. Determined as described in Section **Systems and methods**.

Table 3. Expected per-site variation and posterior probability ratio for selected sites

Alignment[a]	SC-E1[b]	ER–X_1^c	Golgi–X_2^c	R_{12}^d
11	55	0.00	8.04	3.19
119	160	0.00	8.70	4.18
195	235	8.00	0.00	5.59
208	248	8.00	0.00	5.59
366	401	5.87	0.00	2.00
388	423	1.11	11.55	3.04
461	495	6.63	0.00	2.88

[a] Position of the selected residue in the ER–Golgi alignment (Figure 4).
[b] Position of the selected residue in the SC-EI sequence that has been structurally characterized (Vallee *et al.*, 2000b).
[c] Expected level of amino acid variation at selected sites for the ER and Golgi groups respectively.
[d] Posterior probability ratio that a site is likely to have played a role in the functional diversification between the ER and Golgi groups.

possible that such a site was not functionally important in the ancestral sequence and so was not subject to selective constraint. After duplication, the site may have gained a functionally important role in one group, so the critical residue would have been fixed by positive selection in that group. After an initial period of positive selection, the identity of the residue at that functionally important site would be maintained by negative selection in one group.

It is possible to distinguish between these two possibilities (relaxation of functional constraint versus positive selection) by considering the pattern of variation at the selected sites in the two outgroups (extracellular and predicted Figure 1) with respect to the expectations given a parsimonious model of evolution. If the bimodal pattern of variation at a selected site is due to a relaxation of functional constraints, then one expects to observe low

levels of variation at that site in the two outgroups. If, on the other hand, the pattern of variation at a selected site is due to fixation by positive selection, one expects to observe high levels of variation at that site in the two outgroups. These possibilities were evaluated using the multiple sequence alignment (data not shown) of the catalytic domains of all class I α-mannosidases analyzed here (Table 1). The levels of amino acid variation at selected sites were observed in each of the four phylogenetically distinct groups. Five of the seven selected sites (positions 11, 195, 208, 388 and 461 in Table 3 and Figure 4) show a pattern of variation consistent with the fixation of a functionally important residue by positive selection, while the other two sites show variation consistent with a relaxation of functional constraints in one of the two groups. Thus, for the sites deemed most likely to play a role in the functional diversification between the ER and Golgi groups, the predominant mode of evolution after duplication appears to be the fixation of favorable residues by positive selection.

It is worth noting that the presence of such sites, conserved in one phylogenetic group and variable in another, has received substantial treatment as part of the covarion theory of molecular evolution (Fitch and Markowitz, 1970; Miyamoto and Fitch, 1995). This theory holds that sites may change from invariant to variant after a speciation event. The approach developed for identifying such sites as important in functional diversification can be applied to any evolutionary event that generates a bifurcation in a phylogenetic tree (Gu, 1999). This includes gene duplication, as described here, or speciation as in the covarion theory. In this sense, the covarion model can be considered to describe functional divergence after speciation (Gu, 1999).

There is one notable site that was not detected by the Gu method but that may play a role in the functional diversification between the ER and Golgi groups (site 17—yellow in Figure 4). This site has a fixed difference between the two groups with each group having a completely conserved residue of unique identity. The absence of variation at this site in both groups results in a correlation of $r = 1$ between groups and thus no detectable functional divergence. However, such sites have been identified by the evolutionary trace method (Lichtarge *et al.*, 1996) as contributing to functional differences between related groups of proteins. In addition when the pattern of variation at this site is observed for all four class I α-mannosidase groups, it becomes apparent that positive selection is likely to have fixed the difference between the Golgi group and the rest of the groups.

Structural topology of selected residues

The location of these selected sites in the primary (sequence), secondary and tertiary structures of the

```
                    1                                                                                                  100
SC-EI     (45)  MFLESWRDYSKHGWGYDVYGPIEHTSHNM..PRGNQPLGWIIVDSVDTLMLMYNSSTLYKSEFEAEIQRSEHWINDVLDFD.IDAEVNVFETTIRMLGGL
HS-EI    (256)  VFLHAWKGYRKFAWGHDELKPVSRSFSEW..FG....LGLTLIDALDTMWILG........LRKEFEEARKWVSKKLHFE.KDVDVNLFESTIRILGGL
CE-pI.3   (89)  AFQHAWSYKKYAWGHDELRPVSKRFDDS..FG....LGLTIIDSLDTAIIMG........LEEETRDGVEWIRENLNVS.PARSVNVFETTIRVLGGL
GM-pI    (115)  AMLHAWGSYRKYAWGQDELQPQSKNGVNS..FG....GLGATLIDSLDTLYIMG........LNEQFQKAREWVANSLDFN.KDYEASVFETTIRVVGGL
HS-GIA   (174)  MMKHAWNNYKCYAWGKNELKPISKGGHSSSLFG..NIKGATIVDALDTLFIME........MKHEFEEAKSWVEENLDFN.VNAEISVFEVNIRFVGGL
HS-GIB   (187)  MMKHAWDNYRTYGWGHNELRPIARKGHSPNIFGS.SQMGATIVDALDTLYIMG........LHDEFLDGQRWIEDNLDFS.VNSEVSVFEVNIRFIGGL
CE-pI.1  (146)  MMIHAWEGYKNYSWGANELRPMSKKPNSQNIFGG.SQMPATIVDAADTLFIMD........LKDKYKEARDYIENNFSMAKSTSTLSVFETTIRFLGGL
DM-pI.1  (115)  MMMHAWRNYARVVWGYNEFRPISRRVHFGGDFAT.YKLGATIIESLDTLHLMG........LNKELRRSRDWIEKSFHLDRVDEALSVYELTSRLLCPM

                    101                                                                                                200
SC-EI           LSAYHLSDVLEVGNKTVYLNKAIDLGDRLALAFLSTQTGIPYSSINLHSGQAVKNHAD.GGASSTAEFTTLQMEFKYLAYLTGNRTYWELVERVYEPLYK
HS-EI           LSAYHLS......GDSLFLRKAEDFGNRLMPAFR.TPSKIPYSDVNIGTGVAHPPRWTS..DSTVAEVTSIQLEFRELSRLTGDKKFQEAVEKVTQHIHG
CE-pI.3         LSGYHLT......GEEALDEKATQLGDNLIKAFTKSSSLIPKSDVNLDNGNACSPNR...DLSSLAEATTLQLEFRDLTALTGDQKYEDVAFGASEHVHN
GM-pI           LSAYDLS......GDKVFLDKAIEIADRLLPAWN.TPTGIPYNIINLSHGRAHNPSWT.GGESILADSGTEQLEFIVLSQRTGDLKYQQKVENVTAQLNK
HS-GIA          LSAYYLS......GEEIFLRKKAVELGVKLLPAFH.TPSGIPWALLNMKSGIGRNWPASGGSSILAEFGTLHLEFMHLSHLSGNPIFAEKVMNIRTVLNK
HS-GIB          LAAYYLS......QEEIFKIKAVQLAEKLLPAFN.TPTGIPWAMVNLKSGVGRNWGWASAGSSILAEFGTLHMEFIHLSYLTGDLTYYKKVMHIRKLLQK
CE-pI.1         LSLYALT......QESFYIEKAREVGEALLPAFN.TPSGIPKSNLDVASKHASNYGWANGGQSILSEIGSLHLEFLYLSRISNAPIFEKKVKKVRDALEK
DM-pI.1         LTLYSLT......GDSLYMDKAIHIADKILPAFD.TPTGIPRRLVVPKEGSTLTKYL.S.DISRTSEFGSLHLEFYYLSEVSGYPVYRERVDAIREILAK

                    201                                                                                                300
SC-EI           NNDLLNTYMDGLVPIYTFPDTGKFGA.STIRFGSRGDSFYEYLLKQYLLTHET....LYYDLYRKSMEGMKKHLLAQSKPSSLWYIGEREQGLHGQLSPKM
HS-EI           LSG...KKDGLVPMFINTHSGLFTHLGVFTLGARADSYYEYLLKQWIQGGKQE..TQLLEDYVEAIEGVRTHLLRHSEPSKLTFVGELAHGR...FSAKM
CE-pI.3         VGC...KKKLGCPFYIDG.KGEFKK.SSITLGARADSYYEYLIKQWLQTKKS..IDWLRDDFIQSIAAMKKHLYRQSQPNSIWFLGEITELAQ..FYPKM
GM-pI           TF....PDGLLPIYINPHSG.AAGYSPITFGAMGDSFYEYLLKVWIQGNKTSSIKHYRDMWEKSMKGLSS.LIRRSTPSSFTYICEKNGGS...LTDKM
HS-GIA          LE....KKQGLYPNYLNPSSG.QWGQHHVSVGGLGDSFYEYLLKAWLMSDKTD..LEAKKMYFDAVQAIETHLIRKS.SSGLTYIAEWKRGL...LEHKM
HS-GIB          MD....RKNGLYPNYLNPRTG.RWGQYHTSVGGLGDSFYEYLLKAWLMSDKTD..HEARKMYDDAIEAIEKHLIKKS.RGGLTFIGEWKNGH...LEKKM
CE-pI.1         AE....KKNGLYSNYINPDTG.KFTGSHMSLGALGDSFYEYLIKSYVQSNYTD..TQAKNMYWDVSDAIQKHMIKVSKQSNLTYTVELNNGQ...AQHKM
DM-pI.1         TT....RKNGLYPNAYCTKFG.KWENYNCSMHRL....MDTLLKSWIQSGRTD..TQNADTFKEAMLAVAQNLVVIN.PEDVTYVSTFRNGT...LFHRM

                    301                                                                                                400
SC-EI           DHLVCFMGGLLASGSTEGLSIHEARRRPFFSLSLERKSDWDLAKGITDTCYQMYKQSSSGLAPEIVVENDGNIKQDGWWRSSVGDFFVKPLDRHNLQRPE
HS-EI           DHLVCFLPCTLALGVYHGLPA................SHMELAQELMETCYQMNRQMETGLSPEIVHEN.LYPQPG..RR...DVEVKPADRHNLLRPE
CE-pI.3         DHLVCFLSGSLVLSHLNGLDHDN.............EHLEMAKNIGNVCHKMY.ENPTGLGPEIIHEN..MEDSSDMTQA...DTYVKNLDAHSLLRPE
GM-pI           DELACFAPCMIALGSFGYSAADD............SQKFLSLAEELAWTCYNFYQSTPTKLAGENYFGH....SGQD.......MSVC...TSWNILRPE
HS-GIA          GHLTCFAGGMFALGADAAPEGM.............AQHYLELGAEIARTCHESYNRTFMKLGPEAFRFD....GGVE......AIAIRQNEKYYILRPE
HS-GIB          GHLACFAGGMFALGADGSRADK.............AGHYLELGAEIARTCHESYDRTALKLGPESFKFD....GAVE......AVAVRQAEKYYILRPE
CE-pI.1         GHLACFVPGMFALQAINEDTEEE............KLRIMTLAEELAKTCHESYIRSETHIGPEMFYEN....ERDE......ATSRHSENG.YIQRPE
DM-pI.1         RHSDCFAGGLFVLGAAETQMKH............WEKYAHIGIGLTDTCHDSYWSSPTRLGPDTFAFT....EESQ......QEIEPLQRNYYNLRPE

                    401                                                                              501
SC-EI           TVEESIMFMYHLSHDHKYREWGAEIATSFFENTCVDCNDPKLRRFTSLSDCIT.LPTKKSNNMBSFWLAETLKYLYILFLDE...FDFTKVVFNTEAHPFPV (532)
HS-EI           TVEESLFYLYRVTGDRKVQDWGWEILQSFSRFTRVPSGG.....YSSINNVQDPQKPEPRDKMBSFFLGETLKYLFLLFSDDPNLLSLDAYVFNTEAHPLPI (653)
CE-pI.3         AIEAWFYLYRVTKDKKVQDWGWKAFESIEKYAKVETGG.....YSSIDNVLR.KIKRRDKMBSFFLAETLKYLLYLMADDQEILPLPDRWVLNTEAHPLPI (530)
GM-pI           TVEESLFYLWRLTGNKTVQEWGWNIFQAFEKNSRIESG.....YVGLKDVN...SGVKDNMMQSFFLAETLKYFYLLFSPSS.VISLDEWVFNTEAHPLRI (547)
HS-GIA          VMETYMYMWRLTHDPKYRKWAWEAVEALENHCRVNGG.....YSGLRDVYL.LHESYDDVQQSFFLAETLKYLYLIFSDDD.LLPFEHWIFNSEAHLLPI (612)
HS-GIB          VIETTWYLWRFTHDPRYRQWGWEAAALAIEKYCRVNGG.....FSGVKDVYS.STPTHDDVQQSFFLAETLKYLYLLFSGDD.LLPFDHWVFNTEAHPLPV (626)
CE-pI.1         VIEGWFYLWRLTGKTMKRDWVVWDAVQAIEKYCRVDSG.....FTGLQNVYN.PKAGREDVMQSFFLAEFLKYAYLTFADES.LISLDKWVFNTEAHPVPV (587)
DM-pI.1         VAEETYLVLWRITHHPQYRLWGLEMVQAIEKYCRMPYG.....YTGVMDVNN.VTSEPDDYQGSFFLGSTLKYLYLLFSDDD.VVSLEQWVFNSAGHFLPI (547)
```

Fig. 4. Multiple sequence alignment of the catalytic domain of eight representative (four—ER and four—Golgi) class I α-mannosidases. Sites absolutely conserved in all Golgi and ER mannosidase sequences (including those not shown) are shown in blue. Selected sites deemed likely to have played a role in the functional diversification between groups are shown in red. The site indicated in yellow, while not detected by the Gu method, may also play a role in the functional diversification between groups (see text).

α-mannosidase catalytic domain was evaluated and compared with the location of the absolutely conserved sites. The selected sites are mapped onto an alignment of eight representative, four—ER and four—Golgi, class I α-mannosidases catalytic domains (red—Figure 4). Also shown in the alignment are the sites that are absolutely conserved within and between both groups (blue—Figure 4). These absolutely conserved sites are likely to be the most functionally important sites in the catalytic domain. For the most part the selected sites and the conserved sites appear to be associated in the primary structure. However, the spatial relationship between these two classes of residues can be more realistically evaluated by mapping them (see Section **Systems and methods**) onto the secondary and tertiary structures of the enzyme. Secondary structural element identities (helix, sheet and loop) were determined for the conserved and selected sites (data not shown). Comparison of these identities with the overall secondary structure composition of the

S.cerevisiae ER class I α-mannosidase (SC-EI) revealed no significant deviation ($\chi^2 = 8.64$, $P = 0.13$) from the expected values. Thus conserved and selected sites do not appear to be over or under represented in any particular seconday structural environment of the catalytic domain.

A ribbon view of the crystal structure (PDB entry 1DL2) of the catalytic domain of SC-EI reveals that absolutely conserved residues (blue) tend to be located on the inside barrel of the structure close to the active site (Figure 5). Surprisingly, the selected residues (red) appear to be predominantly located on the periphery of the structure. A different view of the same structure, that shows only conserved and selected sites, demonstrates even more clearly that selected sites are located on the periphery of the structure (Figure 6). To assess the robustness of these visual observations, the distances from each conserved and selected residue to the α-carbon of a centrally located residue (T525—Figure 6) were calculated (Table 4). Indeed, selected residues are on
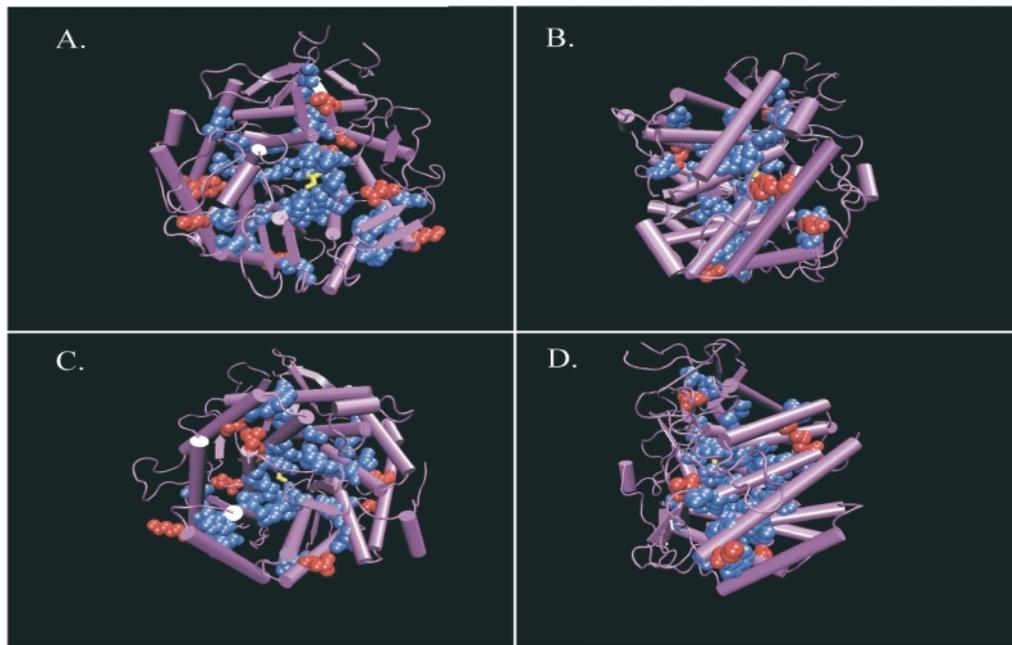
Fig. 5. Crystal structure (PDB entry 1DL2) of the catalytic domain of *S.cerevisiae* ER class I α-mannosidase (SC-EI). Panels (B), (C), and (D) are respective 90°, 180°, 270° rotations about the *y*-axis of the structure presented in panel (A). The protein backbone is drawn as a ribbon structure. Homology modeling (see Section **System and methods**) was used to map absolutely conserved sites (blue) and selected sites likely to have played a role in the functional diversification between groups (red) onto the structure. The atoms of these residue's side chains are presented as van der Waals radii. A molecule of glycerol marking the enzyme active site is depicted in yellow.
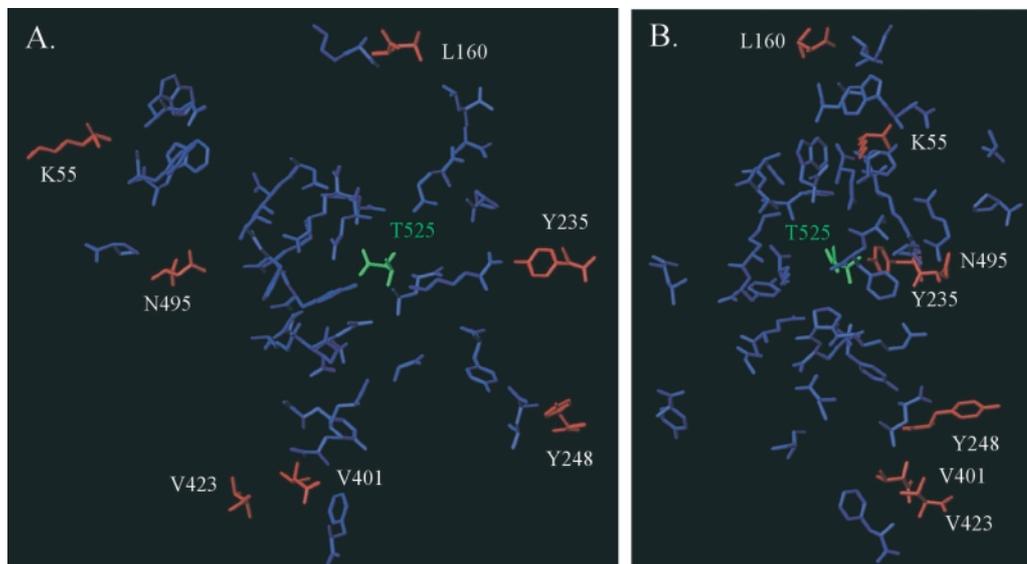
Fig. 6. A neon view (only bonds shown) of the crystal structure (PDB entry 1DL2) of the catalytic domain of *S.cerevisiae* ER class I α-mannosidase (SC-EI). Residues that occupy absolutely conserved sites (blue) and residues that occupy selected sites likely to have played a role in functional diversification between groups (red) are shown to emphasize their structural relationships. The residue in shown green is T525, which is buried near the center of the monomeric structure and serves as a point of reference for distance calculations (Table 4). The orientation is a 90° rotation from Figure 5a.
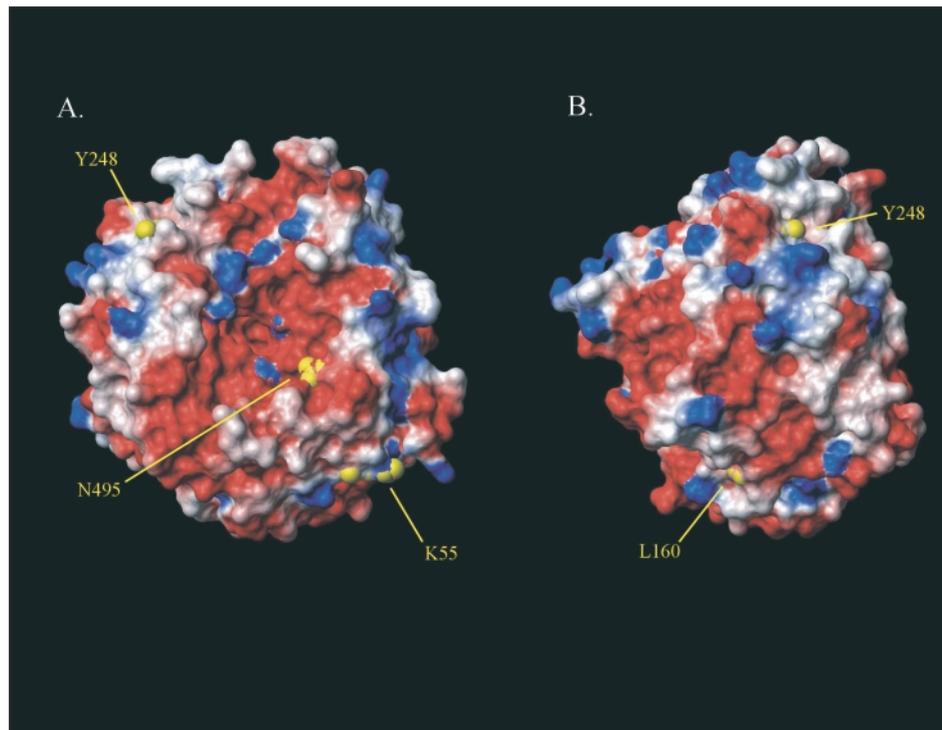
Fig. 7. Electrostatic potential (Nicholls and Honig, 1990) and solvent accessible surface (Thompson and Goldstein, 1996) of the *S.cerevisiae* ER class I α-mannosidase (SC-EI) catalytic domain structure (PDB entry 1DL2), panel (B) is a 90° rotation about the *y*-axis with respect to panel (A). The solvent accessible surface was generated using a 1.4 Å spherical radius for water using the program MOLMOL (Koradi *et al.*, 1996). The electrostatic potential energy surface was then calculated and painted onto the solvent accessible surface. The patches colored white, red, and blue represent regions that are overall neutrally, negatively, and positively charged respectively. The atoms depicted as yellow spheres show those residues that occupy selected sites that have played a role in the functional diversification between groups and are solvent accessible and, thus, on the surface of the protein.

average located further from the center of the structure than are conserved residues (*t*-test $P = 1.14 \times 10^{-4}$). In addition, of the seven selected sites implicated in α-mannosidase functional diversification, three are occupied by residues that are solvent accessible as determined by a mapping of the solvent accessible surface of the SC-EI structure (Figure 7). Thus, these solvent accessible selected residues may be in direct contact with the environment of the organelle where the protein is localized (Golgi or ER) or with other units of a multimeric protein complex.

Given the likelihood that the selected sites play a role in the functional diversification between the two groups, it could have been expected that they would map to the inside of the structure close to the ligand. It is easy to imagine how changes in such internally located sites could affect protein function. In fact, it has been shown that a change in the identity of a single internally located residue is sufficient to change the biochemical specificity of the *S.cerevisiae* ER α-mannosidase (SC-EI) to that of a Golgi

group α-mannosidase (Romero *et al.*, 2000). It is less obvious how changes in sites located on the periphery and near the surface of the structure may affect the biochemical function of the enzymes. None of these sites have been directly implicated in the catalytic mechanism of the enzyme nor do they play any obvious role in the stabilization of the structure of the protein. However, changes in such sites may in fact have a subtle affect on protein function while changes on the inside of the structure would be more prone to result in loss of function and thus be eliminated by purifying selection. Consistent with this idea, diversifying selection has been shown to favor relatively conservative amino acid substitutions (Wyckoff *et al.*, 2000). This type of positive selection for small subtle changes is also consistent with the Darwinian (Darwin, 1859) notion of evolution by a series of gradual changes. It has also been shown that such sites on the surface of proteins are responsible for mediating changes in protein–protein interactions (Sowa *et al.*, 2000). Thus adaptively fixed changes in peripherally located sites may

Table 4. Distances (Å) of selected and conserved sites to the centrally located threonine (T525) residue in the *S.cerevisiae* class I ER α-mannosidase (SC-EI) structure (PDB entry 1DL2)

Residue #[a]	Distance from T525[b]
Selected residues	
55	27.10
61	19.90
160	19.83
235	21.19
248	24.69
401	21.32
423	26.05
495	19.13
Conserved residues	
50	20.29
53	22.62
58	21.96
59	21.51
65	24.82
89	11.77
90	13.26
132	9.88
136	7.69
162	19.48
173	20.27
182	19.43
214	10.55
218	16.29
251	20.24
278	11.37
283	11.62
340	11.61
344	12.13
385	18.23
403	25.17
433	11.17
434	10.58
435	7.97
438	6.54
452	20.65
498	11.76
499	9.76
506	11.66
507	12.11
518	15.61
528	7.36

[a]Residues numbered with respect to the position of the selected residue in the SC-EI sequence that has been structurally characterized (Vallee *et al.*, 2000b).

[b]Distances in Å determined as described in the Section **Systems and methods**.

have driven functional diversification between ER and Golgi α-mannosidases by facilitating interactions between members of macromolecular protein assemblies that are endemic to different organellar environments.

With respect to this assertion, it is of note that each of the solvent accessible selected residues contributes to the electrostatic potential of the surface of the protein (Figure 7). For example, K55 imparts a very localized positively charge isolated in a net neutral, albeit polar, region on the surface of the protein (Figure 7). Such a disruption in surface charge could lead to alternative binding modes of the enzyme for various proteins. Furthermore, of the four selected residues that are solvent accessible, only N495 is poised to make a potential hydrogen-bonding contact with substrate. The active form of the α-mannosidase enzyme is thought to be a multimer (Vallee *et al.*, 2000b). In the crystal structure, the mannose unit of the *N*-linked glycosyl group of a neighboring enzyme monomer is inserted into the putative active site 'mouth' of the enzyme. Figure 7a, depicts the opening of the active site which also contains a bound glycerol molecule in the yeast crystal structure. N495 is relatively close (19.13 Å) to the active site. Substitution of this polar amino acid residue with a non-polar valine as is the case for the all the Golgi specific proteins (Figure 4) could alter substrate specificity and/or protein–protein interaction by changing the polar character of that site. Such an effect would be enhanced by the disruption of the very polar and negatively charged (i.e. red) acidic surface immediately surrounding the residue near the active site of the enzyme (Figure 7).

While this manuscript was in preparation the structure for the human class I ER α-mannosidase (PDB entry 1FMI) was published (Vallee *et al.*, 2000a). When this structure was superimposed on the *S.cerevisiae* class I ER α-mannosidase structure, the root mean square deviation for the distance between all homologous C-α atoms was 1.44 Å (Vallee *et al.*, 2000a). This indicates that the human structure is essentially identical to that of *S.cerevisiae*. Accordingly, the same structural patterns as those obtained for the location of the conserved and selected residues in the yeast structure were observed for the structure of the human ER α-mannosidase (data not shown). The identity between the yeast and human ER α-mannosidase structures is not surprising when you consider the high level of sequence identity (38%) and similarity (55%) between the catalytic domains of these two enzymes. Structure is known to be conserved in evolution for far longer than sequence. For example, it has been shown that homologous proteins with sequence identities as low as 25% have remarkably similar structures (Chothia and Lesk, 1986). Any systematic attempt to further map out the structural space occupied by processing α-mannosidases should employ judicious sampling based on the known phylogenetic relationships (Figure 1; Gonzalez and Jordan, 2000) among family members.

CONCLUSION

The use of position specific variability information from alignments of homologous proteins is emerging as a powerful method for extracting meaningful biological

information from sequence data (Altschul *et al.*, 1997; Gu, 1999; Naylor and Gerstein, 2000). The use of such profiles here enabled the identification of sequence and structural elements that have likely been shaped by natural selection during the functional diversification of organelle-specific groups of class I processing α-mannosidases. Somewhat unexpectedly, such selected residues map to the periphery of the mannosidase catalytic domain structure away from the active site. However, the structural location of these selected residues may not be surprising when the dichotomous role of natural selection is considered. For the most part, at the molecular level, natural selection functions as a conservative force eliminating variants that alter the function of a given protein. Such purifying (negative) selection can be expected to be particularly stringent among residues in or near the active site of an enzyme. More rarely, natural selection may favor variants that change the function of a protein. Such variants are not likely to occur among the severely constrained residues that make up an enzyme's active site. Less constrained sites at the periphery of a structure are freer to explore sequence space and thus more likely to effect tolerable changes in molecular function and organismic fitness. This mode of change, as revealed here for class I α-mannosidases, is consistent with both the Darwinian prediction of evolutionary diversification by a series of small gradual steps (Darwin, 1859) and Wright's shifting balance theory (Wright, 1977) that emphasizes the role of drift in traversing fitness valleys.

REFERENCES

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Ayalon-Soffer,M., Shenkman,M. and Lederkremer,G.Z. (1999) Differential role of mannose and glucose trimming in the ER degradation of asialoglycoprotein receptor subunits. *J. Cell Sci.*, **112**, 3309–3318.

Bateman,A., Birney,E., Durbin,R., Eddy,S.R., Howe,K.L. and Sonnhammer,E.L. (2000) The Pfam protein families database. *Nucleic Acids Res.*, **28**, 263–266.

Bause,E., Breuer,W., Schweden,J., Roeser,R. and Geyer,R. (1992) Effect of substrate structure on the activity of Man9-mannosidase from pig liver involved in N-linked oligosaccharide processing. *Eur. J. Biochem.*, **208**, 451–457.

Beccari,T., Appolloni,M.G., Costanzi,E., Stinchi,S., Stirling,J.L., Della Fazia,M.A., Servillo,G., Viola,M.P. and Orlacchio,A. (1997) Lysosomal alpha-mannosidases of mouse tissues: characteristics of the isoenzymes, and cloning and expression of a full-length cDNA. *Biochem. J.*, **327**, 45–49.

Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18.

Cabral,C.M., Choudhury,P., Liu,Y. and Sifers,R.N. (2000) Processing by endoplasmic reticulum mannosidases partitions a secretion-impaired glycoprotein into distinct disposal pathways. *J. Biol. Chem.*, **275**, 25 015–25 022.

Campbell Dyke,N., Athanassiadis,A. and Herscovics,A. (1997) Genomic organization and chromosomal mapping of the murine alpha1,2-mannosidase IB involved in *N*-glycan maturation. *Genomics*, **41**, 155–159.

Chillaron,J., Adan,C. and Haas,I.G. (2000) Mannosidase action, independent of glucose trimming, is essential for proteasome-mediated degradation of unassembled glycosylated Ig light chains. *Biol. Chem.*, **381**, 1155–1164.

Chothia,C. and Lesk,A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823–826.

Darwin,C. (1859) *On the Origin of Species by Means of Natural Selection.* Murray, London.

Eades,C.J., Gilbert,A.M., Goodman,C.D. and Hintz,W.E. (1998) Identification and analysis of a class 2 alpha-mannosidase from *Aspergillus nidulans*. *Glycobiology*, **8**, 17–33.

Felsenstein,J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Meth. Enzymol.*, **266**, 418–427.

Fitch,W.M. and Markowitz,E. (1970) An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem. Genet.*, **4**, 579–593.

Gonzalez,D.S. and Jordan,I.K. (2000) The alpha-mannosidases: phylogeny and adaptive diversification. *Mol. Biol. Evol.*, **17**, 292–300.

Gonzalez,D.S., Karaveg,K., Vandersall-Nairn,A.S., Lal,A. and Moremen,K.W. (1999) Identification, expression, and characterization of a cDNA encoding human endoplasmic reticulum mannosidase I, the enzyme that catalyzes the first mannose trimming step in mammalian Asn-linked oligosaccharide biosynthesis. *J. Biol. Chem.*, **274**, 21 375–21 386.

Gu,X. (1999) Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol.*, **16**, 1664–1674.

Henrissat,B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, **280**, 309–316.

Herscovics,A. (1999) Importance of glycosidases in mammalian glycoprotein biosynthesis. *Biochim. Biophys. Acta*, **1473**, 96–107.

Herscovics,A., Schneikert,J., Athanassiadis,A. and Moremen,K.W. (1994) Isolation of a mouse Golgi mannosidase cDNA, a member of a gene family conserved from yeast to mammals. *J. Biol. Chem.*, **269**, 9864–9871.

Hiramoto,S., Tamba,M., Kiuchi,S., Jin,Y.Z., Bannai,S., Sugita,Y., Dacheux,F., Dacheux,J.L., Yoshida,M. and Okamura,N. (1997) Stage-specific expression of a mouse homologue of the porcine 135kDa alpha-D-mannosidase (MAN2B2) in type A spermatogonia. *Biochem. Biophys. Res. Commun.*, **241**, 439–445.

Hosokawa,N., Wada,I., Hasegawa,K., Yorihuzi,T., Tremblay,L.O., Herscovies,A. and Nagata,K. (2001) A novel ER alpha-mannosidase-like protein accelerates ER-associated degradation. *EMBO Rep.*, **2**, 415–422.

Howard,S., Braun,C., McCarter,J., Moremen,K.W., Liao,Y.F. and Withers,S.G. (1997) Human lysosomal and jack bean alpha-mannosidases are retaining glycosidases. *Biochem. Biophys. Res. Commun.*, **238**, 896–898.

Hughes,A.L. (1999) *Adaptive Evolution of Genes and Genomes.* Oxford University Press, New York.

Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graph.*, **14**, 33–38, 27–28.

Kimura,M. and Ohta,T. (1974) On some principles governing molecular evolution. *Proc. Natl Acad. Sci. USA*, **71**, 2848–2852.

Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.

Kornfeld,R. and Kornfeld,S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, **54**, 631–664.

Lal,A., Pang,P., Kalelkar,S., Romero,P.A., Herscovics,A. and Moremen,K.W. (1998) Substrate specificities of recombinant murine Golgi alpha1,2-mannosidases IA and IB and comparison with endoplasmic reticulum and Golgi processing alpha1,2-mannosidases. *Glycobiology*, **8**, 981–995.

Lal,A., Schutzbach,J.S., Forsee,W.T., Neame,P.J. and Moremen,K.W. (1994) Isolation and expression of murine and rabbit cDNAs encoding an alpha1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides. *J. Biol. Chem.*, **269**, 9872–9881.

Li,W.H. (1997) *Molecular Evolution*. Sinauer, Sunderland, MA.

Lichtarge,O., Bourne,H.R. and Cohen,F.E. (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.*, **257**, 342–358.

Lipari,F., Gour-Salin,B.J. and Herscovics,A. (1995) The *Saccharomyces cerevisiae* processing alpha1,2-mannosidase is an inverting glycosidase. *Biochem. Biophys. Res. Commun.*, **209**, 322–326.

Miyamoto,M.M. and Fitch,W.M. (1995) Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.*, **12**, 503–513.

Moremen,K.W., Trimble,R.B. and Herscovics,A. (1994) Glycosidases of the asparagine-linked oligosaccharide processing pathway. *Glycobiology*, **4**, 113–125.

Naylor,G.J. and Gerstein,M. (2000) Measuring shifts in function and evolutionary opportunity using variability profiles: a case study of the globins. *J. Mol. Evol.*, **51**, 223–233.

Nicholls,A. and Honig,B. (1990) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comput. Chem.*, **12**, 435–445.

Ohno,S. (1970) *Evolution by Gene Duplication*. Springer, New York.

Ohno,S. (1973) Ancient linkage groups and frozen accidents. *Nature*, **244**, 259–262.

Rivera-Marrero,C.A., Ritzenthaler,J.D., Roman,J. and Moremen,K.W. (2001) Molecular cloning and expression of an alpha-mannosidase gene in *Mycobacterium tuberculosis*. *Microb. Pathog.*, **30**, 9–18.

Romero,P.A., Vallee,F., Howell,P.L. and Herscovics,A. (2000)

Mutation of Arg (273) to Leu alters the specificity of the yeast *N*-glycan processing class I alpha1,2-mannosidase. *J. Biol. Chem.*, **275**, 11 071–11 074.

Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.

Schultz,J., Copley,R.R., Doerks,T., Ponting,C.P. and Bork,P. (2000) SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.*, **28**, 231–234.

Sowa,M.E., He,W., Wensel,T.G. and Lichtarge,O. (2000) A regulator of G protein signaling interaction surface linked to effector specificity. *Proc. Natl Acad. Sci. USA*, **97**, 1483–1488.

Tabas,I. and Kornfeld,S. (1979) Purification and characterization of a rat liver Golgi alpha-mannosidase capable of processing asparagine-linked oligosaccharides. *J. Biol. Chem.*, **254**, 11 655–11 663.

Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.

Thompson,M.J. and Goldstein,R.A. (1996) Predicting solvent accessibility: higher accuracy using Bayesian statistics and optimized residue substitution classes. *Proteins*, **25**, 38–47.

Tremblay,L.O. and Herscovics,A. (1999) Cloning and expression of a specific human alpha1,2-mannosidase that trims Man$_9$GlcNAc$_2$ to Man$_8$GlcNAc$_2$ isomer B during *N*-glycan biosynthesis. *Glycobiology*, **9**, 1073–1078.

Tremblay,L.O., Campbell Dyke,N. and Herscovics,A. (1998) Molecular cloning, chromosomal mapping and tissue-specific expression of a novel human alpha1,2-mannosidase gene involved in *N*-glycan maturation. *Glycobiology*, **8**, 585–595.

Tulsiani,D.R. and Touster,O. (1988) The purification and characterization of mannosidase IA from rat liver Golgi membranes. *J. Biol. Chem.*, **263**, 5408–5417.

Vallee,F., Karaveg,K., Herscovics,A., Moremen,K.W. and Howell,P.L. (2000a) Structural basis for catalysis and inhibition of *N*-glycan processing class I alpha1,2-mannosidases. *J. Biol. Chem.*, **275**, 41 287–41 298.

Vallee,F., Lipari,F., Yip,P., Sleno,B., Herscovics,A. and Howell,P.L. (2000b) Crystal structure of a class I alpha1,2-mannosidase involved in *N*-glycan processing and endoplasmic reticulum quality control. *EMBO J.*, **19**, 581–588.

Wright,S. (ed) (1977) *Experimental Results and Evolutionary Deductions*. University of Chicago Press, Chicago.

Wyckoff,G.J., Wang,W. and Wu,C.I. (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature*, **403**, 304–309.

Zhang,J., Rosenberg,H.F. and Nei,M. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl Acad. Sci. USA*, **95**, 3708–3713.