Strategic protein target analysis for developing drugs to stop dental caries

Jeremy A Horst[1,2,3*], Ursula Pieper[2], Andrej Sali[2], Ling Zhan[1,4], Gaurav Chopra[3,5], Ram Samudrala[3], John DB Featherstone[4]

[1] *Division of Pediatric Dentistry, Department of Orofacial Sciences*; [2] *Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California, San Francisco*; [3] *Department of Microbiology, University of Washington*; [4] *Department of Preventive and Restorative Dental Sciences, University of California, San Francisco*; [5] *Department of Structural Biology, Stanford University*.

[*]Corresponding author: Jeremy A Horst, UCSF Pediatric Dentistry, 707 Parnassus, University of California at San Francisco, San Francisco, CA 94134, USA. tel: +1 415 476 3276; fax: +1 415 514 4231.
E-mail: *horst@salilab.org*

Running title: Drug targets for dental caries

Total number of pages, figures, supplementary tables: 15, 4, 1.

Submitted to: *Advances in Dental Research*

Date: March 23, 2012

Keywords: Dental caries, *Streptococcus mutans*, drug target, protein structure prediction, proteome, druggability, phylogenetic profile, drug side effects.

Abstract

Dental caries is the most common disease to cause irreversible damage in humans. Several therapeutic agents are available to treat or prevent dental caries, but none besides fluoride have significantly influenced the disease burden globally. Etiologic mechanisms of the mutans group streptococci and specific *Lactobacillus* species have been characterized to various degrees of detail, from identification of physiologic processes to specific proteins. Here, we analyze the entire *Streptococcus mutans* proteome for potential drug targets by investigating their uniqueness with respect to noncariogenic dental plaque bacteria, quality of protein structure models, and likelihood of finding a drug for the active site. Our results suggest specific targets for rational drug discovery, including 15 known virulence factors, 16 proteins for which crystallographic structures are available, and 84 previously uncharacterized proteins, with varying levels of similarity to homologs in dental plaque bacteria. This analysis provides an effective map to streamline the process of developing clinically effective multispecies pharmacologic interventions for dental caries.

1. Introduction

Impact of caries. Dental caries affects the vast majority of people in developed nations (WHO Report on oral health, 2003), and annually costs the United States approximately $750 million, 164 million work hours, and 51 million school hours (US 2002 Oral Health Annual Report, NIDCR/CDC). The multifactorial etiology of dental caries includes multiple bacterial species (Loesche *et al.*, 1986) and nutrients that enable bacterial acidogenesis (van Palenstein Helderman *et al.*, 1996). Factors influencing susceptibility include age, immunologic status (Taubman and Smith, 1992), salivary function (Brown *et al.*, 1976), human genetics (Wright, 2010), bacterial genetics (Loesche, 1986), and behavioral practices such as diet (Miller, 1890; Hefferren, 1986; Mundorff *et al.*, 1990) and hygiene (Featherstone, 2000; Milgrom *et al.*, 2009).

Paradigm not sufficient. The most effective approach to preventing dental caries is to completely exclude refined sugars from the diet (*e.g*. sucrose, fructose) and to promote consumption of protein, lipids, and complex carbohydrates (Hefferren, 1986; Mundorff *et al.*, 1990; van Palenstein Helderman *et al.*, 1996; Featherstone, 2000). However, within the constraints of consumer culture, the inability of law, policy, or behavioral psychology to effect dietary changes at the population level, and the difficulty with mass distribution of knowledge, dietary changes are not expected to impact the pandemic of dental caries significantly in the near future. Thus, innovative approaches are needed (Milgrom *et al.*, 2009). Pharmacologic medicaments present one option to globally treat dental caries.

Drug targets needed. The primary targets of contemporary pharmacologic treatment are the etiologic bacteria and the diseased tissues. The tooth is the target of various effective chemically based regimens for prevention and regeneration (Featherstone, 2009). Some natural and synthetic molecular agents show moderate efficacy against cariogenic bacteria, but no clinical panacea has been found. Concerted approaches to rational drug design are rare. As drugs traditionally target protein binding sites, protein structure is required for rational design (Horst *et al.*, 2011). The recent explosion of sequencing technology made available the protein sequences corresponding to all genes of *Streptococcus mutans* and various other dental plaque bacteria (Figure 1b). Combined with comparative structure prediction to model structure from sequence (Sali and Blundell, 1993), we are presented with the novel opportunity to rationally design multitarget multispecies drugs.

Organisms to attack. Although the concept of targeting *S. mutans* alone is attractive, multispecies therapy is essential because multiple species contribute to dental caries. Caries experience seems to depend more on diet than on the prevailing plaque species (van Palenstein Helderman *et al.*, 1996). Additionally, *S. mutans* levels in older patients do not correlate with caries experience (Milgrom *et al.*, 2009), and inverse associations of caries experience with *S. mutans* detection are reported for children with blood dyscrasias (Ou-Yang *et al.*, 2010). Even when *S. mutans* correlates best to caries experience, many other species and genera are also

significantly associated (Tanner *et al.*, 2011). Furthermore, histologically distinct regions of caries lesions have been found to associate with different bacteria: in early lesions, lack of cultivatable *Veillonella* is associated with lack of *S. mutans* (Marsh *et al.*, 1989). Meanwhile, evidence suggests that some bacteria are protective, and should be permitted to thrive. In Figure 1b, we detail the species that appear to be contributory (n=16) or protective (n=7) for dental caries. We investigate them here as target or antitarget species, respectively.

Proteins to attack. In this work we estimate the likelihood of each *S. mutans* protein to be successfully targeted by structure-based drug discovery (Becker et al., 2004; Jenwitheesuk et al., 2008; Fan et al., 2011). We funnel down the entire proteome to those sequences for which highly reliable models can be computed or for which experimentally determined structures are available (modelable), and then to those with binding site features similar to known drug targets (druggable). We then continue to funnel these proteins for uniqueness to *S. mutans* by comparing each to the entire proteomes of 23 dental plaque bacteria, stratified by contribution to dental caries (Figure 1). We predict whether pharmacologic inhibition of any *S. mutans* protein would also selectively inhibit other cariogenic bacteria. This is a guide to strategic target selection for effective long-term preventive and therapeutic pharmacologic interventions. This approach is novel to dental caries, and provides a model for chronic multi-bacterial diseases.

2. Methods

Our approach is outlined in Figure 1. We take a three stage approach to assess the likelihood of a given protein interaction site to bind a drug-like compound (druggability) and of a drug for that protein to target other dental plaque bacteria. In the first stage we build atomic models with all relevant templates. In the second stage we assess the druggability of the template that was used to generate the best model. In the third stage we assess the similarity of each protein to all proteins (proteomes) in dental plaque bacteria.

2.a. Sequences and structures. All available *S. mutans* protein structures were obtained from the Protein Data Bank (PDB; Berman, *et al.*, 2000; accessed October 4th, 2011). All protein sequences ascribed to the reviewed complete proteome sets for *S. mutans* and other dental plaque bacteria were downloaded from UniProtKB (Ajdić *et al.*, 2002; Apweiler *et al.*, 2004; accessed January 16th, 2011).

2.b. Comparative modeling. To generate atomic models for each *S. mutans* protein, we applied the restraint-based comparative modeling program MODELLER-v9.10 (Sali and Blundell, 1993). The model dataset was generated using the automated modeling pipeline ModPipe (Pieper *et al.*, 2008), including template selection and target-template alignment (MODELLER, PSI-BLAST) with crystal structures available in a subset of the PDB with redundancy removed at the 95% sequence identity level, model building, and model evaluation. To select the most accurate model for each sequence from the model pool created by ModPipe, we applied the Z-

score of the DOPE atomic distance-dependent statistical potential (zDOPE; Shen and Sali, 2006), which estimates the reliability of each model. zDOPE < -1 indicates that the modeling process identified the native fold topology, which is deemed "modelable."

2.c. Druggability. To predict proteins that bind compounds which satisfy Lipinski's Rule of 5 (Lipinski *et al.*, 1997) and have ≤10 rotatable bonds, we applied the DrugEBIlity analysis (Agüero *et al*., 2008). The DrugEBIlity score is calculated as the mean of 11 machine learning algorithms, separately trained with 25 physicochemical descriptors of all known drug binding sites. To obtain predictions of high specificity, we applied the threshold of satisfying at least 8 of the 11 algorithms (drugEBIlity score > 0.5; Figure 2).

2.d. Targeting other dental plaque bacteria. To anticipate analogous targeting of other relevant bacteria, we built HHsearch HMM-based phylogenetic profiles for each *S. mutans* to all proteins in other dental plaque bacteria. We built an HMM with HHsearch for each protein in each proteome by comparing similarity patterns found in the 70% and 90% nonredundant NCBI protein sequence database by fold family hierarchically, and calibrating (normalizing) against a set of HMMs including one for each fold family in SCOP. We compared the HMM for each *S. mutans* protein to all 23 cariogenic and noncariogenic bacterial proteomes using HHSearch. HHsearch evaluates protein similarity by maximizing the co-emission log-odds probability for a pair of HMMs, which represent position-specific insertion-deletion probabilities of multiple sequence alignment profiles (Söding, 2005). We plotted the proportion of matching HMM alignment columns for the most similar protein in each proteome (Figures 3, 4).

3. Results and Discussion

3.a. Druggable modelable proteins found.

15 known virulence factors, 16 proteins for which crystal structures are available, and 84 previously unidentified proteins were identified as modelable and druggable. All comparative models are available through ModBase (http://modbase.compbio.ucsf.edu).

We illustrate protein druggability using six proteins with highly reliable models (Figure 2). First, as predicted by the drugEBIlity score of -0.71, the proton/lactate pump (P50976) has no detectable pocket large enough for any drug-like compound (Figure 2a). Second, the large central cavity of the multiple sugar binding protein (Q00749; drugEBIlity score 0.76) is large enough to fit galactose (shown), other sugars, and most drug families (Figure 2b). Third, the cell-surface adhesin presents a shallow cleft predicted to bind an RNA strand, like the template does (1ddl; Figure 2c). Fourth, glucosamine-6-phosphate deaminase illustrates a druggable pocket from a crystal structure, with suitable geometry and chemistry to bind the glucosamine-6-phosphate, other physiologic riboses, and other drug analogs (Figure 2d). Fifth, among all *S. mutans* proteins for which a crystal

structure is not yet available, uracil-diphosphate acetyl-glucosamine epimerase (Q8DTB7) bears the binding site predicted with the highest confidence to be pharmacologically inhibited (Figure 2e). The fit of the UDP from template structure 3beo suggests accurate modeling of the binding site: the long, narrow pocket, and the hydrophobic patch at the end (red) are favorable conditions to enable drug-induced inhibition. Sixth, a completely uncharacterized protein exemplifies a protein predicted to be modelable, druggable, and is relatively unique to *S. mutans* (Figure 2f). All modelable and druggable proteins represent potential drug targets.

3.b. Virulence factors annotated.

61 proteins contributing to cariogenesis were identified from the literature, in general because inhibition has reduced some parameter of cariogenicity (Supplemental table 1). The strength of evidence for each protein being a virulence factor corresponds to the clinical relevance of the model system in which experiments were performed, the method by which the protein was inhibited, and the magnitude of impact on surrogate markers of cariogenesis. We annotated 22 of these proteins with highly reliable (zDOPE < -1) or moderately reliable (zDOPE <-0.5) atomic models; the druggability of discovering a drug for the template protein; and comparison of phylogenetic profiles among cariogenic or protective bacterial species (Figure 3). These proteins are categorized according to etiologic mechanisms and physiologic processes essential to bacterial colonization and thriving (Figure 3; Supplemental Table 1).

3.b.i. Targeting metabolism. Physiologic responses optimized to environment-specific nutrients may be exploitable by rational drug discovery. The potential drug targets within this set of physiologic processes include multiple sugar-binding protein (UniProt Q00749; Figure 2b), fructose phosphotransferase (Q8DUN3), purine nucleoside phosphorylase (Q8DTU4), glycogen synthase (Q8CWX0), signal recognition particle (Q54431), formyltetrahydrofolate ligase (Q59925), and panthothenate flavoprotein (Q8DU74). For all these proteins, homologs are identified in the vast majority of bacteria sampled, suggesting general cross reactivity (Figure 3).

3.b.ii. Targeting attachment. The mutans group streptococci secrete a sticky matrix that adheres to many surfaces and proteins to attach to the matrix. Modelable and druggable proteins of this category include glycogen phosphorylase (Q8DT55), another phosphorylase (Q8DT31), dextran glucosidase (Q99040), secreted peptidoglycan hydrolase (Q8DWM3), glucan-binding protein-C response regulator (Q9S151), and cell surface adhesin (P11657). Most of these proteins are present in all sampled proteomes. The hydrolase is present more in protective than cariogenic bacteria, and therefore is not a good target, whereas the adhesin is relatively unique to *S. mutans* and is a good target.

3.b.iii. Targeting coordination. Environmental adaptation is facilitated by proteins that signal changes via quorum sensing. Bromodomain-containing RNA-binding protein-2 response regulator (Q8DVJ8) and oxidative stress sensor kinase (Q8DT64) are ubiquitous and predicted to be potential drug targets.

3.b.iv. No preferential profiles. None of the putative virulence factors are more prevalent in cariogenic bacteria. Rather, all are either ubiquitous in the set, common to all *Streptococci* and *Bifidobacteria* but absent from *Lactobacilli*, or relatively unique to *S. mutans* (Figure 3). Specific analyses of binding site residues may reveal more specificity than estimated by this ortholog prediction.

3.c. Targeting known structures.

Crystal structures provide the most globally accurate models currently obtainable, and are generally preferable for drug discovery (Baker and Sali, 2001), although comparative models can also be useful (Becker *et al*., 2004; Fan *et al.*, 2009). We predict 14 out of 81 known *S. mutans* to be highly amenable to drug discovery, and present their phylogenetic profiles to aid design of specificity (Figure 4a).

3.d. The most modelable and druggable proteins in *S. mutans*.

Our *S. mutans* proteome modeling and druggability experiment discovered 84 novel high quality models (zDOPE < -1) with highly druggable template structures (> 0.5; Figures 1a, 3b). While functional annotations have been made by sequence comparison, most of these proteins are not well studied. We assert these proteins as suitable targets for rational drug discovery. Future work on these proteins could include crystallography with physiologic ligand analogs, high throughput screening, or using computational multitarget molecular docking studies (CANDO: http://cando.compbio.washington.edu/wiki).

4. Conclusions

The character of a bacterial species is found in the divergent structural features and the differential physiologic responses to environmental shifts. To inform a strategic plan against *S. mutans* we assessed the accessibility of its structural features to rational drug discovery, and the uniqueness of its proteins with respect to other relevant bacteria in the dental plaque. We performed this analysis to inform discovery of pharmacologic inhibitors for dental caries.

Unfortunately, no druggable proteins were found to be differentially abundant in cariogenic bacteria. It seems the probability of developing a highly accurate model for a given protein is greatly increased for well-studied protein families, as more template structures are available for them; physiologically central roles are of high interest for study, but centrality equates to ubiquity, so modelable proteins tend to be common.

Inability to produce accurate models with the current PDB makes no statement about the druggability of the protein; it is simply not possible to perform structural analysis without a structure: many currently unmodelable proteins are expected to be drug targets. Bench assays and crystallography are indicated for proteins with no template that correlate closely with cariogenecity. Meanwhile, the 15 virulence factors predicted to be modelable and druggable validate the funnel approach we took to analyze the full proteome.

The information explosion in sequence and structural data can be cross-referenced with epidemiologic data that identify differential gene presence (Zhang *et al.*, 2009) or *in vitro* studies of gene expression (Sol *et al.*, 2011). These, and environment-specific phylogenetic analyses will become more meaningful as sequencing data expands to the many yet unrepresented dental plaque bacterial species.

A subset of the targets identified here will progress to virtual screening, which have resulted in selecting verifiable hits with 40-60% accuracy when applied with our recent protocols to crystal structures or comparative models constructed from templates as low as 30% sequence identity (Fan et al., 2009; Horst et al., 2011). In our experience a week's worth of effort is sufficient to model, dock, and select compounds for one protein. Thereafter, virtual hits must be tested at the bench. It is expected that application to the modelable and druggable proteins identified here will lead to *in vitro* hits for at least some of these proteins. Focusing on proteins that are at least moderately unique to *S. mutans* (rare, Figure 1a) will add specificity over other dental plaque bacteria, enabling a shift in the microbial ecology. Selecting compounds that are predicted to target multiple proteins have been successful in other disease models (Jenwitheesuk *et al.*, 2008). Elevating the search for specific multispecies inhibition would make dental caries a useful study model for other biofilm-mediated diseases, such as periodontitis, ulcers, enteritis, and gluten sensitivity.

6. Acknowledgements

This work was supported by NIH grants DP1-OD006779, U54-GM094662, and P01-GM71790.

5. References

Agüero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, *et al* (2008). Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov* 7(11):900-7.

Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, *et al* (2002). Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proc Natl Acad Sci USA* 99(22):14434-9.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, *et al* (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 32:D115-9.

Baker D, Sali A (2001). Protein structure prediction and structural genomics. *Science* 294:93-6.

Becker OM, Marantz Y, Shacham S, Inbal B, Heifetz A, Kalid O, *et al* (2004). G protein-coupled receptors: in silico drug discovery in 3D. *Proc Natl Acad Sci USA*. 101 (31):11304-9.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al* (2000). The Protein Data Bank. *Nucleic Acids Res* 28(1):235-42.

Brown LR, Dreizen S, Handler S (1976). Effects of selected caries preventive regimens on microbial changes following irradiation-induced xerostomia in cancer patients. In Stiles HJ, Loesche WJ, O'Brien TC. *Proceedings: microbial aspects of dental caries*.

Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A (2009). Molecular docking screens using comparative models of proteins. *J Chem Inf Model* 49(11):2512-27.

Featherstone JD (2000). The science and practice of caries prevention. *J Am Dent Assoc* 131:887–899.

Featherstone JD (2009). Remineralization, the natural caries repair process--the need for new approaches. *Adv Dent Res* 21(1):4-7.

Hefferren JJ (1986). Proceedings, scientific consensus conference on methods for assessment of the cariogenic potential of foods. *J Dent Res* 65(s):1473:1544.

Horst JA, Laurenzi A, Bernard B, Samudrala R (2011). Computational multitarget drug discovery. In JE Peters, ed. *Polypharmacology*.

Jenwitheesuk E, Horst JA, Rivas KL, Van Voorhis WC, Samudrala R (2008). Novel paradigms for drug discovery: computational multitarget screening. *Trends Pharmacol Sci* 29:62-71.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* 23:3-25.

Loesche WJ (1986). Role of Streptococcus mutans in human dental decay. *Microbiol Rev* 50:353–380.

Marsh PD, Featherstone A, McKee AS, Hallsworth AS, Robinson C, Weatherell JA, *et al* (1989). A microbiological study of early caries of approximal surfaces in schoolchildren. *J Dent Res* 68(7):1151-4.

Milgrom P, Zero DT, Tanzer JM (2009). An examination of the advances in science and technology of prevention of tooth decay in young children since the Surgeon General's Report on Oral Health. *Acad Pediatr* 9(6):404-409.

Miller WD (1890). The micro-organisms of the human mouth. *The SS White Manufacturing Co.*, Philadelphia.

Mundorff SA, Featherstone JD, Bibby BG, Curzon MEJ, Eisenberg AD, Espeland MA (1990). Cariogenic potential of foods: I. caries in the rat model. *Caries Res* 24:344-355.

Ou-Yang LW, Chang PC, Tsai AI, Jaing TH, Lin SY (2010). Salivary microbial counts and buffer capacity in children with acute lymphoblastic leukemia. *Pediatr Dent* 32(3):218-22.

Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, Braberg H, *et al* (2011). ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res* 39:D465-74.

Sali A, Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779-815.

Söding J (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951-960.

Shen M-Y, Sali A (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15:2507–2524

Sol A, Feuerstein O, Featherstone JD, Steinberg D (2011). Effect of sublethal CO2 laser irradiation on gene expression of streptococcus mutans immobilized in a biofilm. *Caries Res* 45(4):361-9.

Tanner AC, Mathney JM, Kent RL, Chalmers NI, Hughes CV, Loo CY, *et al* (2011). Cultivable anaerobic microbiota of severe early childhood caries. *J Clin Microbiol* 49(4):1464-74.

Taubman MA, Smith MA (1992). Significance of immune responses to oral antigens in dental diseases. In: *Genetically Engineered Vaccines*. Ciardi JE et al, eds. New York: Plenum Press, 1992:274.

van Palenstein Helderman WH, Matee MI, van der Hoeven JS, Mikx FH (1996). Cariogenicity depends more on diet than the prevailing mutans streptococcal species. *J Dent Res* 75:535–545.

Wright JT (2010). Defining the contribution of genetics in the etiology of dental caries. *J Dent Res* 89(11):1173-4.

Zhang L, Foxman B, Drake DR, Srinivasan U, Henderson J, Olson B, *et al* (2009). Comparative whole-genome analysis of Streptococcus mutans isolates within and among individuals of different caries status. *Oral Microbiol Immunol* 24(3):197-203.

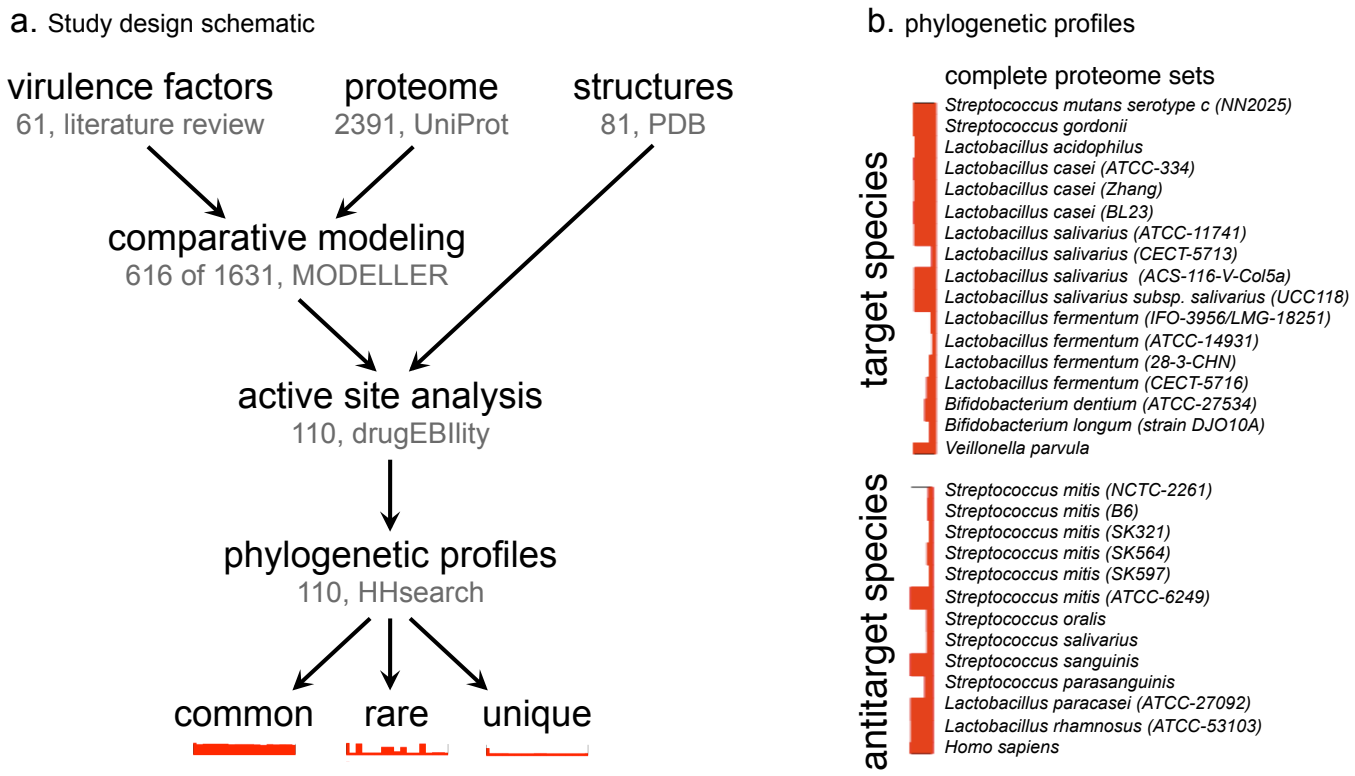
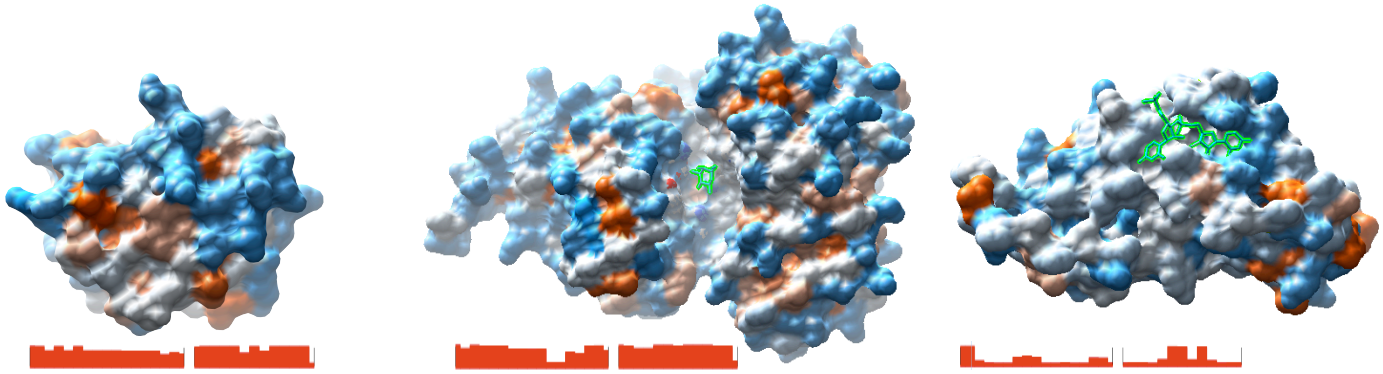


Figure 1. Study design. a. This study is designed to filter the *S. mutans* proteome for useful drug targets by searching for proteins that are modelable, druggable, and differentially abundant in cariogenic bacteria. The analytic steps are shown, including the abundance of proteins remaining at each step, the source databases (top row), and the analytic methods applied. Of 2391 sequences, models were produced for 1631 (68%), including 616 (26%) highly reliable models (zDOPE < -1). Along with 81 known structures, 110 proteins (18%) are predicted to be druggable. Prevalence among other dental plaque bacteria for the druggable proteins informs target selection for multispecies rational drug discovery. **b.** Phylogenetic profiles of *S. mutans* proteins amongst dental plaque bacteria that contribute to dental caries (target species) and those that are protective (antitarget species; references in Supplement). The protein sequences in each complete proteome set for each listed strain were converted into a queryable hidden Markov model database, and searched by HHSearch. The magnitude of similarity for the most similar protein in each proteome is represented by each notch of the profile. Target species for which complete proteome sets are not yet available include *Actinomyces naeslundii*, *Actinomyces gerensceriae*, *Parascardovia denticolens*, *Scardovia wiggsiae*, *Streptococcus cristatus*, and *Streptococcus sobrinus*.

Examples of modelable *S. mutans* virulence factors

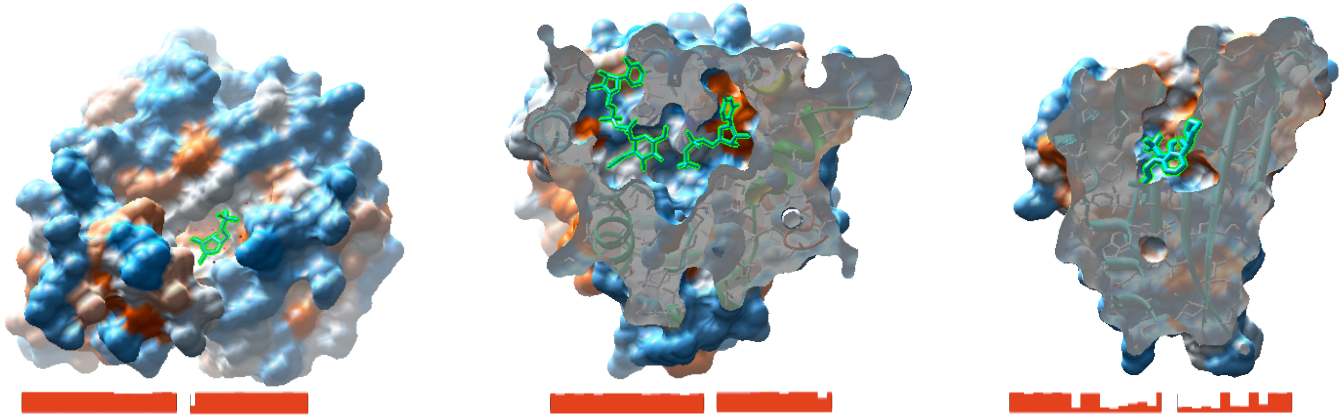


a. not druggable
proton/lactate pump
zDOPE (model reliability) = -1.73
drugEBIlity score = -0.71

b. druggable
mutltiple sugar binding protein
zDOPE (model reliability) = -1.57
drugEBIlity score = 0.76

c. druggable and favorable profile
major cell-surface adhesin
zDOPE (model reliability) = -1.50
drugEBIlity score = 0.41

Examples of modelable and druggable *S. mutans* proteins

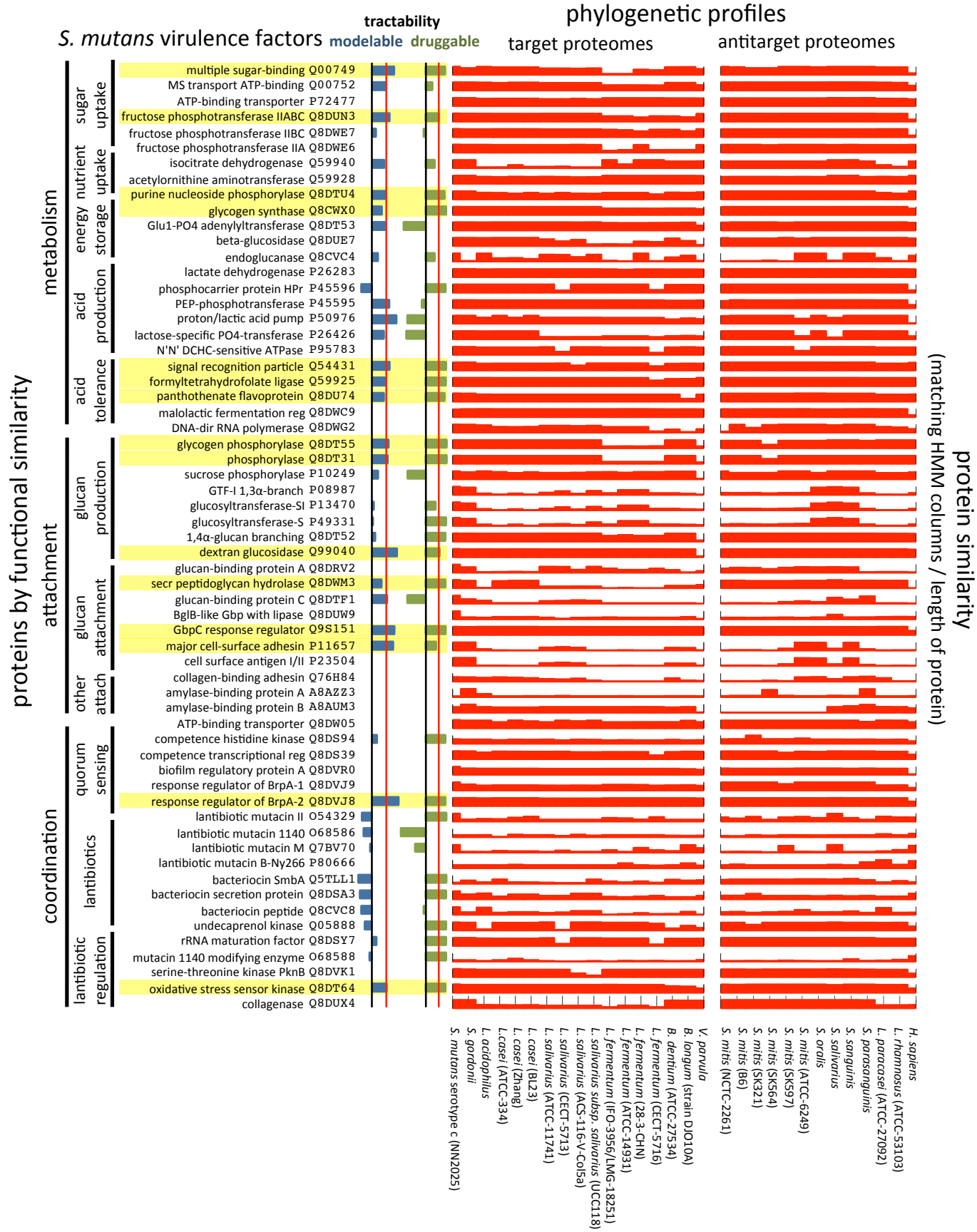


d. crystal structure
glucosamine-6-PO_4 deaminase
resolution = 2.03 Å
drugEBIlity score = 0.77

e. comparative model
nonhydrolyzing UDP acetyl-glucosamine epimerase
zDOPE (model reliability) = -1.75
drugEBIlity score = 0.82

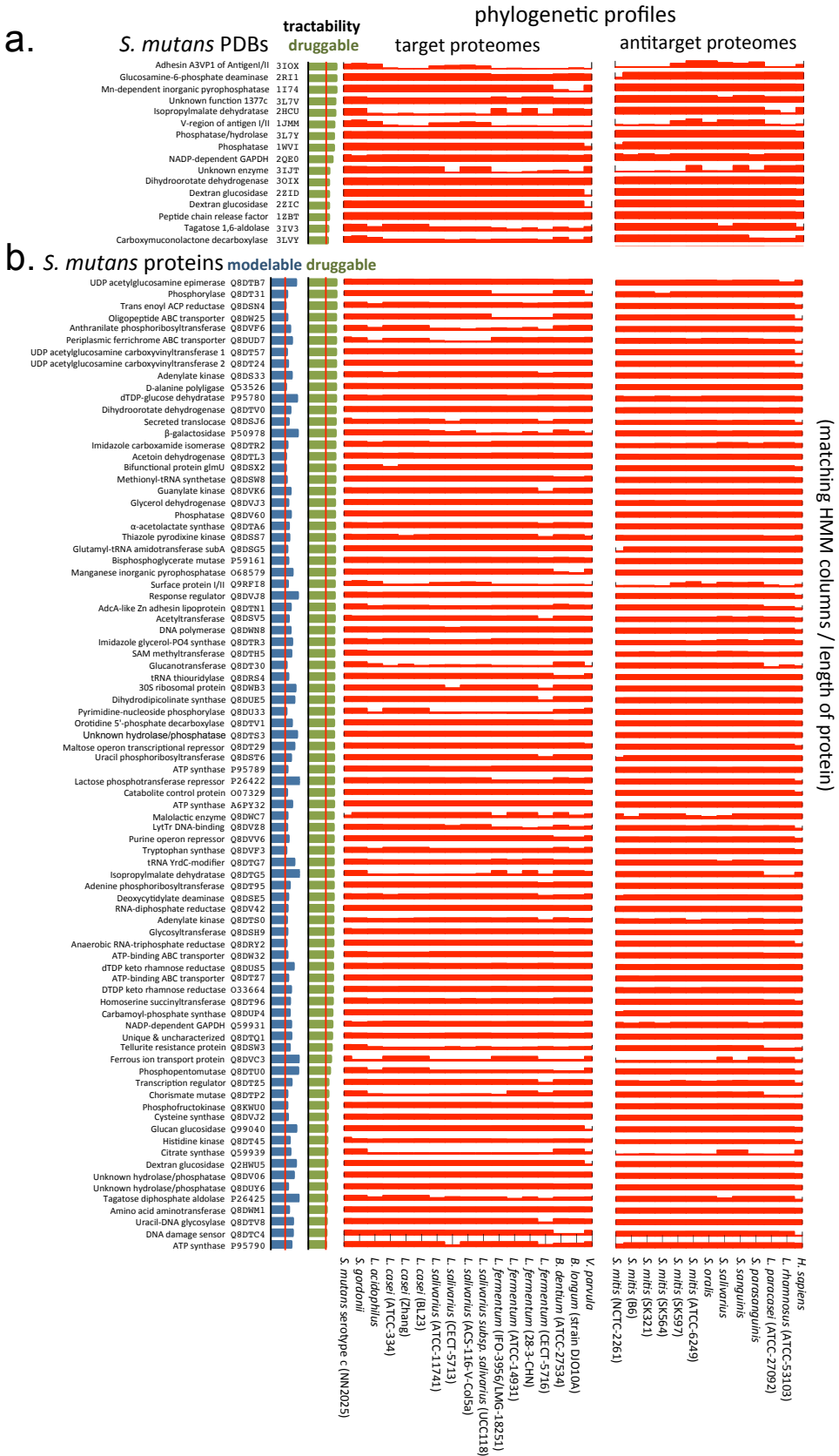
f. favorable profile
unknown enzyme
resolution = 2.38 Å
drugEBIlity score = 0.57

Figure 2. Structural features of target protein druggability. Protein surfaces shown with hydropathy plot (red as hydrophobic, blue as hydrophilic), and ligands highlighted to exemplify patterns of favorable drug binding sites. Phylogenetic profiles (red) display prevalence in other cariogenic (target, left) and protective bacteria (antitarget, right). **a.** A virulence factor with no detectable pocket is not expected to be an effective drug target. *S. mutans* proteins previously **b-c.** identified or **d-f.** unidentified as useful drug targets that are relatively **b,d,e.** common or **c,f.** rare to dental plaque bacteria. Each presents a cavernous pocket of favorable size, shape, and chemical composition for rational drug design. **e,f.** Surface cut to reveal binding site. Mapped ligands: **b.** galactose (PDB=2b3f), **c.** diuracil (1ddl), **e.** uracil-diphosphate (3beo), **f.** cephalosporins (1cef, 1hvb).


proteins by functional similarity
S. mutans virulence factors
tractability
modelable
druggable
phylogenetic profiles
target proteomes
antitarget proteomes
protein similarity
(matching HMM columns / length of protein)
metabolism
coordination
sugar uptake
nutrient uptake
energy storage
acid production
acid tolerance
attachment
glucan production
glucan attachment
other attach
quorum sensing
lantibiotics
lantibiotic regulation
multiple sugar-binding Q00749
MS transport ATP-binding Q00752
ATP-binding transporter P72477
fructose phosphotransferase IIABC Q8DUN3
fructose phosphotransferase IIBC Q8DWE7
fructose phosphotransferase IIA Q8DWE6
isocitrate dehydrogenase Q59940
acetylornithine aminotransferase Q59928
purine nucleoside phosphorylase Q8DTU4
glycogen synthase Q8CWX0
Glu1-PO4 adenylyltransferase Q8DT53
beta-glucosidase Q8DUE7
endoglucanase Q8CVC4
lactate dehydrogenase P26283
phosphocarrier protein HPr P45596
PEP-phosphotransferase P45595
proton/lactic acid pump P50976
lactose-specific PO4-transferase P26426
N'N' DCHC-sensitive ATPase P95783
signal recognition particle Q54431
formyltetrahydrofolate ligase Q59925
panthothenate flavoprotein Q8DU74
malolactic fermentation reg Q8DWC9
DNA-dir RNA polymerase Q8DWG2
glycogen phosphorylase Q8DT55
phosphorylase Q8DT31
sucrose phosphorylase P10249
GTF-I 1,3α-branch P08987
glucosyltransferase-SI P13470
glucosyltransferase-S P49331
1,4α-glucan branching Q8DT52
dextran glucosidase Q99040
glucan-binding protein A Q8DRV2
secr peptidoglycan hydrolase Q8DWM3
glucan-binding protein C Q8DTF1
BglB-like Gbp with lipase Q8DUW9
GbpC response regulator Q9S151
major cell-surface adhesin P11657
cell surface antigen I/II P23504
collagen-binding adhesin Q76H84
amylase-binding protein A A8AZZ3
amylase-binding protein B A8AUM3
ATP-binding transporter Q8DW05
competence histidine kinase Q8DS94
competence transcriptional reg Q8DS39
biofilm regulatory protein A Q8DVR0
response regulator of BrpA-1 Q8DVJ9
response regulator of BrpA-2 Q8DVJ8
lantibiotic mutacin II O54329
lantibiotic mutacin 1140 O68586
lantibiotic mutacin M Q7BV70
lantibiotic mutacin B-Ny266 P80666
bacteriocin SmbA Q5TLL1
bacteriocin secretion protein Q8DSA3
bacteriocin peptide Q8CVC8
undecaprenol kinase Q05888
rRNA maturation factor Q8DSY7
mutacin 1140 modifying enzyme O68588
serine-threonine kinase PknB Q8DVK1
oxidative stress sensor kinase Q8DT64
collagenase Q8DUX4
S. mutans serotype c (NN2025)
S. gordonii
L. acidophilus
L.casei (ATCC-334)
L. casei (Zhang)
L. casei (BL23)
L. salivarius (ATCC-11741)
L. salivarius (CECT-5713)
L. salivarius (ACS-116-V-Col5a)
L. salivarius subsp. salivarius (UCC118)
L. fermentum (IFO-3956/LMG-18251)
L. fermentum (ATCC-14931)
L. fermentum (28-3-CHN)
L. fermentum (CECT-5716)
B. dentium (ATCC-27534)
B. longum (strain DJO10A)
V. parvula
S. mitis (NCTC-2261)
S. mitis (B6)
S. mitis (SK321)
S. mitis (SK564)
S. mitis (SK597)
S. mitis (ATCC-6249)
S. oralis
S. salivarius
S. sanguinis
L. paracasei (ATCC-27092)
L. rhamnosus (ATCC-53103)
H. sapiens

(*previous page*) **Figure 3. Putative *S. mutans* virulence factor modelability, druggability, and phylogenetic profiles.** Previously identified etiologic proteins were categorized according to etiologic mechanisms and physiologic processes essential to colonization and thriving (Supplementary Table 1). Each protein was compared to the proteomes of cariogenic (target) and noncariogenic (antitarget) dental plaque bacteria (listed at bottom). Red bars depict the magnitude of similarity for the most similar protein in each proteome, illustrating the uniqueness of the protein in the context of the floral environment, and therefore the potential impact of targeting this protein. Shown in columns are the model quality (zDOPE, blue) and drugEBIlity score of the best template structure (green). Threshold scores for drugEBIlity (0.5) and model quality (-0.5) are indicated with red lines. Proteins with scores above both are highlighted. The relation of these data can guide the selection of protein targets for rational drug discovery to treat dental caries.

(*next page*) **Figure 4. The most druggable proteins of known and unknown structure in *S. mutans*.** Druggable *S. mutans* proteins with **a.** available crystallographic structures or **b.** reliable models were assessed for similarity to the proteomes of cariogenic and noncariogenic dental plaque bacteria. The thresholds and depictions are as described in Figure 3. Many highly reliable models are amenable to rational drug discovery, and may have tunable side effects on other dental plaque bacteria.


a.
S. mutans PDBs
tractability
druggable
phylogenetic profiles
target proteomes
antitarget proteomes
Adhesin A3VP1 of Antigen I/II 3IOX
Glucosamine-6-phosphate deaminase 2RI1
Mn-dependent inorganic pyrophosphatase 1I74
Unknown function 1377c 3L7V
Isopropylmalate dehydratase 2HCU
V-region of antigen I/II 1JMM
Phosphatase/hydrolase 3L7Y
Phosphatase 1WVI
NADP-dependent GAPDH 2QBQ
Unknown enzyme 3LJT
Dihydroorotate dehydrogenase 3OIX
Dextran glucosidase 2ZID
Dextran glucosidase 2ZIC
Peptide chain release factor 1ZBT
Tagatose 1,6-aldolase 3IV3
Carboxymuconolactone decarboxylase 3LVY
b. S. mutans proteins modelable druggable
UDP acetylglucosamine epimerase Q8DTB7
Phosphorylase Q8DT31
Trans enoyl ACP reductase Q8DSN4
Oligopeptide ABC transporter Q8DW25
Anthranilate phosphoribosyltransferase Q8DVF6
Periplasmic ferrichrome ABC transporter Q8DUD7
UDP acetylglucosamine carboxyvinyltransferase 1 Q8DT57
UDP acetylglucosamine carboxyvinyltransferase 2 Q8DT24
Adenylate kinase Q8DS33
D-alanine polyligase Q53526
dTDP-glucose dehydratase P95780
Dihydroorotate dehydrogenase Q8DTV0
Secreted translocase Q8DSJ6
β-galactosidase P50978
Imidazole carboxamide isomerase Q8DTR2
Acetoin dehydrogenase Q8DTL3
Bifunctional protein glmU Q8DSX2
Methionyl-tRNA synthetase Q8DSW8
Guanylate kinase Q8DVK6
Glycerol dehydrogenase Q8DVJ3
Phosphatase Q8DV60
α-acetolactate synthase Q8DTA6
Thiazole pyrodixine kinase Q8DSS7
Glutamyl-tRNA amidotransferase subA Q8DSG5
Bisphosphoglycerate mutase P59161
Manganese inorganic pyrophosphatase Q68579
Surface protein I/II Q9RFI8
Response regulator Q8DVJ8
AdcA-like Zn adhesin lipoprotein Q8DTN1
Acetyltransferase Q8DSV5
DNA polymerase Q8DWN8
Imidazole glycerol-PO4 synthase Q8DTR3
SAM methyltransferase Q8DTH5
Glucanotransferase Q8DT30
tRNA thiouridylase Q8DRS4
30S ribosomal protein Q8DWB3
Dihydrodipicolinate synthase Q8DUE5
Pyrimidine-nucleoside phosphorylase Q8DU33
Orotidine 5'-phosphate decarboxylase Q8DTV1
Unknown hydrolase/phosphatase Q8DTS3
Maltose operon transcriptional repressor Q8DT29
Uracil phosphoribosyltransferase Q8DST6
ATP synthase P95789
Lactose phosphotransferase repressor P26422
Catabolite control protein O07329
ATP synthase A6PY32
Malolactic enzyme Q8DWC7
LytTr DNA-binding Q8DVZ8
Purine operon repressor Q8DVV6
Tryptophan synthase Q8DVF3
tRNA YrdC-modifier Q8DTG7
Isopropylmalate dehydratase Q8DTG5
Adenine phosphoribosyltransferase Q8DT95
Deoxycytidylate deaminase Q8DSE5
RNA-diphosphate reductase Q8DV42
Adenylate kinase Q8DTS0
Glycosyltransferase Q8DSH9
Anaerobic RNA-triphosphate reductase Q8DRY2
ATP-binding ABC transporter Q8DW32
dTDP keto rhamnose reductase Q8DUS5
ATP-binding ABC transporter Q8DTZ7
DTDP keto rhamnose reductase O33664
Homoserine succinyltransferase Q8DT96
Carbamoyl-phosphate synthase Q8DUP4
NADP-dependent GAPDH Q59931
Unique & uncharacterized Q8DTQ1
Tellurite resistance protein Q8DSW3
Ferrous ion transport protein Q8DVC3
Phosphopentomutase Q8DTU0
Transcription regulator Q8DTZ5
Chorismate mutase Q8DTP2
Phosphofructokinase Q8KWU0
Cysteine synthase Q8DVJ2
Glucan glucosidase Q99040
Histidine kinase Q8DT45
Citrate synthase Q59939
Dextran glucosidase Q2HWU5
Unknown hydrolase/phosphatase Q8DV06
Unknown hydrolase/phosphatase Q8DUY6
Tagatose diphosphate aldolase P26425
Amino acid aminotransferase Q8DWM1
Uracil-DNA glycosylase Q8DTV8
DNA damage sensor Q8DTC4
ATP synthase P95790
S. mutans serotype c (NN2025)
S. gordonii
L. acidophilus
L. casei (ATCC-334)
L. casei (Zhang)
L. casei (BL23)
L. salivarius (ATCC-11741)
L. salivarius (ECT-5713)
L. salivarius (ACS-116-V-Col5a)
L. salivarius subsp. salivarius (UCC118)
L. fermentum (IFO-3956/LMG-18251)
L. fermentum (ATCC-14931)
L. fermentum (28-3-CHN)
L. fermentum (CECT-5716)
B. dentium (ATCC-27534)
B. longum (strain DJ010A)
V. parvula
S. mitis (NCTC-2261)
S. mitis (B6)
S. mitis (SK321)
S. mitis (SK564)
S. mitis (SK597)
S. mitis (ATCC-6249)
S. oralis
S. salivarius
S. sanguinis
S. parasanguinis
L. paracasei (ATCC-27092)
L. rhamnosus (ATCC-53103)
H. sapiens
protein similarity
(matching HMM columns / length of protein)