THEMATICS is Effective for Active Site Prediction in Comparative Model Structures

Ihsan A. Shehadi[a], Alper Uzun[b,d], Leonel F. Murga[c,d], Valentin Ilyin[b,d] and Mary Jo Ondrechen[c,d]

Department of Chemistry, United Arab Emirates University, Al-Ain, UNITED ARAB EMIRATES[a] and Department of Biology[b], Department of Chemistry and Chemical Biology[c], and Institute for Complex Scientific Software[d], Northeastern University, Boston, MA 02115 USA

mjo@neu.edu

Abstract

THEMATICS (Theoretical Microscopic Titration Curves) is a simple, reliable computational predictor of the active sites of enzymes from structure. Our method, based on well-established Finite Difference Poisson-Boltzmann techniques, identifies the ionisable residues with anomalous predicted titration behaviour. A cluster of two or more such perturbed residues is a very reliable predictor of the active site. The power of the method is that it only requires the three-dimensional structure as input. The protein does not have to bear any resemblance in sequence or structure to any previously characterized protein. The disadvantage of the method is that it does require the structure. We now present evidence that THEMATICS can also locate the active site in structures built by comparative modelling from similar structures. Results are given for three sets of orthologous proteins (Triosephosphate isomerase, 6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase, and Aspartate aminotransferase) and for one set of human homologues of Aldose reductase with different functions. In all of the cases studied, THEMATICS correctly locates the active site in the model structures. This suggests that the method can be applicable to proteins for which an experimentally determined structure is unavailable.

Keywords: Protein function; functional genomics; THEMATICS; electrostatics; titration.

1 Introduction

Theoretical Microscopic Titration Curves (THEMATICS) is a new technique, based on established computational methods, for locating the active sites of proteins from the three-dimensional structure alone (Ondrechen, Clifton et al. 2001; Shehadi 2002). The technique utilizes Finite Difference Poisson-Boltzmann methods (Bashford and Karplus 1991; Bashford and Gerwert 1992; Gilson 1993; Yang 1993; Sampogna and Honig 1994; Beroza 1995; Antosiewicz 1996; Carlson, Briggs et al. 1999) to obtain the electrical potential function for the protein, followed by calculation of the predicted titration curves for all of the ionisable residues in the protein structure. The shapes of the predicted titration curves are analysed to identify those residues with elongated, non-sigmoidal titration behaviour. A cluster of two or more such anomalous residues in physical proximity is a highly reliable predictor of the active site.

The power of THEMATICS is that it only requires the three-dimensional structure of the query protein as input. The query protein does not have to bear any resemblance in sequence or in structure to any previously characterized protein. However, herein lies the disadvantage of the method: One must have the three-dimensional structure of the protein. This raises the obvious question: Does one have to have an experimentally determined structure, or is a theoretical model structure good enough?

As a first step to answer this question, we decided to focus initially on structures built from sequence homology (Sali 1998). In the present work, we show how THEMATICS can predict active site location in structures built by comparative modelling (Sali 1998), starting with an experimentally determined template structure (Berman 2000; Westbrook 2003) and with the sequence of the query protein. THEMATICS is applied to four different sets of enzymes. Three of these are sets of the same enzyme (and therefore the same chemistry) from different organisms, while the fourth set is made up of homologous enzymes from the same species (human) but with different chemical functions. Sequence alignments were performed using ClustalW, http://www.ebi.ac.uk/ (Thompson, Higgins et al. 1994). The comparative (Sali 1995; Fiser 2000) model structures were constructed using the program MODELLER (Fiser 2000; Sanchez 2000), http://salilab.org/modeller/, and the program SWISS-MODEL (Schwede 2003), http://www.expasy.org/swissmod/SWISS-MODEL.html from the Swiss Institute of Bioinformatics.

The titration curves (mean net charge as a function of pH) are calculated for all of the ionisable residues in each of the template and model structures. The curves are analysed to select the ones that deviate from the typical sigmoidal shape. Most of the curves do possess the characteristic sigmoidal shape, with a sharp fall-off in charge in the region around the midpoint, as predicted by

the Henderson-Hasselbalch equation. Only a small fraction (about 3 to 7%) of the ionisable residues deviate from the typical behaviour; these are the residues identified by THEMATICS as potentially important. We then search for a cluster of residues with deviant titration behaviour that are in physical proximity. A residue is deemed to belong to a cluster if it is a nearest neighbour, or is within 6 Å, of another cluster member. Such clusters, which we call THEMATICS positive clusters, are highly reliable predictors of active site location in the protein structure.

THEMATICS calculations are simple and fast. On a single-processor desktop personal computer, total real time to analyse one protein ranges from less than one hour for the smallest enzymes to about one day for a large multimeric structure with thousands of residues. Total time is roughly a few hours for an average sized enzyme.

2 Application and results

2.1 Triosephosphate isomerase (TIM) orthologs.

Triosephosphate isomerase (TIM) catalyses the conversion of D-glyceraldehyde 3-phosphate (GAP) to dihydroxyacetone phosphate (DHAP) and has the α/β barrel ("TIM barrel") fold. The x-ray crystal structure data for TIM from chicken (PDB code 1TPH) (Zhang 1994) is obtained from the Protein Data Bank with a resolution of`1.8 A. Since TIM is active as a dimer, the calculations are performed on the dimer.

The first of the four structures homologous to the chicken TIM structure 1TPH is built from the sequence for *Schistosoma japonicum* (Liu 2003) with 60% sequence identity in the pair wise alignment and 0.16 Å RMSD value for the model structure. The second model is determined for the sequence for *Enterococcus faecalis* (Paulsen 2003) with 40.2 % sequence identity, resulting in a 0.29 Å RMSD value with the template structure. The third model is built from the sequence of *Bartonella henselae* with 38.7 % identity and RMSD value of 0.31 Å, while the last model is built from the sequence of *Mycoplasma genitalium* (Fraser 1995) with 33 % identity and RMSD value of 1.73 Å. These structures are all obtained with MODELLER and are summarized in Table 1.

Table 1 gives the THEMATICS result for the active site cluster for each template structure and the orthologous model structures. Known active site residues are shown in **boldface** and "second shell" residues (those immediately adjacent to active site residues but not considered to be in the active site) are underlined.

For the TIM structure from chicken (1TPH), four neighboring residues with anomalous titration behaviour are identified as the active site cluster. Two of these residues, H95 and E165, are well established by experiment as catalytically active residues (Lodi 1991; Zhang 1994)

Two other residues, C126 and Y164, are located in the active site cleft but any possible catalytic role for these residues has not been investigated experimentally. Upon alignment of the sequences and superposition of the structures, it is confirmed that all four of these residues are conserved, both in the sequence and in the spatial arrangement of the active site cleft, in all of the four model structures. Sequence alignment across a wider range of species again reveals high conservation of all four of these residues.

For all four of the model structures, THEMATICS finds the active site. THEMATICS identifies (by pronounced perturbed shape of the predicted titration curves) the two catalytic residues H95 and E165, plus C126, in all four of the model structures. For two of the four models, *S. japonicum* and *B. hensalae*, Y164 also exhibits perturbed titration behaviour. Y164 does not show significant perturbed titration behaviour in the model structures for *E. faecalis* and *M. genitalium*. Thus THEMATICS identifies the correct active site cluster for all of the model structures, but Y164 is not always included in the predicted active site cluster.

2.2 6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) orthologs.

6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase, HPPK, is a monomeric pyrophosphate transferase with α-β plaits topology. Its crystal structure for *E. coli* (PDB code 1HKA) was obtained from the protein data bank with 1.5 A resolution (Xiao 1999). Four homologous models to the *E. coli* structure 1HKA are built using the MODELLER software from the sequences for the following organisms: *Vibrio vulnificus* (Rhee 2002) (with 63% sequence identity with *E. coli* and 0.34 Å RMSD), *Vibrio parahaemolyticus* (Makino 2003) (with 57% sequence identity and 0.22 Å RMSD), *Pseudomonas aeruginosa* (Stover 2000) (with 51% sequence identity and 0.36 Å RMSD) and *Pseudomonas putida* (Nelson 2002) (with 48% sequence identity and 0.50 Å RMSD).

The predicted titration curves for residues D95, D97 and H115 in the template *E. coli* structure do not show the usual sigmoidal behaviour and are identified as the active site cluster. All three of these residues have been identified as active site residues in an experimentally determined inhibitor structure (Stammers 1999). All of them are conserved across the four species for which the model structures are built and are also generally well conserved across bacterial kinases. When the four model structures are overlaid onto the template *E. coli* structure, the positions of these residues are conserved in the active site pocket with similar orientations.

For the HPPK case, THEMATICS identifies the same cluster for all four of the model structures as for the *E. coli* template structure (see Table 1).

2.3 Aspartate aminotransferase (AspAT) orthologs.

The structure of the pyridoxamine 5'-phosphate (vitamin B6) dependent enzyme Aspartate aminotransferase (AspAT) from *E. coli* at 2.2 Å resolution (PDB code 1AMR) (Miyahara 1994) is used as the template. Its fold is a unique aminotransferase fold. AspAT is active as a homodimer and the calculations are performed on the dimer structure.

Using MODELLER software (Fiser 2000; Sanchez

Table 1

Summary of Orthologous Model Structures and Results

Enzyme	Species	% Identity	RMSD	THEMATICS Result
TIM	*S. japonicum*	60%	0.16	[**H95, E165,** C126, Y164]
	E. faecalis	40%	0.29	[**H95, E165,** C126]
	B. hensalae	39%	0.31	[**H95, E165,** C126, Y164]
	M. genitalium	33%	1.73	[**H95, E165,** C126]
	Template structure:	Chicken (1TPH)		[**H95, E165,** C126, Y164]
HPPK	*V. vulnificus*	63%	0.34	[**D95, D97, H115**]
	V. parahaemolyticus	57%	0.22	[**D95, D97, H115**]
	Ps. aeruginosa	51%	0.36	[**D95, D97, H115**]
	Ps. putida	48%	0.50	[**D95, D97, H115**]
	Template structure:	E. coli (1HKA)		[**D95, D97, H115**]
AspAT	*V. cholerae*	62%	1.52	[**H189, Y225, K258**, C191, C192]
	Oryza sativa	44%	0.64	[**H189, Y225, K258, R266**, C191, Y295]
	N. meningitides	41%	1.28	[**H189, Y225, K258,** C191, C192]
	C. perfringens	22%	3.67	[**H189, Y225, K258, R266**]
	Template structure:	E. coli (1AMR)		[**H189, Y225, K258, R266**, C191, C192, Y256]

Table 1: For each model structure, % pairwise identity with the template and the RMSD value in Å are given. THEMATICS results for the active site cluster are given with known active site residues shown in boldface and second shell residues underlined. Sequence numbers for the models are adjusted to match those of the template structures.

2000), four model structures homologous to the AspAT template from *E. coli* are constructed from the sequences for the following organisms: *Vibrio cholera* (Heidelberg 2000) (with 62 % pairwise identity and 1.52 Å RMSD), *Oryza sativa* (Sasaki 2001) (with 44% identity and 0.64 Å RMSD), and *Neiserria meningitides* (Tetelin 2000) (41 % identity and 1.28 Å RMSD), and *Clostridium perfringens* (22% identity and 3.67 Å RMSD).

Figure 1 illustrates the predicted titration curves for some of the lysine residues of the template structure for *E. coli*. Predicted mean net charge as a function of pH is shown for all of the lysine residues in the sequence between 144 and 355 (inclusive) of the A chain of the homodimer. The titration curves are given for K144 (+), K215 (×), K248 (*), K258 (□), K288 (■), K344 (○), and K355 (●). Note the elongated, highly non-sigmoidal shape of the catalytic lysine residue K258. The other lysine residues all have sigmoidal or nearly sigmoidal shape, with the characteristic sharp fall-off in charge in the region where the charge is approximately equal to one-half.

For the template *E. coli* structure, THEMATICS finds a cluster of residues with perturbed titration behaviour, consisting of the following: H189, Y225, K258, R266, C191 and C192. It has been established experimentally that K258 is the catalytic base that initiates transamination, that H189, Y225 and R266 are also in the active site pocket, and that nearby residues outside the active site, such as C191, also play a role in the catalytic activity (Miyahara 1994; Jeffery 1998; Jeffery 2000; Mizuguchi 2001). When sequential alignment is performed on all of the model sequences and the *E. coli* sequence, the identified residues are all conserved, except that C191 and C192 are not present in *Clostridium perfringens*. Superposition of the model structures onto

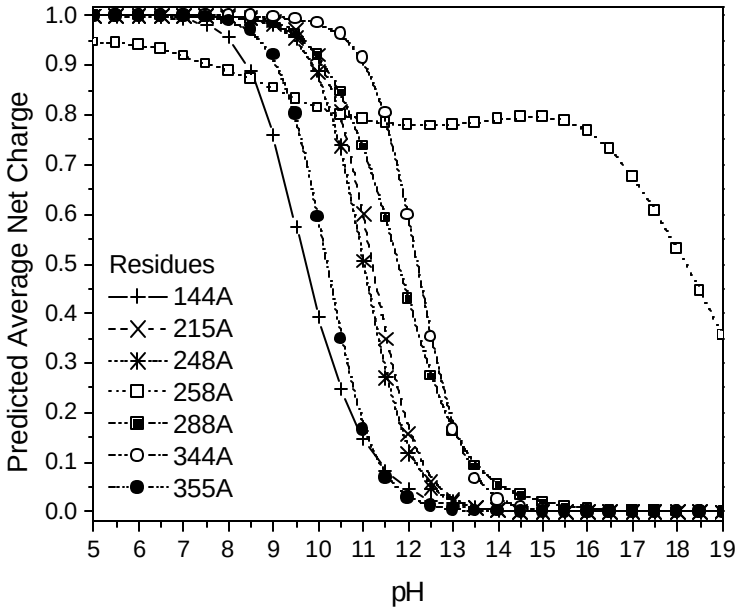

Figure 1: Predicted average net charge as a function of pH for the lysine residues from 144 through 355 of the A chain of the template structure for aspartate aminotransferase (1AMR). K144 (+), K215 (´), K248 (*), K258 (□), K288 (■), K344 (○), K355 (●).

---

the template structure reveal that the identified residues are located in the same region of the protein with similar orientations.

For all four models and for the template, THEMATICS finds the active site cluster, although the list of identified residues is a little different for each species for the AspAT case (see Table 1).

2.4 Human homologues of aldose reductase.

Aldose reductase is an NADH-dependent enzyme that catalyses the reduction of the aldehyde group in an aldose to an alcohol. It has the α/β barrel ("TIM barrel") fold and is active as a monomer. The x-ray crystal structure data are obtained from the protein data bank (PDB code 2ACS) with 1.76 Å resolution (Harrison 1994).

Homologous structures to human aldose reductase are built using SWISS-MODEL (Schwede 2003). Model structures are constructed using homologous protein sequences from the human and the template structure 2ACS for aldose reductase. These model protein structures all have the same fold as aldose reductase but they perform different functions and catalyse different reactions. The following model protein structures have been constructed: aldehyde reductase (Wermuth 1987), (with 50 % pair wise identity), bile acid binding protein (Stolz 1993) (with 48% identity), 3-oxo-5-beta steroid dehydrogenase (Kondo 1994) (50% identity), and chlordecone reductase (Qin 1993) (48% identity). The x-ray structures for aldehyde reductase (2ALR) and bile acid binding protein (1IHI) are available from the PDB and were used to check our results for the model structures.

THEMATICS computation on the template aldose reductase structure identifies the following cluster of residues as important: C298, H110, K77, Y48, Y209, E185 and K21. C298, H110, K77, Y48, Y209 are known active site residues while E185 and K21 are just behind the active site in the second shell surrounding the reacting substrate (Harrison 1994). Results are summarized in Table 2. For the models, the residues that occupy equivalent positions in the structure are vertically aligned in Table 2. Only the residues that are predicted to have perturbed titration behaviour are shown. As these enzymes have different functions, not all of the residues in the THEMATICS active site cluster are conserved, although K77, Y48 and E185 are conserved.

THEMATICS correctly locates the active site cluster for these human homologues of aldose reductase, in spite of

Table 2

Human Homologues of Aldose Reductase

Enzyme / % Identity	**THEMATICS result**						
Template:							
Aldose reductase	**[C298,**	**H110,**	**K77,**	**Y48,**	**Y209,**	E185,	K21]
Models:							
Aldehyde reductase / 50%	Y297	H112	K79	Y49	Y209	E185	K22
Bile acid binding protein / 48%	np[1]	E117[5]	K84	Y55	Y216	E192	K27
3-oxo-5-beta steroid dehydrogenase / 50%	np[2]	H113	K80	Y51	Y212	E188	np[3]
Chlordecone reductase / 48%	Y304	H116	K83	Y54	H215	E191	np[4]

Table 2: THEMATICS results are given for human aldose reductase and four human homologues with different chemical functions. Residues occupying the same position in the structure are aligned vertically in the table. In the template, known active site residues are shown in boldface; second shell residues are underlined. Only perturbed residues (those identified by THEMATICS) are shown.

np = the residue in that position is not perturbed and is not identified by THEMATICS;

np[1] = R304; np[2] = Y301; np[3] = E24; np[4] = E27

[5] There is sequence variability at this position. E117 is sometimes H117. Even when E is in this position, it is a THEMATICS positive.

---

some sequence variability in the active site and in spite of differences in function. Note that residues in equivalent positions are sometimes identified by THEMATICS even when there is amino acid substitution at that location.

It is also instructive to compare the THEMATICS results for the model structures of aldehyde reductase and bile acid binding protein with the results for the experimentally determined x-ray crystal structures. For aldehyde reductase, five out of the seven residues identified for the model structure are also identified for the crystal structure (H112, K79, Y49, Y209 and E185); two of the seven (K22 and Y297) are identified for the model structure but not for the crystal structure. For bile acid binding protein, the model structure and the crystal structure both identify five active site cluster residues: E117, K84, Y55, Y216, and E192. K27 is identified for the model structure but not for the crystal structure. Both structures give negative results for R304 (which occupies the position corresponding to C298 in aldose reductase).

3 Discussion and conclusions

For the model structures studied, THEMATICS successfully locates the active site. In some cases, there is some minor variation in the list of important residues, but the catalytically active residues always seem to be properly identified. The present findings are significant because they broaden the scope of proteins that can be studied with THEMATICS.

Comparison of results for model structures with the experimentally determined crystal structures for aldehyde reductase and for bile acid binding protein reveals some minor variations in the list of identified residues, but the ability to locate the active site from the model structure nevertheless is clear for the cases studied here.

The present results demonstrate that THEMATICS can work effectively on structures built from comparative modelling. The next step is to establish how effective is THEMATICS on structures obtained from threading. As fold libraries become more extensive, the reliability of threaded structures improves. The opportunity is emerging to predict information on chemical function from the genome sequence using a combination of theoretical model structures and THEMATICS.

4 Acknowledgments

This work was supported by the National Science

Foundation under grant MCB-0135303 and by the Institute for Complex Scientific Software at Northeastern University.

5 References

Antosiewicz, J., Briggs, J.M., Elcock, A.H., Gilson, M.K. , McCammon, J.A. (1996). "Computing the Ionization States of Proteins with a Detailed Charge Model." J. Comp. Chem. **17**: 1633-1644.

Bashford, D. and K. Gerwert (1992). "Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin." J Mol Biol **224**(2): 473-86.

Bashford, D. and M. Karplus (1991). "Multiple-site Titration Curves of Proteins: An Analysis of Exact and Approximate Methods for Their Calculation." J. Phys. Chem. **95**: 9556-9561.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-42.

Beroza, P., Fredkin, D. R., Okamura, M. Y., Feher, G. (1995). "Electrostatic calculations of amino acid titration and electron transfer, Q-AQB-->QAQ-B, in the reaction center." Biophys J **68**(6): 2233-50.

Carlson, H. A., J. M. Briggs, et al. (1999). "Calculation of the pKa values for the ligands and side chains of Escherichia coli D-alanine:D-alanine ligase." J Med Chem **42**(1): 109-17.

Fiser, A., R.K. Do, and A. Sali (2000). "Modeling of loops in protein structures." Protein Sci **9**: 1753-1773.

Fraser, C. M., J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kerlavage, G. Sutton, J.M. Kelley, et al. (1995). "The minimal gene complement of Mycoplasma genitalium." Science **270**: 397-403.

Gilson, M. K. (1993). "Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins." Proteins **15**(3): 266-82.

Harrison, D. H., K.M. Bohren, D. Ringe, G.A. Petsko, K.H. Gabbay (1994). "An anion binding site in human aldose reductase: Mechanistic implications for the binding of citrate, cacodylate and glucose-6-phosphate." Biochemistry **33**: 2011-2020.

Heidelberg, J. F., J.A. Eisen, W.C. Nelson, R.A. Clayton, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, L.A. Umayam, et al. (2000). "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae." Nature **406**: 477-483.

Jeffery, C. J., L.M. Gloss, G.A. Petsko, D. Ringe (2000). "The Role of Residues Outside the Active Site: Structural Basis for Function of C191 Mutants of Escherichia coli Aspartate Aminotransferase." Protein Engineering **13**: 105-112.

Jeffery, C. J., T. Barry, S. Doonan, G.A. Petsko, D. Ringe (1998). "Crystal Structure of Saccharomyces cerevisiae Cytoso lic Aspartate Aminotransferase." Protein Sci **7**: 1380-1387.

Kondo, K. H., M.H. Kai, Y. Setoguchi, G. Eggertsen, P. Sjoeblom, T. Setoguchi, K.I. Okuda, I. Bjoerkhem (1994). "Cloning and expression of cDNA of human delta 4-3-oxosteroid 5 beta-reductase and substrate specificity of the expressed enzyme." Eur J Biochem **219**: 357-363.

Liu, F., W. Hu, Q. Yan, X. Xu, Z. Zhu, X. Zhang, J. Wang, J. Sun, Z. Wang, L. Zeng, et al. (2003). "The full-length cDNA of S. japonicum genes." EMBL GenBank DDBJ databases.

Lodi, P. J., J.R. Knowles (1991). "Neutral imidazole is the electrophile in the reaction catalyzed by triosephosphate isomerase: Structural origins and catalytic implications." Biochemistry **30**: 6948-6956.

Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, et al. (2003). "Genome sequence of Vibrio parahaemolyticus: A pathogenic mechanism distinct from that of V. cholerae." Lancet **361**: 743-749.

Miyahara, I., K. Hirotsu, H. Hayashi, H. Kagamiyama (1994). "X-ray Crystallographic Study of Pyridoxamine 5'-Phosphate-Type Aspartate Aminotransferases from Escherichia coli in Three Forms." J. Biochem. (Tokyo) **116**: 1001.

Mizuguchi, H., H. Hayashi, K. Okada, I. Miyahara, K. Hirotsu, H. Kagamiyama (2001). "Strain Is More Important than Electrostatic Interaction in Controlling the pKa of the Catalytic Group in Aspartate Aminotransferase." Biochemistry **40**: 353-360.

Nelson, K. E., C. Weinel, I.T. Paulsen, R.J. Dodson, H.Hilbert, V.A.P. Martins dos Santos, D.E. Fouts, S.R. Gill, M. Pop, M. Holmes, et al. (2002). "Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440." Environ. Microbiol. **4**: 799-808.

Ondrechen, M. J., J. G. Clifton, et al. (2001). "THEMATICS: A simple computational predictor of enzyme function from structure." Proc. Natl. Acad. Sci. (USA) **98**: 12473-12478.

Paulsen, I. T., L. Banerjei, G.S.A. Myers, K.E. Nelson, R. Seshadri, T.D. Read, D.E. Fouts, J.A. Eisen, S.R. Gill, J.F. Heidelberg, et al. (2003). "Role of

mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis." Science **299**: 2071-2074.

Qin, K. N., M.I. New, K.C. Cheng (1993). "Molecular cloning of multiple cDNAs encoding human enzymes structurally related to 3 alpha-hydroxysteroid dehydrogenase." J. Steroid Biochem. Mol. Biol. **46**: 673-679.

Rhee, J. H., S.Y. Kim, S.S. Chung, J.J. Kim, Y.H. Moon, H. Jeong, and H.E. Choy (2002). "Complete genome sequence of Vibrio vilnificus." EMBL GenBank DDBJ databases.

Sali, A. (1995). "Modeling mutations and homologous proteins." Curr Opin Biotechnol **6**(4): 437-51.

Sali, A. (1998). "100,000 protein structures for the biologist." Nature Struct Biol **5**: 1929-1932.

Sampogna, R. V. and B. Honig (1994). "Environmental effects on the protonation states of active site residues in bacteriorhodopsin." Biophys J **66**(5): 1341-52.

Sanchez, R., A. Sali (2000). "Comparative protein structure modeling. Introduction and practical examples with modeller." Methods Mol. Biol. **143**: 97.

Sasaki, T., T. Matsumoto, K. Yamamoto (2001). "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC clone." EMBL GenBank DDBJ databases **FEB-2001**.

Schwede, T., J. Kopp, N. Guex, M.C. Peitsch (2003). "SWISS-MODEL: An automated protein homology-modeling server." Nucleic Acids Res **31**: 3381-3385.

Shehadi, I. A., H. Yang and M.J. Ondrechen (2002). "Future directions in protein function prediction." Mol Biol Reports **29**: 329-335.

Stammers, D. K., A. Achari, D.O. Somers, P.K. Bryant, J. Rosemond, D.L. Scott, J.N. Champness (1999). "2.0A X-ray structure of the ternary complex of 7,8-Dihydro-6-Hydroxymethylpterinpyrophosphokinase from Escherichia coli with ATP and a substrate analogue." FEBS Lett **456**: 49.

Stolz, A., L. Hammond, H. Lou, H. Takikawa, M. Ronk, J.E. Shively (1993). "cDNA cloning and expression of the human hepatic bile-acid binding protein. A member of the monomeric reductase gene family." J. Biol. Chem. **268**: 10448-10457.

Stover, C. K., et al. (2000). "Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen." Nature **406**: 959-964.

Tetelin, H., N.J. Saunders, J. Heidelberg, A.C. Jeffries, K.E. Nelson, J.A. Eisen, K.A. Ketchum, D.W. Hood, J.F. Peden, R.J. Dodson, et al. (2000). "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58." Science **287**: 1809-1815.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Wermuth, B., A. Omar, A. Forster, C. di Francesco, M. Wolf, J.P. von Wartburg, B. Bullock, K.H. Gabbay (1987). "Primary structure of aldehyde reductase from human liver." Prog. Clin. Biol. Res. **232**: 297-307.

Westbrook, J., Z. Feng, L. Chen, H. Yang, and H.M. Berman (2003). "The Protein Data Bank and structural genomics." Nucleic Acids Res **31**: 489-491.

Xiao, B., G. Shi, X. Chen, H. Yan, X. Ji (1999). "Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents." Structure Fold Des **7**: 489-496.

Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K., Honig, B. (1993). "On the calculation of pKas in proteins." Proteins **15**(3): 252-65.

Zhang, Z., Sugio, S., Komives, E. A., Liu, K. D., Knowles, J. R., Petsko, G. A., Ringe, D. (1994). "Crystal structure of recombinant chicken triosephosphate isomerase-phosphoglycolohydroxamate complex at 1.8-A resolution." Biochemistry **33**(10): 2830-7.