HUMAN MUTATION 17:263–270 (2001)

RESEARCH ARTICLE

SNPs, Protein Structure, and Disease

Zhen Wang and John Moult*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland

For the SNP 2000 Special Issue

Inherited disease susceptibility in humans is most commonly associated with single nucleotide polymorphisms (SNPs). The mechanisms by which this occurs are still poorly understood. We have analyzed the effect of a set of disease-causing missense mutations arising from SNPs, and a set of newly determined SNPs from the general population. Results of in vitro mutagenesis studies, together with the protein structural context of each mutation, are used to develop a model for assigning a mechanism of action of each mutation at the protein level. Ninety percent of the known disease-causing missense mutations examined fit this model, with the vast majority affecting protein stability, through a variety of energy related factors. In sharp contrast, over 70% of the population set are found to be neutral. The remaining 30% are potentially involved in polygenic disease. Hum Mutat 17:263–270, 2001. © 2001 Wiley-Liss, Inc.

KEY WORDS: SNP; missense mutation; protein structure; disease; modeling; structural biology

DATABASES:
http://www.SNPS3D.org (SNP Project);http://www.ncbi.nlm.nih.gov/SNP (dbSNP/NCBI); http://waldo.wi.mit.edu/cvar_snps/ (Whitehead/MIT cSNP Data); http://genome.cwru.edu/candidates/candidates.html (Case-Western Reserve University); http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html (HGMD)

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variations in humans, accounting for about 90% of sequence differences [Collins et al., 1998] at an overall frequency of about one per 1000 bases [Taillon-Miller et al., 1998]. New techniques for the large-scale identification of SNPs in the human population [Wang et al., 1998] are resulting in an exponential expansion in the number known, and already the NIH SNP database (http://www.ncbi.nlm.nih.gov/SNP) contains approximately 2.8 million cases. There are high hopes that knowledge of an individual's SNP genotype will provide a basis for assessing susceptibility to disease and the optimal choice of therapies [Masood, 1999]. A major challenge in realizing these expectations is understanding how and when the variants cause disease. SNPs in coding regions (cSNPs) and regulatory regions are most likely to affect gene function [Collins et al., 1997; Chakravarti, 1998; Syvanen et al., 1999]. The first systematic studies of SNPs in sets of human genes show that each gene contains an average of about four cSNPs with a frequency of greater than about 1% in the population.[†] About half of cSNPs cause missense mutations (missense cSNPs) in the corresponding proteins, while the other half are silent [Cargill et al., 1999; Halushka et al., 1999]. Missense cSNPs may be neutral, that is, not detectably altering the phenotype, or they may result in differences in individual characteristics, different responses to drugs, or different susceptibility to

Received 1 December 2000; accepted revised manuscript 19 January 2001.

*Correspondence to: John Moult, CARB, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: jmoult@tunc.org

[†]Some authors suggest the formal definition of a SNP requires a frequency in the population of greater than 1% [Brookes, 1999]. Others, for example, the NIH SNP database, maintain a general definition. We follow the latter practice.

©2001 WILEY-LISS, INC.

disease. According to the Human Gene Mutation Database [Krawczak et al., 2000], missense mutations account for almost half of all DNA mutations that are known to cause genetic disease. Many of these are single causative factors of relatively rare monogenic inherited disorders. It is expected that some more frequent missense mutations arising from cSNPs will be associated with common polygenic diseases such as heart disease and hypertension [Brookes, 1999; Pratt and Dzau, 1999]. We have focused on developing a model of missense cSNP action that provides a basis for understanding the impact of a cSNP at the molecular level, and thus a means of predicting which cSNPs are potentially involved in disease. A related study has recently appeared [Sunyaev et al., 2000].

Which missense cSNPs will affect the phenotype? At the DNA level, linkage studies are a powerful means of detecting monogenic traits, but are more problematic for polygenic effects [Pratt and Dzau, 1999], and require SNP information from an appropriate population sample. At the amino acid sequence level, the missense mutations affecting function might be expected to be those producing the least conservative replacements. We find this correlation to be low (data not shown). At the protein structure level, the wealth of experimental data on the effect of site directed mutagenesis on structure and function makes it possible to predict the likely changes in molecular function from the structural context of each missense mutation produced by a cSNP.

It is the connection between SNPs and protein structure that we exploit in the present study. In vitro site directed mutagenesis studies, together with the data on the structural context of known disease-causing missense mutations are used to develop a model of missense cSNP functional impact. Each mutation is associated with an effect on one or more roles of the residue concerned. Roles that may be affected are: protein stability or folding, ligand binding, catalysis, regulation by allosteric and other mechanisms, post-translational modification. Some mutations are considered to be essentially neutral. The types of interactions that are changed are also identified: hydrophobic [Eriksson et al., 1992], hydrogen bond (H-bond) [Shirley et al., 1992], van der Waals [Xu et al., 1998], electrostatic interactions [Horovitz et al., 1990], and disulfide bonds [Betz, 1993]. Figure 1 shows some examples. These observations were used to devise rules that allow assignment of missense mutations to different categories and severity of effect. The rules may be applied to any missense mutation for which adequate information on the structural context is available. There are some rare effects, such as on transcription, translation, and some forms of post-translational modification, that are relatively poorly understood, and difficult to deduce from sequence or structure information. These will not be detected by the present model.

We have used the rules in the model to analyze the structural and functional effect of missense cSNPs in two classes of human proteins: One set (*SNP-disease*) contains a sample of proteins with clinically identified missense mutations, each of which is a cause of a monogenic inherited disorder. The other, termed the *SNP-population* set, contains proteins with missense mutations existing in the general population that are not known to cause any inherited disease. The *SNP-disease* set consists of 23 proteins containing 262 missense mutations. All homozygous missense mutations in this set are known to cause disease independently, a few are dominant. The *SNP-population* set consists of 22 proteins containing 42 missense mutations. All the proteins in both sets have either known three-dimensional (3-D) structures, or have 40% or higher sequence identity to a protein with known structure. Comparative models of the wild type structures were built for those proteins without experimentally known 3-D structures. Simple models of protein missense mutants were built, analyzed, and compared to protein wild type structures. The effects of missense mutations on protein function and stability were assigned using the rules in the model.

COMPUTATIONAL METHODS

Construction of the *SNP-Disease* Set

The 4360 SNPs in the NIH dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) as of May 1999 were subjected to a BLASTX [Altschul et al., 1990] search against the NR database, to identify those in open reading frames. The resulting set of proteins IDs was then used to query the Human

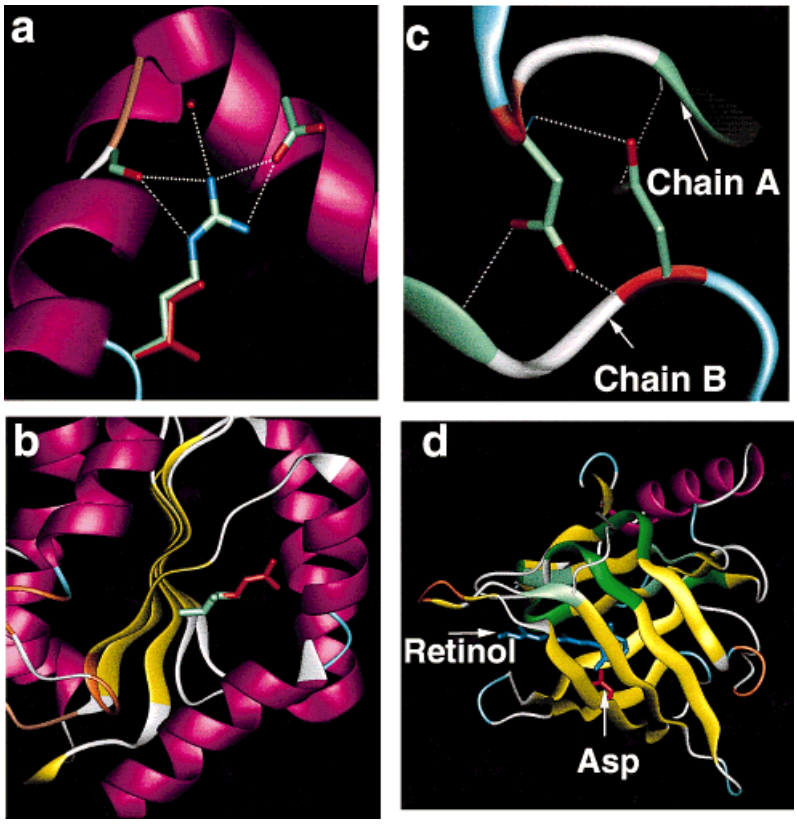

FIGURE 1. **Examples of the variety of ways in which the structural effects of missense mutations cause inherited monogenic disease. Four cases from the set of 262 mutations analyzed are shown. For each one, the model provides an explanation of the cause of disease, in three cases by destabilizing the folded state (83% of all cases fall in that category), and in the fourth interfering with ligand binding (~5% of all cases). Wild type side chains are colored by element, and mutation side chains are colored red (for Gly mutations, the backbone is shown in red). a: Loss of a salt bridge and hydrogen bonds in coagulation factor XIIIA, the last enzyme in the blood clot formation cascade, a transglutaminase that crosslinks fibrin and stabilizes clots [Board et al., 1990]. The wild type R252 forms a salt bridge and three hydrogen bonds with adjacent loops on the surface of the protein. These interactions are lost in the R252L mutant, destabilizing the folded structure, and resulting in poor clot formation (OMIM# 134570) [Mikkola et al., 1996]. b: Burying of a charged residue in aldehyde dehydrogenase 10. The mutation L106R introduces a positive charge into the hydrophobic core of the protein, destabilizing the folded structure. The mutant leads to an accumulation of long chain fatty alcohol intermediates, with associated myelination defects [van Domburg et al., 1999]. The phenotype is Sjogren-Larsson syndrome (OMIM# 270200) [Sillen et al., 1998], the primary symptoms of which are mental retardation and spasticity. c: Destabilization of the multimeric structure of aldolase A, a glycolytic pathway enzyme. The mutant D128G removes four hydrogen bonds between a pair of subunits, and introduces a higher entropy cost on folding, destabilizing the folded state. There is no detectable effect on most cell types. However, red blood cells exhibit weakened cell membranes, resulting in a phenotype of congenital nonspherocytic hemolytic anemia (OMIM# 103850) [Kishi et al., 1987]. Mature red blood cells are unable to manufacture additional protein, suggesting a compensating mechanism in other cell types. d: Hindrance of ligand binding. Retinol (vitamin A) binds in a hydrophobic cavity in the center of the beta barrel of retinol binding protein. The mutant G75D introduces a negative charge into the cavity, interfering with retinol binding both electrostatically and sterically. The result is vitamin A deficiency (OMIM# 180250), with a phenotype of night blindness [Seeliger et al., 1999].**

Genome Mutation Database, (HGMD, http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html). Sequences of the 557 proteins reported as containing one or more disease-causing missense cSNPs were then screened against the set of protein sequences in the protein data bank (PDB), using BLAST. A total of 157 proteins with a structure or a homologous structure with 25% or more sequence identity were found. A representative subset of 23 proteins was chosen randomly for this study.

Construction of *SNP-Population* Set

Proteins in the *SNP-population* set were identified by interrogating the Case Western hyper-

tension candidate genes database [Halushka et al., 1999] (http://genome.cwru.edu/candidates/candidates.html), containing 150 genes possibly involved in blood-pressure homeostasis; and the Whitehead cSNP Database [Cargill et al., 1999] (http://waldo.wi.mit.edu/cvar_snps/), containing 106 genes related to cardiovascular disease, endocrinology, or neuropsychiatry. Sequences were compared with the PDB to identify those having experimental structures or those that could be modeled with acceptable accuracy. Twenty-two proteins containing 42 missense mutations not so far associated with disease were found.

Structure Modeling

For the *SNP-disease* set, 18 proteins have known 3-D x-ray crystallography structures and five are homologs of known 3-D structures, with sequence identities between 42–77%. In the *SNP-population* set, 13 proteins have known 3-D structures and nine have homologs of known structure with sequence identities between 61–92%. All protein models were built automatically using a well established procedures [Marti-Renom et al., 2000] and an in-house comparative modeling program. Wild type protein structures were modeled as follows: Sequences were aligned using FASTA [Pearson and Lipman, 1988] and the main chain conformation was constructed by copying the conserved main chain coordinates of a template. There were no insertion or deletions between these targets and their templates. Side chains were added using SCWRL [Dunbrack and Cohen, 1997]. Mutant side chains were initially positioned using SCWRL and possible alternative side chain conformations were examined manually.

Rules for Assigning the Effects of Missense Mutations on Molecular Function

I. Protein stability. Protein stability is affected by one or more of the following: 1) Loss of one or more hydrogen bonds: A hydrogen bond is defined as a donor to acceptor distance of ≤2.5Å, and an angle at the acceptor ≥90.0°. 2) Reduced hydrophobic interaction: loss of burial of 50Å^2 or more of non-polar area on folding. 3) Loss of a salt bridge: A salt bridge is defined as at least one pair of atoms on oppositely charged groups within 4.5Å. 4) Buried charged residue: Introduction of a zero-accessibility electrostatically isolated charge. 5) Over-packing: Introduction of a larger side chain, making unavoidable short atomic contacts (<2.5 Å in the interior of a protein or <2.0 Å on the surface). 6) Internal cavity: Replacement of one or more zero-solvent accessibility groups with a smaller side chain. 7) Electrostatic repulsion: Introduction of a charged group, such that at least one pair of atoms on like-charged groups are unavoidably within 4.5Å. 8) Buried polar residue: One or more polar groups unavoidably have zero solvent accessibility and no hydrogen bonds. 9) Disruption of metal binding: A metal liganding group is replaced by a non-liganding group. 10) Breakage of a disulfide bond: A cysteine residue in an S-S bond is replaced by a non-cysteine residue. 11) Backbone strain: Gly is changed to any other residue, with phi/psi values not in the Ramachandran general sterically allowed region, or any other residue is changed to Pro, where phi is unfavorable. (Permitted phi for Pro = –60 ± 15°); or a cis Pro (omega = 0 ± 60°) is changed to any other residue. 12) Destabilization of a protein multimer: Any of the above rules, involving atoms on a neighboring subunit**.**

II. Ligand binding. Any of the above protein stability rules can apply, where ligand atoms interact with the mutated side chain.

III. Catalysis. The mutated residue is directly involved in a catalytic process, or is critical in positioning such a residue.

IV. Allosteric regulation. The mutated residue is involved in an allosteric mechanism.

V. Post-translational modification. An N-X-S/T sequence pattern for N-glycosylation (X = any residue except Pro) is disrupted.

RESULTS AND DISCUSSION

Ninety Percent of Missense Mutations in the *SNP-Disease* Set Fit the Model With the Vast Majority Affecting Protein Stability

A sample of results is shown Table 1. Full data are available at http://www.SNPS3D.org. The distribution of the effects of missense mutations on protein molecular function for the *SNP-disease* set is shown in Figure 2A. The large majority of these mutations (83%) are found to affect

TABLE 1. Sample Results of the Effect of Missense Mutations in *SNP-Disease* Set*

Gene name, protein name	Missense mutation	Structural effect	Molecular function effect	Phenotypic effect (disease)
RBP4 Retinol (vitamin A) binding protein	I41N	Disruption of hydrophobic interaction	Ligand binding	Vitamin A deficiency: night blindness
	G75D	Charged residue buried	Ligand binding	
PPGB Protective protein for beta galactosidases	Q21R	Over-packing	Protein stability	Lysosomal storage disease: galactosialidosis
	S23Y	Over-packing	Protein stability	
	W37R	Charged residue buried	Protein stability	
	S62L	Over-packing	Protein stability	
	V104M	Over-packing	Protein stability	
	L208P	Loss of H-bond	Protein stability	
	Y221N	No effect	No effect	
	Y367C	Disruption of hydrophobic interaction	Protein stability	
	G411S	Over-packing	Protein stability	
	F412V	Multimer destabilization	Protein stability	
	M378T	New glycosylation site	Post-translational modification	
ALDOB Aldolase B, fructose-bisphosphate aldolase	C134R	Charged residue buried	Protein stability	Hereditary fructose intolerance
	W147R	Charged residue buried	Protein stability	
	A149P	Backbone strain	Protein stability	
	A174D	Charged residue buried	Protein stability	
	L256P	Loss of H-bond	Protein stability	
	N334K	Charged residue buried	Protein stability	
	R303W	Over-packing	Ligand binding	

*Mutations known to be causes of monogenic inherited disorders.

protein stability. The predominance of stability effects is surprising, for two reasons. First, significant effects on function would be expected from mutations of residues directly involved in binding and catalysis, yet only 5% of mutations implicate residues in these categories. A likely explanation is that a much larger fraction of residues are involved in stability than any other role. Second, many of the mutations that affect stability would not be expected to have a significant effect on function, as judged by in vitro experiments. These missense mutations appear to decrease stability by approximately 1, 2, or 3 Kcal/mol range [Horovitz et al., 1990; Eriksson et al., 1992; Shirley et al., 1992; Xu et al., 1998], not enough to seriously destabilize a folded state with a typical free energy of about 10 Kcal/mol [Creighton, 1993], and they have no direct impact on binding, catalysis, or regulation. Proteins that traffic through the endoplasmic reticulum (ER) may be trapped and eliminated by mechanisms that detect excessive fluctuations [Ellgaard et al., 1999]. However, comparison of our results for proteins that pass through the ER and those that do not show any significant difference.

Effects on stability are distributed over all

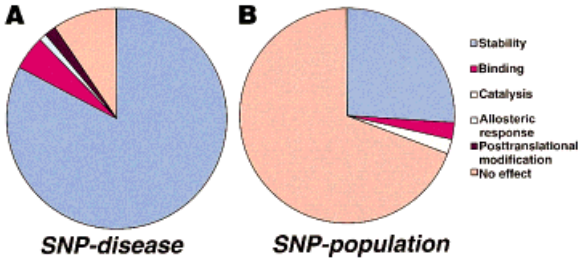

FIGURE 2. **Analysis of the effect of missense cSNPs on protein function. A: A set of 262 disease-causing missense cSNPs. The pie chart shows the fraction of these mutations that are assigned to different classes of functional effect, using the model described in this paper. By far the largest fraction (83%) affects protein stability in some way. The next largest category is effects on ligand binding, for only 5% of the mutations. Ten percent of the mutations could not be assigned any functional effect with the model. B: A set of missense cSNPs in the human population. The pie chart shows the breakdown of the effects found in missense mutations in two sets of proteins. One set is from genes that may be associated with blood-pressure homeostasis and hypertension [Halushka et al., 1999], and the other set from genes that may be associated with cardiovascular disease, endocrinology, or neuropsychiatry [Cargill et al., 1999]. Seventy percent of these mutations are assigned no effect by the model. The other 30% almost all have effects on stability. A striking exception is a mutation in heart chymase, replacing the histidine forming a part of the serine protease catalytic triad. This and the others causing instability are candidates for previously unidentified polygenic disease-associated mutations.**

TABLE 2. The Set of New Potential Disease-Associated Mutations Identified in This Study

Gene name, protein name	Missense mutation	Structural effect	Molecular function effect	Database
ANX3, inositol 1,2-cyclic phosphate 2-phosphohydrolase	I220N	New H-bonds	Protein stability	Whitehead
	P252L	Over-packing	Protein stability	cSNP
PAI2, plasminogen activator inhibitor-2	S413C	Loss of H-bonds	Protein stability	Whitehead cSNP
HMGCR HMG-CoA reductase	I638V	Loss of hydrophobic interaction	Protein stability	Whitehead cSNP
ICAM2, intercellular adhesion molecule-2	A37T	New H-bonds	Protein stability	Case Western
	R199H	Loss of salt bridge	Protein stability	confirmed by sequencing
ALDR1 Aldehyde reductase	I14F	Over-packing	Protein stability	Case Western confirmed by sequencing
COX2 Cyclooxygenase-2	E488G	Loss of salt bridge	Protein stability	Case Western
	V511A	Loss of hydrophobic interaction	Protein stability	not confirmed by sequencing
COX1 Cyclooxygenase-1	R108Q	Loss of H-bonds	Protein stability	Case Western not confirmed by sequencing
ELAM Human E-selectin	C130W	Loss of disulfide bond	Protein stability	Case Western not confirmed by sequencing
GH2 Growth hormone	R90W	Electrostatic interactions	Ligand binding	Case Western not confirmed by sequencing
CYH Human chymase	H66R	Catalytic triad	Catalysis	Case Western not confirmed by sequencing

possible causes, with over-packing (22%), loss of H-bonds (21%) (approximately a quarter of these lost H-bonds involve a charged group, and about a quarter are main chain to main chain), backbone strain (19%), and buried charged residues (14%) the most common.

Only 10% of disease-causing mutations are not assigned an effect on function by the model. In some of these cases, the model is probably unable to correctly predict the effect of a missense mutation on protein structure and function. In others, unknown molecular function may be affected; for example a mutation may occur in an unidentified macromolecular binding site. An implication of the high percentage of missense mutations fitting the model is that most inherited monogenic disease caused by missense mutations is a consequence of straightforward defects in protein stability and function.

Seventy Percent of the Missense Mutations in the *SNP-Population* Set Are Neutral

A summary of the effects of the missense mutations in the *SNP-population* set is shown in Figure 2B. In sharp contrast to those in the *SNP-disease* set, 70% of these mutations have no effect on protein function, according to the model. This is consistent with the idea that most accepted polymorphisms in the population are essentially neutral [Kimura and Takahata, 1983]. The missense mutations with apparent effects on function are again dominated by stability effects, with only one exception, involving catalysis. A small set of new potentially common disease-associated cSNPs is identified (Table 2). Some of these may be experimental false positives (estimated at around 21% of SNPs in the Case Western data [Halushka et al., 1999], but much lower in the Whitehead set [Cargill et al., 1999], and await confirmation by sequencing. Some may have an effect at the molecular level but no significant phenotypic effect for one of number of reasons, such as functionally redundant protein domains, overlapping protein functions [Labow et al., 1994], alternative pathways, compensating feedback mechanisms, or simply a minor role for the protein in the viability of the organism. Some are expected to represent previously unknown common disease-related SNPs, probably requiring

other mutations in other genes to have a detectable phenotypic effect. That is, they may contribute to a common polygenic disease.

CONCLUSIONS

These results show that a simple model of the effects of cSNPs on protein function is able to distinguish between those that affect molecular function and those that do not, and provide insight into the mechanisms by which missense cSNPs cause disease. There are some important caveats, of course. First, there is a false negative rate of around 10% in *SNP-disease* set, arising from effects that are not easily identified from structure alone. Second, there is some false positive rate, arising partly from the fact that the model identifies effects on the molecular function of a protein, rather than in the phenotype. Third, the model is based on the assumption that all disease arising from missense mutations is caused by defects in proteins.

ACKNOWLEDGMENTS

We thank Eugene Melamud for extensive assistance with data handling and analysis, and Dr. Michael Gilson, Dr. Osnat Herzberg, Dr. Arlin Stalzfus, and Dr. Ron Unger for many constructive comments on the initial manuscript.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Mol Biol 215:403–410.

Betz S. 1993. Disulfide bonds and the stability of globular proteins. Protein Sci 2:1551–1558.

Board PG, Pierce K, Coggan M. 1990. Expression of functional coagulation factor XIII in Escherichia coli. Thromb Haemost 12:235–240.

Brookes AJ. 1999. The essence of SNPs. Gene 234:177–186.

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 22:231–238.

Chakravarti A. 1998. It's raining SNPs, hallelujah? Nat Genet 19:216–217.

Collins FS, Guyer MS, Charkravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581.

Collins FS, Brooks LD, Chakravarti A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 8:1229–1231.

Creighton TE. 1993. Proteins, structures and molecular properties. Proteins in solution and in membranes: stability of the folded conformation. New York: W.H. Freeman and Company. p 299.

Dunbrack RL Jr, Cohen FE. 1997. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci 6:1661–1681.

Ellgaard L, Molinari M, Helenius A. 1999. Setting the standards: quality control in the secretory pathway. Science 286:1882–1888.

Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Blaber M, Baldwin EP, Matthews BW. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. Science 255:178–183.

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N. 1999. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 22:239–247.

Horovitz A, Serrano L, Avron B, Bycroft M, Fersht AR. 1990. Strength and co-operativity of contributions of surface salt bridges to protein stability. J Mol Biol 216:1031–1044.

Kimura M, Takahata N. 1983. Selective constraint in protein polymorphism: study of the effectively neutral mutation model by using an improved pseudosampling method. Proc Natl Acad Sci 80:1048–1052.

Kishi H, Mukai T, Hirono A, Fujii H, Miwa S, Hori K. 1987. Human aldolase A deficiency associated with a hemolytic anemia: thermolabile aldolase due to a single base mutation. Proc Natl Acad Sci USA 84:8623–8627.

Krawczak M, Ball EV, Fenton I, Stenson PD, Abeysinghe S, Thomas N, Cooper DN. 2000. Human gene mutation database—a biomedical information and research resource. Hum Mutat 15:45–51.

Labow MA, Norton CR, Rumberger JM, Lombard-Gillooly KM, Shuster DJ, Hubbard J, Bertko R, Knaack PA, Terry RW, Harbison ML. 1994. Characterization of E-selectin-deficient mice: demonstration of overlapping function of the endothelial selectins. Immunity 1:709–720.

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. 2000. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29:291–325.

Masood E. 1999. As consortium plans free SNP map of human genome. Nature 398:545–546.

Mikkola H, Yee VC, Syrjala M, Seitz R, Egbring R, Petrini P, Ljung R, Ingerslev J, Teller DC, Peltonen L, Palotie A. 1996. Four novel mutations in deficiency of coagulation factor XIII: consequences to expression and structure of the A-subunit. Blood 87:141–151.

Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448.

Pratt RE, Dzau VJ. 1999. Genomics and hypertension. Concepts, potentials and opportunities. Hypertension 33:238–245.

Seeliger MW, Biesalski HK, Wissinger B, Gollnick H, Gielen S, Frank J, Beck S, Zrenner E. 1999. Phenotype in retinol deficiency due to a hereditary defect in retinol binding protein synthesis. Invest Ophthalmol Vis Sci 40:3–11.

Shirley BA, Stanssens P, Hahn U, Pace CN. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. Biochem 31:725–731.

Sillen A, Anton-Lamprecht I, Braun-Quentin C, Kraus CS, Sayli BS, Ayuso C, Jagell S, Kuster W, Wadelius C. 1998. Spectrum of mutations and sequence variants in the FALDH gene in patients with Sjogren-Larsson syndrome. Hum Mutat 12:377–384.

Sunyaev S, Ramensky V, Bork P. 2000. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet 16:198–200.

Syvanen AC, Landegren U, Isaksson A, Gyllensten U, Brookes AJ. 1999. Enthusiasm mixed with skepticism about single-nucleotide polymorphism markers for dissecting complex disorders. Eur J Hum Genet 7:98–101.

Taillon-Miller P, Gu Z, Li Q, Hillier L, Kwok PY. 1998. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. Genome Res 8:748–754.

van Domburg PH, Willemsen MA, Rotteveel JJ. 1999. Sjogren-Larsson syndrome: clinical and MRI/MRS findings in FALDH-deficient patients. Neurology 52:1345–1352.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES, Lipshutz R, Chee M. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082.

Xu J, Baase WA, Baldwin E, Matthews BW. 1998. The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. Protein Sci 7:158–177.