

# Three-Dimensional Structural Location and Molecular Functional Effects of Missense SNPs in the T Cell Receptor V $\beta$ Domain

Zhen Wang<sup>1,2</sup> and John Moulton<sup>1\*</sup>

<sup>1</sup>Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland

<sup>2</sup>Molecular and Cell Biology Graduate Program, University of Maryland, College Park, Maryland

**ABSTRACT** The mechanisms by which human single nucleotide polymorphisms (SNPs) influence susceptibility to disease are not yet well understood. In a previous study, we developed a structure-based model that may be used to identify which missense SNPs are neutral and which are deleterious to protein function and so potentially involved in disease (Wang and Moulton, *Hum Mutat* 2001;263–270). The model has now been applied to a set of 54 missense cSNPs in the 46 functional T-cell receptor V $\beta$ -genes. Most of these missense cSNPs are found to be neutral, but 10 are identified as likely deleterious to protein function. Only one was previously associated with disease. We suggest that the others may be disease related but that redundancy in the T-cell response prevents any simple, monogenic effect. Therefore, these SNPs are the most likely contributors to complex, polygenic disease traits. It has been noted that there is a surprisingly high (74%) fraction of nonsynonymous SNPs in these genes. Contrary to expectation, the analysis shows that these are not associated with an unusually high fraction of deleterious SNPs, nor do they significantly contribute to a larger range of antigen recognition or a reduced superantigen-binding repertoire. *Proteins* 2003;53:748–757. © 2003 Wiley-Liss, Inc.

**Key words:** T-cell receptor; human SNPs; inherited disease; protein structure

## INTRODUCTION

Single nucleotide polymorphisms (SNPs) are the most common form of genetic difference between individuals, and are believed to be responsible for most inherited traits, including a large fraction of inherited disease susceptibility. The link between a single nucleotide change and monogenic disease has been explored in detail for some cases, and there are about 1000 proteins that are known to be associated with this process.<sup>1</sup> The large number of human SNPs that have now been identified [in excess of 4 million unique SNPs (<http://www.ncbi.nlm.nih.gov/SNP/index.html>)], together with knowledge of the genome sequence and other proteome information, should open the way for a much broader understanding of the link between genotype and phenotype at this level. Nevertheless, so far the detailed mechanisms by which a SNP may cause a phenotypic difference are largely unknown. In particular, much disease susceptibility is believed to be inherited as a

complex trait: The combination of SNPs in a number of genes affect the likelihood of an individual developing a particular disease, such as Alzheimer, asthma, diabetes, and many others.<sup>2</sup> Despite intense effort,<sup>3</sup> it has so far not been possible to track these correlations for any disease.

SNPs may affect the function of a protein through a variety of mechanisms. However, all but 2% of known single nucleotide variants associated with monogenic disease are nonsynonymous SNPs in protein-coding regions (i.e., SNPs that change a single amino acid in a protein molecule).<sup>1</sup> Thus, it is expected that this class of SNPs will be very relevant to complex inherited disease traits as well. Recent work has established that for these SNPs, insight into functional impact can be provided by analysis of protein sequence profiles and the structural environment of the affected amino acid.<sup>4–9</sup> In previous work,<sup>4</sup> we developed a structure-based model that determines the likely impact of these nonsynonymous SNPs in protein-coding regions. A preliminary analysis of a small set of SNPs found in the general population and not known to be associated with disease suggested that about 30% of these do significantly damage protein function.

Although there are many SNPs known throughout the genome, there are so far rather few sets of genes where all the SNPs down to some defined level of penetration in the population have been identified. One set of genes for which this has been done is the T-cell receptor (TCR) V $\beta$ -genes.<sup>10</sup> We have now applied the SNP model to these data to determine how the nonsynonymous cSNPs affect protein structure and function. The analysis addresses the issues of what fraction of these SNPs are deleterious to TCR function, what aspects of function are affected, and the possible causes of the very high fraction of nonsynonymous SNPs in these genes.

Antigen recognition by T cells is necessary for T-cell activation, proliferation, cytokine secretion, and the cytolytic and regulatory function of T cells.<sup>11,12</sup> Specific antigen recognition is accomplished by the binding of a pro-

Grant sponsor: National Library of Medicine; Grant number: R01 LM07174.

\*Correspondence to: John Moulton, Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: [jmoulton@tunc.org](mailto:jmoulton@tunc.org)

Received 13 January 2003; Accepted 21 March 2003

cessed antigen: a peptide bound to a major histocompatibility complex (MHC) molecule (reviewed in Ref. 13). The TCRs are highly variable transmembrane glycoproteins. More than 90% consist of a TCR  $\alpha$ -chain and  $\beta$ -chain. Each chain has a variable (V) domain and a constant (C) domain. The V domains of the  $\alpha$ - and  $\beta$ -chains bind the complex of peptides and MHC molecules (pMHCs) and play an important role in antigen recognition. The recombination mechanisms for germline TCR variant (V) segment, diverse (D) segment, and joint (J) segment<sup>11</sup> provide a large antigen recognition repertoire.<sup>14,15</sup>

The TCR  $\beta$ -V region also interacts with superantigens (SAGs). SAGs are bacterial or viral proteins that have the ability to activate a large T-cell population and cause T cells to secrete cytokines, resulting in inflammation and disease, such as toxic shock syndrome.<sup>11,16</sup> SAGs may be involved in some autoimmune diseases such as diabetes mellitus,<sup>17</sup> multiple sclerosis (MS),<sup>18</sup> and rheumatoid arthritis.<sup>19</sup> The activation of a large T-cell population requires the simultaneous interaction of SAGs with a TCR  $\beta$ -V region and a pMHC.

In recent years, the availability of the X-ray crystal structures of the TCR  $\beta$ -chain,<sup>20</sup> TCR  $\alpha\beta$ -heterodimer,<sup>21–23</sup> the TCR  $\alpha\beta$ -heterodimer complex with a pMHC,<sup>24–28</sup> and the TCR complex with a SAG<sup>29</sup> have provided extensive knowledge of the three-dimensional (3D) structure of the TCR and its interaction with pMHC and SAG. The CDR1, CDR2, and CDR3 in the V domains of the  $\alpha$ - and  $\beta$ -chains are the binding sites for pMHC. The CDR1, CDR2, and HV4 in the V domain of TCR  $\beta$ -chain are the binding sites for SAG.<sup>29–31</sup> The crystal structures of TCR and SAG complex have shown that SAGs bind to the CDR1, CDR2, and HV4 of TCR V $\beta$  through H-bonds and van der Waals interactions.<sup>30,32–34</sup>

Knowledge of the 3D structures of the TCR and its interactions with pMHC and SAGs provides a basis for analyzing the structural and molecular functional effects of missense cSNPs on the TCR  $\beta$ -V region and the binding of antigen and SAG. Because the TCR V $\beta$ -gene does not encode the CDR3 region, SNP impact on this region is not discussed in this study.

A total of 279 SNPs in the 65 TCR V $\beta$ -genes (46 functional genes and 17 pseudogenes) have been identified by using DNA sequencing by the Nickerson group at the University of Washington.<sup>10</sup> 137 SNPs are located in exons, two are in the promoter, 34 are in introns, and 17 are in the recombination signal sequence (RSS). Seventy-four percent of the SNPs in exons (cSNPs) are nonsynonymous (missense mutation and nonsense mutation), much higher than that found in other recent SNP surveys, where only about half of the SNPs are nonsynonymous.<sup>35–37</sup> A total of 72 missense cSNPs have been identified in the functional V $\beta$ -genes. Fourteen of these are located in the leader peptide region (LDR). The other 58 are in the framework regions (FR) and CDRs. The missense mutations in the leader peptide cannot be analyzed by the SNP model and may or may not affect protein transportation. The impact on structure and molecular function of the

TCR V $\beta$ s of the 58 missense cSNPs in the V $\beta$  regions of the framework and CDRs has been investigated.

## MATERIALS AND METHODS

### Source of SNP Data

The SNP location and frequency data were taken from the TCR SNP database: <http://droog.mbt.washington.edu/>

### Identification of Residue Positions Corresponding to TCR V $\beta$ -SNPs

The TCR SNP database provides the base changes (SNPs), the base position of each SNP, nonsynonymous residue changes arising from SNPs and the accession numbers of TCR V $\beta$ -segments in the Entrez nucleotide database (<http://www.ncbi.nlm.nih.gov/Entrez/>). These data were related to protein level information, as follows: 1) For each of the 58 SNPs, retrieve the wild-type DNA sequence of each V $\beta$ -segment from Entrez by using the accession number; 2) Mutate the wild-type nucleotide base into a SNP manually at the position of each missense cSNP; 3) Translate each V $\beta$ -DNA sequence to amino acid sequence by using BLASTX<sup>38</sup>; 4) Align the translated protein sequence with the wild-type protein sequence using ClustalW,<sup>39</sup> confirm the sequences are identical except for one residue, and identify the residue incorporating the missense cSNP; 5) Confirm that the nucleotide base change induces the amino acid change reported in the SNP database. (Note that residue numbers of missense mutations used here refer to the mature TCR V $\beta$ -subunit, after removal of the leader peptide.)

### Structure Modeling

All TCR V $\beta$ -domains are 29% or higher identical in sequence to an experimentally determined structure, so that standard homology-modeling techniques could be used<sup>40</sup> for the framework regions. Available homologous structures for each wild-type TCR V $\beta$ -protein sequence were identified by searching against the sequences in the protein database (PDB) with use of BLASTP.<sup>38</sup> Where more than one homolog was identified, the crystal structure with the highest sequence identity and the best resolution was chosen as the template to build the main-chain of the framework region and HV4. Side-chains were built by using SCRWI.<sup>41</sup>

Model building of the CDRs of TCR V $\beta$  was less straightforward. It has been shown that the hypervariable regions of immunoglobulins usually adopt one of a small number of different main-chain conformations, termed canonical structures.<sup>42,43</sup> Analysis of six known crystal structures for TCR V $\beta$ s has shown that similarly there are three primary canonical structures for each of the CDR1 and CDR2 hypervariable regions, termed  $\beta$ 1-1,  $\beta$ 1-2,  $\beta$ 1-3, and  $\beta$ 2-1,  $\beta$ 2-2,  $\beta$ 2-3, respectively.<sup>44</sup> Key residues responsible for each conformation have been identified,<sup>44</sup> and it has been observed that these are conserved in the V $\beta$ -subfamilies.<sup>20,21,29</sup> Canonical structures were assigned on the basis of these sequence signals.

## Conservation Study

A PERL script incorporating FASTA and Clustal was used to search against the NR database to identify and align homologous sequences with an E-score < 0.001. The relative conservation of amino acid type at each position in a resulting alignment was calculated as:

$$S(j) = \sum_{i=1}^{20} P(i) \ln P(i)^{45}$$

$$P(i) = n(i) / \sum n(i) \quad (i = 1, 20)$$

where  $S(j)$  is the sequence entropy at position  $j$  in the multiple-sequence alignment and  $n(i)$  is the number of occurrences of residue type  $i$  in the family.

The conservation scores were scaled to the range of normal crystallographic temperature factors, so that standard software GRASP<sup>46</sup> could be used to display the protein surface conservation by color. It has been observed that protein–protein interaction sites are generally characterized by a high level of residue conservation on the protein surface.<sup>47</sup> Surface inspection using the GRASP interface was used to identify any conserved patches that might constitute protein or other interaction sites.

## Assignment of SNP Effects on the Structure and Molecular Function of the TCR V $\beta$ -Domains

The structural and molecular functional effects of missense cSNPs on the TCR V $\beta$ -domains were assigned according to the rules in the SNP model.<sup>4</sup> The environment of each mutated amino acid was examined to determine if any of the modeling rules applied. These rules are divided into five classes: effects on protein stability, ligand binding, catalysis, allosteric regulation and posttranslational modification. Within each class, additional rules may be used. In particular, for mutations affecting protein stability, precise parameters are used for assessing the impact of a range of energetic factors, such as H-bonds, salt bridges, hydrophobic interactions, and so on. Details are in Ref. 4 and at <http://www.SNPS3D.org/SNPS-model.pdf>

## RESULTS

### Structure Modeling

The structure modeling is summarized in Tables I and II. Forty-one CDR1s and 44 CDR2s out of the total of 46 functional V $\beta$ -domains<sup>15</sup> were assigned canonical structures 1–3 on the basis of conserved key residues and loop length. A few were assigned putative canonical structures 4 and 5, and a few could not be assigned any canonical structure. Thirty-five of the 46 functional V $\beta$ -genes contain one or more missense cSNPs, and suitable models could be built for all of these.

The models include 54 of the 58 missense cSNPs. The other four lie in the N-terminal region of the mature domain not included in any experimental structure, and so these could not be analyzed.

### Location of 54 Missense Mutations in the 3D Structures of TCR $\beta$ -chain V Domains

Forty-two of the 54 cSNP missense mutations are located in the framework regions of TCR V $\beta$ -domains, with

most exposed on the protein surface. The other 12 are in the loop regions, with four in CDR1, two in CDR2, and six in HV4. Table III lists the expected effect of these 54 cSNPs on protein structure and function.

### Effects of 42 Missense Mutations in Framework Regions of TCR $\beta$ -chain V Domains

Thirty-five of the 42 framework cSNPs are not expected to have significant effects on protein structure and function. The other seven are all expected to affect protein stability and are discussed in detail below.

#### Framework mutations affecting protein stability

Seven missense mutations in the framework region of TCR V $\beta$ s are expected to affect protein stability. These are C111R (BV6S1), D105Y (BV6S1), R55Q (BV6S5), R63Q (BV15S1), T105A (BV24S1), G35R (BV13S7), and G35V (BV26S1). The locations of these in the structure are shown in Figure 1.

**C111R in BV6S1.** Cys-111 and Cys-44 form a disulfide bond. C111R breaks that bond and introduces the positively charged Arg-111 into the hydrophobic core. Therefore, it falls under two rules affecting protein stability. Literature reports show that this disulfide bond is conserved in the superfamily and is critical for the antibody-like structure.<sup>48</sup> It is believed that the missense mutation C111R is related to junior rheumatoid arthritis (JRA).<sup>49</sup>

**D105Y in BV6S1.** Asp-105 is partly buried and forms two salt bridges with surface Arg-83 (4.4 Å) and Arg-55 (4.1 Å). D105Y removes these two salts bridges and creates overpacking with C $\beta$  of Arg-83 (1.1 Å).

**R55Q in BV6S5.** This change causes loss of two H-bonds with the backbone of Glu-104 and one salt bridge with Asp-105 (4.2 Å).

**R63Q in BV15S1.** This causes loss of one surface salt bridge with Asp-57 (3.1 Å).

**T105A in BV24S1.** This causes loss of one H-bond with backbone of Val-116.

**G35R in BV13S7 and G35V in BV26S19.** Gly-35 plays a critical role in forming a type II  $\beta$ -hairpin structure and is conserved in almost all the TCR V $\beta$ -subfamily. The  $\Phi$  and  $\Psi$  of Gly-35 are not in the Ramachandran sterically allowed region for non-Gly residues, so backbone strain is expected for these mutants. There are three additional missense mutations in the TCR V $\beta$ s where Gly is mutated to another residue, without the introduction of backbone strain.

#### Possible role of protein–protein interactions

Most of the 35 “no effect” mutations are located on the protein surface. Our previous study<sup>4</sup> showed that unidentified macromolecular binding sites might cause false-negative results in this situation. It is known that the cell surface proteins CD4 and CD8 have physical and functional interactions with the TCR,<sup>50</sup> but it is not clear whether CD4 and CD8 bind to the V $\beta$ -domain. How likely is it that the no effect missense mutations may in fact be involved in binding to these or other proteins? We have searched for possible protein-binding sites by inspecting

**TABLE I. V $\beta$ -Genes, Number of Missense cSNPs, and Information for Model Building**

A	B	C	D	E	F
BV1S1	1	1nfd (2.8 Å)	57%	2	2
BV2S1	2	1ao7 (2.6 Å)	29%	5	5
BV4S1	1	1bd2 (2.5 Å)	31%	4	5
BV5S1	1	1nfd	54%	2	2
BV5S2	2	1nfd	54%	2	2
BV5S3	1	1nfd	56%	2	2
BV5S4	1	1nfd	49%	2	2
BV5S5	1	1nfd	52%	2	2
BV5S6	2	1nfd	51%	Not assigned	2
BV5S7	3	1nfd	54%	2	2
BV6S1	7	1nfd	40%, gap 1%	2	2
BV6S3	1	1nfd	42%	2	2
BV6S4	2	1nfd	40%, gap 1%	2	2
BV6S5	3	1nfd	45%, gap 1%	2	2
BV6S6	1	1nfd	43%, gap 1%	2	2
BV6S8	1	1bec (1.7 Å), 1ao7 1nfd	37%, gap 1%	1	2
			41%, gap 1%		
BV7S1	1	1bec, 1ao7 1nfd	31%, gap 1%	1	2
			36%, gap 4%		
BV8S1	1	1nfd	43%, gap 1%	2	2
BV8S2	1	1nfd	43%, gap 1%	2	2
BV8S3	2	1bec, 1ao7 1nfd	38%, gap 1%	Not assigned	2
			37%, gap 1%		
BV9S1	1	1ao7	29%	1	2
		1nfd	35%		
BV11S1	1	1bec, 1ao7	55%	1	1
BV12S1	1	1bec, 1ao7	65%	Not assigned	1
BV12S2	1	1bec, 1ao7	65%	1	1
BV13S1	1	1bec, 1ao7	60%	1	1
BV13S5	1	1bd2	67%	1	1
BV13S6	2	1ao7	88%	1	1
BV13S7	2	1bd2	86%	1	1
BV15S1	1	1bec, 1ao7	55%	1	1
BV18S1	1	1nfd	36%, gap 1%	2	2
BV21S1	2	1nfd	41%, gap 1%	2	2
BV21S2	1	1nfd	42%, gap 1%	2	2
BV21S3	1	1nfd	44%, gap 1%	2	2
BV24S1	1	1bec, 1nfd	36%, 28%	1	2
BV26S1	2	1nfd	36%	2	2

Column A: name of the human V $\beta$ -gene. B: total number of missense cSNPs in the mature product of each TCR V $\beta$ . C: PDB codes for model building templates and the resolution of the structure. D: sequence identity between each V $\beta$  and the template. E: canonical structure of CDR1. F: canonical structure of CDR2.

the surface of the V $\beta$ -domain for patches of conserved residues using GRASP. None were found, suggesting there are no direct interactions with CD4 and CD8 or other proteins.

### **Mutations in the V $\alpha$ -V $\beta$ interface**

Four of the 35 no effect mutations are located in the V $\alpha$ -V $\beta$  interface. These are Q67H (BV1S1), G61V (BV5S5), E77K (BV9S1), and Q67R (BV21S1). These residues are all at the edges of the interface, and the side-chains point toward the surface and are not fully buried. In G61V (BV5S5), the side-chain of Val-61 in V $\beta$  and that of the interacting residue in V $\alpha$  can be accommodated parallel to each other, so there is no overpacking effect. For the other

residues, Q67H (BV1S1), E77K (BV9S1), and Q67R (BV21S1), the distances to the nearest residues in V $\alpha$  are so large (>5.5 Å) that overpacking and unfavorable electrostatic interactions can be avoided. Several residues in the V $\alpha$ -V $\beta$  interface are conserved across all human and mouse TCRs and appear to stabilize the  $\alpha\beta$  heterodimer structure.<sup>21</sup> None of the four missense mutations is at any of these positions.

### **Missense Mutations in Hypervariable Regions (CDR1, CDR2, and HV4)**

Missense mutations in hypervariable regions (CDR1, CDR2, and HV4) are of special interest because they may affect ligand binding. CDR1 and CDR2 are involved in

**TABLE II. Summary of Structure Modeling**

Name	Count
V $\beta$ -domains	65
Functional V $\beta$ -domains	46
Pseudogenes	19
Functional V $\beta$ -domains containing missense cSNPs	35
Missense cSNPs in the functional genes	72
Missense cSNPs in the leader peptide regions of the functional genes	14
Missense cSNPs in framework and loop regions	58
Missense cSNPs in framework and loop regions included in a model	54

The 54 missense cSNPs require modeling of 35 V $\beta$ -domains. The models are based on sequence identities ranging from 29 to 88%. A list of the 35 V $\beta$ -domains, the number of missense cSNPs in each domain, the template for model building of each V $\beta$ , the sequence identity between the template and model, and the canonical structure of CDR1 and CDR2 are shown in Table I. The higher the sequence identity, and the better resolution of the template's crystal structure, the more straightforward the interpretation of SNP impact. Sequence alignment of targets and templates shows that none of the missense cSNPs are located in an insertion or deletion area. The reliability of the analysis also depends on the location of SNPs: SNPs located in the framework regions or in CDR1/CDR2s that have canonical structures can generally be reliably interpreted. SNPs in the HV4 region, for which neither canonical structures nor an appropriate template is available, are less reliably interpreted. The rules in the SNP model<sup>4</sup> were applied to these 54 missense mutations. The results are shown in Table III.

binding of pMHC and SAG. HV4 is involved in binding of SAG. SNP impact on stability and binding is analyzed separately in each region.

#### **Effect of four missense mutations in CDR1**

Four missense mutations are located in CDR1: P44L (BV8S1), D49H (BV8S2), M45V (BV18S1), and H48R (BV21S1).

All four of these CDR1s are assigned canonical structure  $\beta$ 1-2.<sup>44</sup> The principal conformational determinants are the conserved residues Pro-25, His-29, Val-32, and Ser-94, and the H-bond between His-29 and Ser-94.<sup>44</sup> Models of the four CDR1 mutations were built on the basis of the crystal structure of 1nfd, which has the  $\beta$ 1-2 canonical structure. This template is a TCR  $\alpha\beta$ -heterodimer with no bound pMHC or SAG. The lack of a crystal structure complex with pMHC and SAG limits the analysis of SNP impact on binding.

On the basis of our model, both P44L (BV8S1) and H48R (BV21S1) are expected to change the conformation of CDR1 and so affect the binding of pMHC and SAG. P44L (residue 25) causes overpacking with Ser-94. H48R (residue 29) of BV21S1 removes the key H-bond between His-29 and Ser-94 in CDR1, as shown in Figure 2.

The effect of D49H (BV8S2) on binding to pMHC and SAG cannot be determined because of the lack of a model with ligand. M45V is not expected to have a significant effect on binding because it is far from binding sites and the side-chain faces toward the constant domain.

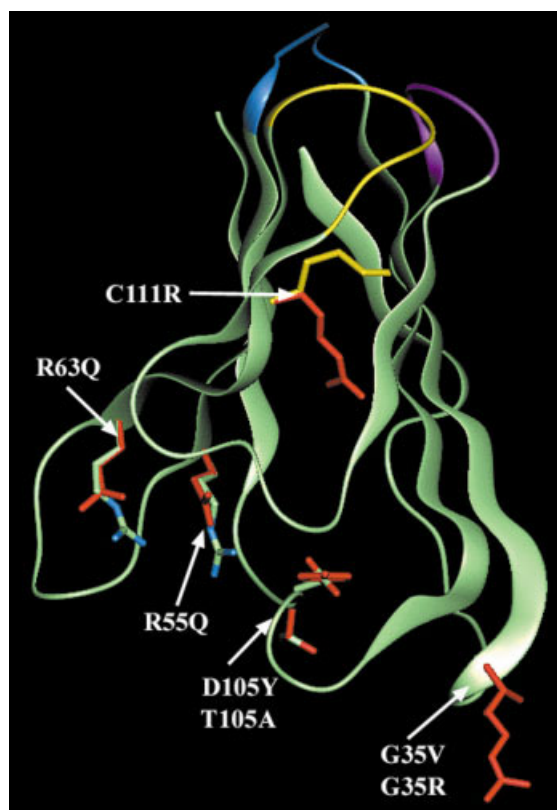


Fig. 1. Seven missense mutations located in the framework regions of TCR V $\beta$ s that are expected to affect protein stability, shown on the structure of PDB entry 1bec. The framework regions are green; CDR1, CDR2, and HV4 are blue, yellow, and magenta, respectively. CDR3 is not shown. Wild-type side-chains are colored by element, and the corresponding mutated side-chains are red. The disulfide bond is yellow. C111R (BV6S1) breaks a conserved disulfide bond and buries a charge. D105Y (BV6S1) causes loss of two salt bridges and overpacking. R55Q (BV6S5) causes loss of two H-bonds and one salt bridge. R63Q (BV15S1) causes loss of one salt bridge. T105A (BV24S1) causes loss of one H-bond. G35R (BV13S7) and G35V (BV26S1) causes backbone strain.

#### **Effect of two missense mutations in CDR2**

There are two missense mutations located in CDR2: E71K (BV5S6) and Q72H (BV12S2).

**E71K in BV5S6.** The CDR2 of BV5S6 is assigned the canonical structure  $\beta$ 2-2. E71K is on the surface and does not appear to affect stability or loop conformation. The lack of a crystal structure complex with pMHC or SAG containing this CDR2 conformation prevents analysis of the effect of the mutant on binding.

**Q72H in BV12S2.** The CDR2 of BV12S2 has conformation  $\beta$ 2-1.<sup>44</sup> There are a total of seven crystal structures of TCR/pMHC-I and TCR/pMHC-II complexes with the  $\beta$ 2-1 CDR2 conformation. Each of them was used to build models to analyze the effects of Q72H of BV12S2. The models show that the minimal atom–atom distance of Q72H to a pMHC atom is 5 Å. Thus, Q72H (BV12S2) is not expected to have a significant effect on binding of pMHC.

Q72H is also not expected to have a significant effect on SAG binding. Two crystal structure complexes of mouse BV8S2 and its SAG ligand SEB (pdb code: 1sbb)<sup>30</sup> and SEC3 (pdb code: ljck)<sup>30</sup> were used as templates. BV12S2

**TABLE III. Structural and Functional Effects of 54 Missense cSNPs in the TCR  $\beta$ -chain V Domain**

A	B	C	D	E	F
BV1S1	Q67H	48	No effect	FR and in the interface of the $\alpha\beta$ -chain	0.24
BV2S1	W24R	10	No effect	Both on surface of FR	0.49
	Q56K		No effect		
BV4S1	M98L	82	No effect	Partially buried in FR	0.39
BV5S1	S93F	75	Stability: loss of 1 H-bond and overpacking	Ser-93 is buried and lies between the FR and HV4	0.03
BV5S2	F88Y	70	Possible effects on SAG binding cannot be assessed	HV4	0.16
	L101W	83	No effect	Surface of FR	0.08
BV5S3	Q37H	18	No effect	Surface of FR	0.42
BV5S4	A38V	19	No effect	Partially buried in FR	0.34
BV5S5	G61V	42	No effect	Interface of the $\alpha\beta$ -chain	0.12
BV5S6	E71K	52	No effect on pMHC binding; possible effects on	CDR2	0.02
	Y91D	73	SAG binding cannot be determined for both SNPs	HV4	0.08
BV5S7	Q36H	17	No effect	All on surface of FR	0.18
	F79L	60	No effect		0.03
	G103E	85	No effect		0.18
BV6S1	Y37D	18	No effect	All on surface of FR except D105Y, which is partially buried and C111, which is in the core	0.26
	N81K	62	No effect		0.26
	R103Q	84	No effect		0.26
	D105Y	86	Stability: loss of 2 salt bridges and overpacking		0.02
	V108M	89	No effect		0.05
	V108A	89	No effect		0.14
	C111R	92	Stability: breaking of a disulfide bond and introduction of a buried charge		0.23
BV6S3	P89A	70	Possible effect on SAG binding cannot be assessed	HV4	0.02
BV6S4	N26D	7	No effect	Surface of FR	0.38
	R100H	81	No effect		0.03
BV6S5	R55Q	36	Stability: lost 2 H-bonds with the backbone and 1 salt bridge	Surface of FR	0.02
	S57R	38	No effect	Surface of FR	0.5
	G91E	72	Possible effect on SAG binding cannot be assessed	HV4	0.5
BV6S6	P80L	60	No effect	Surface of FR	0.05
BV6S8	R100C	81	No effect	Surface of FR	0.05
BV7S1	P80R	61	No effect.	Surface of FR	0.03
BV8S1	P44L	25	Binding. P44L probably changes the conformation of CDR1	CDR1	0.14
BV8S2	D49H	30	No effect on pMHC binding; possible effect on SAG binding cannot be assessed	CDR1	0.05
BV8S3	H29D	10	No effect	All on surface of FR	0.45
	M59G	40	No effect		0.05
BV9S1	E77K	58	No effect	Interface of the $\alpha\beta$ chain	0.03
BV11S1	T90M	72	No effect on SAG binding	HV4	0.05
BV12S1	H60R	41	No effect	Surface of FR	0.02
BV12S2	Q72H	53	No effect on pMHC and SAG binding	CDR2	0.5
BV13S1	N81S	62	No effect	Surface of FR	0.02
BV13S5	R37S	18	No effect	Surface of FR	0.18
BV13S6	R30H	11	No effect	Both on surface of FR	0.14
	T43A	24	No effect		0.39
BV13S7	G35R	16	Stability backbone strain	Both on surface of FR	0.16
	Q103R	86	No effect		0.20
BV15S1	R63Q	44	Stability; loss of a surface salt bridge	Surface of FR	0.25
BV18S1	M45V	26	No effect on pMHC binding; possible effect on SAG binding cannot be assessed.	CDR1	0.22
BV21S1	H48R	29	Binding of both pMHC and SAG.	CDR1	0.14
	Q67R	48	No effect	In the interface of the $\alpha\beta$ -chain	0.14
BV21S2	L55R	36	No effect	Partially buried in FR	0.43
BV21S3	R28G	9	No effect	Surface of FR	0.15
BV24S1	T105A	87	Stability loss of a H-bond	Surface of FR	0.16
BV26S1	G35V	16	Stability backbone strain; Gly-35 is in a type II $\beta$ hairpin turn	Both on surface of FR	0.02
	D43N	24	No effect		0.02

Column A: gene name. B: type and position of each resulting missense mutation (residue numbers for the unprocessed TCR  $\beta$ -chain V domain precursor, including the leader peptide region). C: Residue number in the mature TCR  $\beta$ -chain V domain. D: Predicted structural and molecular functional effect. E: 3D structural location: framework region (FR), complementarity determining region (CDR) or fourth hypervariable region (HV4). F: Population frequency of each missense cSNP.<sup>10</sup>

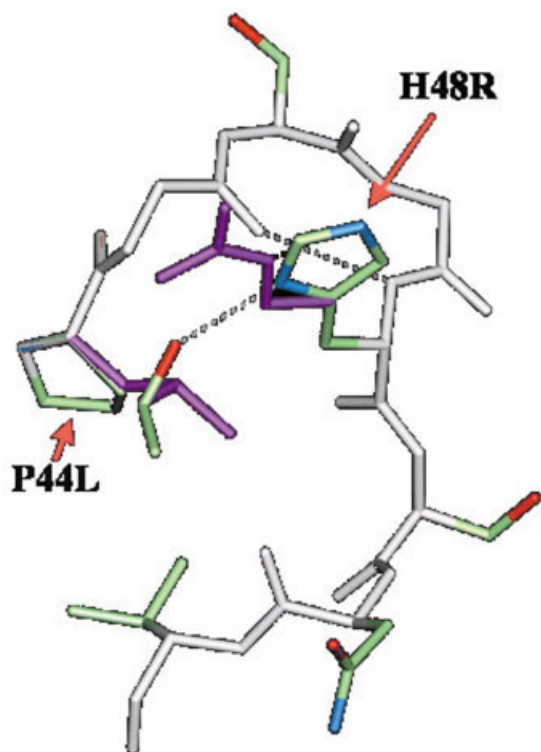


Fig. 2. Two CDR1 missense mutations expected to affect the binding of both pMHC and SAG. The main-chain of the loop is gray. Wild-type side-chains are colored by element, and mutated side-chains are magenta. For clarity, not all side-chains are shown. The conformation of this CDR1 is mainly determined by the conserved residues Pro-25, His-29, Val-32, P44L (residue 25) of BV8S1 causes overpacking with Ser-94. H48R (residue 29) of BV21S1 replaces the conserved residue His-29 with Arg-29, so the conserved H-bond between His-29 and Ser-94 is removed. Both of these mutations are expected to change the conformation of CDR1 and thus affect the binding of pMHC and SAG.

shares 65% sequence identity with mouse BV8S2, and both have the  $\beta$ 2-1 canonical structure for CDR2, suggesting that BV12S2 binds SEB and SEC3, so that these experimental complexes provide suitable templates. Furthermore, it has been shown that SEB and SEC3 are able to expand T cells bearing human V $\beta$ 12.<sup>16,51</sup> The models indicate that the backbone atoms of Q72H of BV12S2 form one H-bond with SEC3, but the side-chain atoms have no interaction with SEB or SEC3. Thus, no significant effect on SAG binding is expected for this mutant.

#### Effect of six missense mutations in HV4

There are six missense mutations located in HV4: S93F (BV5S1), F88Y (BV5S2), Y91D (BV5S6), P89A (BV6S3), G91E (BV6S5), and T90M (BV11S1). S93F is buried, and the others are on the protein surface. HV4 is not involved in pMHC binding, so that only possible effects on stability and SAG binding need be considered. Even though the overall sequence identity of these BVs and their templates are fairly high (42–54%), the local sequence alignment for HV4 region is not very satisfactory, as shown in Table IV. There are also no crystal structures of SAG binding.

**S93F in BV5S1.** Ser-93 is buried between the framework and the end of HV4 and forms an H-bond with

TABLE IV. Sequence Alignment of Missense Mutations in the HV4 Regions of BV5S1, BV5S2, BV5S6, BV6S3, BV6S5, and BV11S1 and Their Homologs With Known 3-D Crystal Structures

N15 (1nfd_B)	OFDDYHS
BV5S1	QFSNSRS
BV5S1_S93F	QF <b>F</b> NSRS
BV5S2	QFPNYSS
BV5S2_F88Y	Q <b>Y</b> PNYSS
BV5S6	QFPNYSS
BV5S6_Y91D	QFP <b>N</b> DSS
BV6S3	RPEGSIS
BV6S3_P89A	R <b>A</b> EGSIS
BV6S5	RTGGSVS
BV6S5_G91E	RT <b>E</b> GGSVS
<b>14. 3. d (1bec)</b>	<b>RPSQEQF</b>
BV11S1	RIR <b>T</b> EHF
BV11S1_T90M	RIR <b>M</b> EHF

The template for model building of BV5S1, BV5S2, BV5S6, BV6S3, and BV6S5 is 1nfd\_B. The overall protein sequence identity is 54%, 54%, 51%, 42%, and 45%, respectively. The template for BV11S1 is 1bec with overall sequence identity of 55%. The templates are in bold. The wild-type residues at each mutation position are blue, and mutated residues are red.

TABLE V. Summary of SNP Impact on TCR V $\beta$ s

Location of mutant	Functional effect	Effect cannot be determined		
		No effect	Effect cannot be determined	Total
FR	7 (Stability)	35	0	42 (78%)
CDR1	2 (Binding)	1	1	4 (7%)
CDR2	0	1	1	2 (4%)
HV4	1 (Stability)	3	2	6 (11%)
Total	10 (19%)	40 (74%)	4 (7%)	54 (100%)

Approximately 19% of missense cSNPs are expected to have an effect on molecular function of TCR V $\beta$ s, with most affecting stability. The effect of another four mutations (7%) cannot be determined; 74% are expected to have no effect.

Trp-53. S93F causes loss of this H-bond as well as overpacking with Trp-53 and the conserved disulfide bond between Cys-110 and Cys-42, and thus is expected to affect stability. It is unlikely to directly interfere with SAG binding.

**F88Y in BV5S2, Y91D in BV5S6, P89A in BV6S3, G91E in BV6S5, and T90M in BV11S1.** Uncertainty of backbone conformation of HV4 complicates the interpretation of these mutants. However, they appear to be located on the surface, and so we tentatively conclude they do not affect stability. Some conclusions concerning SAG binding can be drawn for F88Y, Y91D, and P89A. These three mutants are distant from the SAG-binding site, so they probably can be categorized as no effect. No conclusions can be drawn for G91E in BV6S5 and T90M in BV11S1.

## DISCUSSION

### Fraction of Missense cSNPs That Are Deleterious

Table V summarizes the effects of missense SNPs on the structure and function of the V $\beta$ s. Nineteen percent of missense cSNPs have some effect on the molecular function of TCR V $\beta$ s, according to the model, with most

affecting stability. Seventy-four percent are not expected to have any significant effect; 7% do not affect stability, but the effects on pMHC or SAG binding cannot be determined. These results are similar to those obtained by applying the model to other gene sets<sup>4</sup> where about 30% of cSNPs were found to be deleterious. Other groups have reported similar values for different protein sets with somewhat different methods.<sup>5</sup>

### Severity of SNP Impact

The seven mutations affecting stability cover a range of likely severity. At one extreme, disruption of the conserved disulfide bond in the core of the structure would be expected to destroy the folded state *in vitro* and is known to be associated with disease *in vivo* (see below). At the other, T105A results in the loss of single H-bond between a polar group and the backbone. *In vitro* mutagenesis studies<sup>52</sup> and analysis of monogenic disease-associated mutants<sup>4</sup> both suggest that under some circumstances, loss of one H-bond may have a significant impact on stability and function *in vivo*.

### Class of Effect on Protein Structure and Function

Most of the apparently deleterious cSNPs affect protein stability, consistent with our observations for other proteins.<sup>4</sup> That analysis suggests lower stability will substantially reduce the level of functioning of the affected TCR molecules *in vivo*. Mechanisms responsible for this are known for proteins that traffic through the ER and in the cytoplasm.<sup>53,54</sup> However, as discussed below, the situation with TCR molecules may be more complicated. P44L in BV8S1 and H48R in BV21S1 affect pMHC binding and may also weaken the immune response and so are potentially detrimental. On the other hand, these two mutants also affect SAG binding, so they may decrease the susceptibility to SAG-related disease and thus are also potentially beneficial.

### Relationship to Disease

The stability mutant C111R in BV6S1 is the only deleterious SNP identified in this study so far associated with disease: juvenile rheumatoid arthritis (JRA).<sup>49</sup> It has been shown that there is substantially reduced mRNA in peripheral blood and the thymus.<sup>55</sup> Presumably, during maturation, T cells carrying this allele fail to be positively selected by the normal weak self-recognition criterion.<sup>56</sup> The symptoms of JRA would be expected to result from self-recognition and a resulting autoimmune response. It is difficult to see how that could be caused by the absence of a particular set of TCR molecules. Perhaps, overproduction of other TCRs compensates for the gap in the response repertoire, thus increasing susceptibility to autoimmune effects. If this were the case, we would expect that similar consequences might be associated with the other stability-damaging mutants, but none are known so far. More straightforward disease-related effects are also likely. Low levels of TCRs containing any of the stability mutants may lead to a weakened immune response and thus increased susceptibility to particular sorts of infection. Now that the

likely SNPs have been identified, it should be possible to look for such correlations.

The other deleterious SNPs, not known to be associated to disease, are candidates for involvement in complex, polygenic disease.

### Population Frequency of Deleterious SNPs

It might be expected that deleterious SNPs will have a relatively low incidence in the population. The sample size here is too small to assess this reliably, but there is a tendency in that direction. Frequencies of the 10 identified potentially deleterious SNPs range from 0.02 to 0.23. The highest value happens to be for the one SNP known to be associated with disease: C111R. Only one other frequency is >0.2 (R63Q) and four are <0.05.

### Total Versus Deleterious Missense cSNPs

The most surprising finding from the experimental study of TCR V $\beta$ s is the high fraction of SNPs that are nonsynonymous: 74%. As noted by Subrahmanyam et al.,<sup>10</sup> this is close to the value expected in the absence of any selection pressure (76%) and substantially larger than that found in other studies of gene sets, where it is typically close to 50%.<sup>35–37</sup> Subrahmanyam et al.<sup>10</sup> suggest that the explanation for the high fraction of nonsynonymous SNPs is that the TCRs form a large family with overlapping function, so that selection pressure on any single member is weak. That is, in the absence of any other information, a high fraction of nonsynonymous SNPs is assumed to imply a high fraction of accepted non-neutral SNPs. The structure-based analysis allows us to address this question more directly, because it identifies which SNPs degrade function at the protein level. As noted earlier, we find that the fraction of nonsynonymous cSNPs that affect function is in fact similar to that found in the other gene sets (between 19 and 26% versus 30%). If a high degree of redundancy were a significant factor, we would expect to see a higher fraction.

### No Evidence of Hypervariability

A second possible explanation for a high fraction of nonsynonymous SNPs is hyperselection. That is, some of these nonsynonymous SNPs cause useful variation in function, rather than being neutral or deleterious. We would expect such variability to be primarily (but not entirely) in binding to pMHC and so to occur in CDR1 and CDR2. There might also be selection for TCRs that are resistant to SAG binding, and SNPs for these would be found in CDR1, CDR2, and HV4. There are a total of 12 missense cSNPs in these loops, roughly the same as the overall density of missense cSNPs in non-CDR-coding sequences, suggesting that the higher fraction of nonsynonymous SNPs does not contribute to a larger antigen or a reduced SAG-binding repertoire. Two mutants in CDR1 affect pMHC and SAG binding, one in HV4 affects stability, four have no effect, and the effect of other four cannot be reliably analyzed. These numbers are small, but they certainly do not suggest any role for hypervariability.

### Limitations of the Analysis

TCR molecules operate in a complex molecular environment. Not only do they form complexes with pMHC molecules and SAGs, they also interact with other cell surface factors, such as CD4 and CD8. In addition, there must be a mechanism by which pMHC binding triggers an intracellular response, leading to cell proliferation. These factors complicate interpretation of the impact of some SNPs. The binding sites for CD4 and CD8 are not known. We have used analyses of protein surface conservation to search for possible macromolecular interaction sites on the V $\beta$ -domains and find no evidence of any, supporting the conclusion that all the framework surface mutants are effectively neutral.

Several classes of TCR SNPs are not included in this study. TCR molecules are transported to the cell surface with the aid of leader peptide (LDR). There are 14 nonsynonymous SNPs in this region; any of these may affect transport. A second set of 17 SNPs is found in the RSS region and may affect combination with D or J segments. Two SNPs are also found in the promoter region.

### ACKNOWLEDGEMENT

We thank Eugene Melamud for extensive assistance with data handling and molecular modeling.

### REFERENCES

- Wang Z, Moulton J. SNPs, protein structure, and disease. *Hum Mutat* 2001;17:263–270.
- Krawczak M, Ball EV, Fenton I, Stenson PD. Human gene mutation database—a biomedical information and research resource. *Hum Mutat* 2000;15:45–51.
- Peltonen L, McKusick VA. Genomics and medicine: dissecting human disease in the postgenomic era. *Science* 2001;291:1224–1229.
- Emahazion T, Feuk L, Jobs M, Sawyer SL, Fredman D, St Clair D, Prince JA, Brookes AJ. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet* 2001;17:407–413.
- Sunyaev S, Ramensky V, Koch I, Lathe W III, Kondrashov AS, Bork P. Prediction of deleterious human alleles. *Hum Mol Genet* 2001;10:591–597.
- Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol* 2001;307:683–706.
- Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;11:863–874.
- Ferrer-Costa C, Orozco M, de la Cruz X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J Mol Biol* 2002;315:771–786.
- Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;30:3894–3900.
- Subrahmanyam L, Eberle MA, Clark AG, Kruglyak L, Nickerson DA. Sequence variation and linkage disequilibrium in the human T-cell receptor beta (TCRB) locus. *Am J Hum Genet* 2001;69:381–395.
- Marrack P, Kappler J. The T cell receptor. *Science* 1987;238:1073–1079.
- Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature* 1988;334:395–402.
- Rudolph MG, Wilson IA. The specificity of TCR/pMHC interaction. *Curr Opin Immunol* 2002;14:52–65.
- LaRocque R, Robinson MA. Diversity in the human T cell receptor beta chain. *Hum Immunol* 1996;48:3–11.
- Rowen L, Koop BF, Hood L. The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science* 1996;272:1755–1762.
- Kotzin BL, Leung DY, Kappler J, Marrack P. Superantigens and their potential role in human disease. *Adv Immunol* 1993;54:99–166.
- Conrad B, Weissmahr RN, Boni J, Arcari R, Schupbach J, Mach B. A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* 1997;90:303–313.
- Brocke S, Gaur A, Piercy C, Gautam A, Gijbels K, Fathman CG, Steinman L. Induction of relapsing paralysis in experimental autoimmune encephalomyelitis by bacterial superantigen. *Nature* 1993;365:642–644.
- Cole BC, Griffiths MM. Triggering and exacerbation of autoimmune arthritis by the Mycoplasma arthritis superantigen MAM. *Arthritis Rheum* 1993;36:994–1002.
- Bentley GA, Boulton G, Karjalainen K, Mariuzza RA. Crystal structure of the beta chain of a T cell antigen receptor. *Science* 1995;267:1984–1987.
- Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA. An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 1996;274:209–219.
- Housset D, Mazza G, Gregoire C, Piras C, Malissen B, Fontecilla-Camps JC. The three-dimensional structure of a T-cell antigen receptor V alpha V beta heterodimer reveals a novel arrangement of the V beta domain. *Embo J* 1997;16:4205–4216.
- Wang J, Lim K, Smolyar A, Teng M, Liu J, Tse AG, Hussey RE, Chishti Y, Thomson CT, Sweet RM, Nathenson SG, Chang HC, Sacchetti JC, Reinherz EL. Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. *Embo J* 1998;17:10–26.
- Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 1996;384:134–141.
- Teng MK, Smolyar A, Tse AG, Liu JH, Liu J, Hussey RE, Nathenson SG, Chang HC, Reinherz EL, Wang JH. Identification of a common docking topology with substantial variation among different TCR-peptide-MHC complexes. *Curr Biol* 1998;8:409–412.
- Garcia KC, Teyton L. T-cell receptor peptide-MHC interactions: biological lessons from structural studies. *Curr Opin Biotechnol* 1998;9:338–343.
- Ding YH, Smith KJ, Garboczi DN, Utz U, Biddison WE, Wiley DC. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity* 1998;8:403–411.
- Reinherz BL, Tan K, Tang L, Kern P, Liu J, Xiong Y, Hussey RE, Smolyar A, Hare B, Zhang R, Joachimiak A, Chang HC, Wagner G, Wang J. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* 1999;286:1913–1921.
- Fields BA, Malchiodi EL, Li H, Ysern X, Stauffer CV, Schlievert PM, Karjalainen K, Mariuzza RA. Crystal structure of a T-cell receptor beta-chain complexed with a superantigen. *Nature* 1996;384:188–192.
- Li H, Llera A, Tsuchiya D, Leder L, Ysern X, Schlievert PM, Karjalainen K, Mariuzza RA. Three-dimensional structure of the complex between a T cell receptor beta chain and the superantigen staphylococcal enterotoxin B. *Immunity* 1998;9:807–816.
- Li H, Llera A, Malchiodi EL, Mariuzza RA. The structural basis of T cell activation by superantigens. *Annu Rev Immunol* 1999;17:435–466.
- Patten P, Yokota T, Rothbard J, Chien Y, Arai K, Davis MM. Structure, expression and divergence of T-cell receptor beta-chain variable regions. *Nature* 1984;312:40–46.
- Kline JB, Collins CM. Analysis of the interaction between the bacterial superantigen streptococcal pyrogenic exotoxin A (SpeA) and the human T-cell receptor. *Mol Microbiol* 1997;24:191–202.
- Papageorgiou AC, Collins CM, Gutman DM, Kline JB, O'Brien SM, Tranter HS, Acharya KR. Structural basis for the recognition of superantigen streptococcal pyrogenic exotoxin A (SpeA1) by MHC class II molecules and T-cell receptors. *Embo J* 1999;18:9–21.
- 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. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999;22:231–238.
- Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A,

- Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 1999;22:239–247.
37. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Morre HM, Naik AK, Narayan VA, Neelam B, Nuskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. The sequence of the human genome. *Science* 2001;291:1304–1351.
  38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
  39. Higgins DG, Thompson JD, Gibson TJ. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 1996;266:383–402.
  40. Moult J. Predicting protein three-dimensional structure. *Curr Opin Biotechnol* 1999;10:583–588.
  41. Dunbrack RL, Jr, Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* 1997;6:1661–1681.
  42. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 1987;196:901–917.
  43. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al. Conformations of immunoglobulin hypervariable regions. *Nature* 1989;342:877–883.
  44. Al-Lazikani B, Lesk AM, Chothia C. Canonical structures for the hypervariable regions of T cell alphabeta receptors. *J Mol Biol* 2000;295:979–995.
  45. Shannon C. A mathematical theory of communication. *Bell System Techn J* 1948;27:379–423, 623–656.
  46. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science* 1995;268:1144–1149.
  47. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 1996;257:342–358.
  48. Ioerger TR, Du C, Linthicum DS. Conservation of cys-cys trp structural triads and their geometry in the protein domains of immunoglobulin superfamily members. *Mol Immunol* 1999;36:373–386.
  49. Maksymowyc WP, Gabriel CA, Luyrink L, Melin-Aldana H, Elma M, Giannini EH, Lovell DJ, Van Kerckhove C, Leiden J, Choi E, et al. Polymorphism in a T-cell receptor variable gene is associated with susceptibility to a juvenile rheumatoid arthritis subset. *Immunogenetics* 1992;35:257–262.
  50. Konig R. Interactions between MHC molecules and co-receptors of the TCR. *Curr Opin Immunol* 2002;14:75–83.
  51. Deringer JR, Ely RJ, Stauffacher CV, Bohach GA. Subtype-specific interactions of type C staphylococcal enterotoxins with the T-cell receptor. *Mol Microbiol* 1996;22:523–534.
  52. Pace CN. Evaluating contribution of hydrogen bonding and hydrophobic bonding to protein folding. *Methods Enzymol* 1995;259:538–554.
  53. Plemper RK, Wolf DH. Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem Sci* 1999;24:266–270.
  54. Ciechanover A, Orian A, Schwartz AL. Ubiquitin-mediated proteolysis: biological regulation via destruction. *Bioessays*, 2000;22: 442–451.
  55. Luyrink L, Gabriel CA, Thompson SD, Grom AA, Maksymowyc WP, Choi E, Glass DN. Reduced expression of a human V beta 6.1 T-cell receptor allele. *Proc Natl Acad Sci USA* 1993; 90: 4369–4373.
  56. Bjorkman PJ. MHC restriction in three dimensions: a view of T cell receptor/ligand interactions. *Cell* 1997. 89:167–170.