

## The Role of Structure in Antibody Cross-reactivity Between Peptides and Folded Proteins

Lisa Craig<sup>1</sup>, Paul C. Sanschagrin<sup>2</sup>, Annett Rozek<sup>1</sup>, Steve Lackie<sup>3</sup>  
Leslie A. Kuhn<sup>2\*</sup> and Jamie K. Scott<sup>4\*</sup>

<sup>1</sup>Department of Chemistry  
Institute of Molecular Biology  
and Biochemistry, Simon Fraser  
University, Burnaby, BC  
Canada V5A 1S6

<sup>2</sup>Protein Structural Analysis  
and Design Laboratory  
Department of Biochemistry  
Michigan State University  
East Lansing, MI 48824, USA

<sup>3</sup>Säpidyne Instruments Inc.  
967 E. ParkCenter Blvd. 445  
Boise, ID 83706-6700, USA

<sup>4</sup>Biochemistry Program and  
Department of Biological  
Sciences, Institute of Molecular  
Biology and Biochemistry  
Simon Fraser University  
Burnaby, BC, Canada V5A 1S6

Peptides have the potential for targeting vaccines against pre-specified epitopes on folded proteins. When polyclonal antibodies against native proteins are used to screen peptide libraries, most of the peptides isolated align to linear epitopes on the proteins. The mechanism of cross-reactivity is unclear; both structural mimicry by the peptide and induced fit of the epitope may occur. The most effective peptide mimics of protein epitopes are likely to be those that best mimic both the chemistry and the structure of epitopes. Our goal in this work has been to establish a strategy for characterizing epitopes on a folded protein that are candidates for structural mimicry by peptides. We investigated the chemical and structural bases of peptide-protein cross-reactivity using phage-displayed peptide libraries in combination with computational structural analysis. Polyclonal antibodies against the well-characterized antigens, hen egg-white lysozyme and worm myohemerythrin, were used to screen a panel of phage-displayed peptide libraries. Most of the selected peptide sequences aligned to linear epitopes on the corresponding protein; the critical binding sequence of each epitope was revealed from these alignments. The structures of the critical sequences as they occur in other non-homologous proteins were analyzed using the Sequery and Superpositional Structural Assignment computer programs. These allowed us to evaluate the extent of conformational preference inherent in each sequence independent of its protein context, and thus to predict the peptides most likely to have structural preferences that match their protein epitopes. Evidence for sequences having a clear structural bias emerged for several epitopes, and synthetic peptides representing three of these epitopes bound antibody with sub-micromolar affinities. The strong preference for a type II  $\beta$ -turn predicted for one peptide was confirmed by NMR and circular dichroism analyses. Our strategy for identifying conformationally biased epitope sequences provides a new approach to the design of epitope-targeted, peptide-based vaccines.

© 1998 Academic Press

**Keywords:** antibody cross-reactivity; epitope mimicry; phage-displayed peptide libraries; peptide structural analysis; peptide vaccines

\*Corresponding authors

Abbreviations used: Ab, antibody; HEL, hen eggwhite lysozyme; MHr, myohemerythrin; DTT, dithiothreitol; BSA, bovine serum albumin; TBS, Tris-buffered saline; PDB, Brookhaven Protein Data Bank; SSA, Superpositional Structural Assignment; ELISA, enzyme-linked immunosorbent assay; IC<sub>50</sub>, inhibitory concentration required to reduce ELISA signal to 50%; RMSD, root-mean-square deviation; NMR, nuclear magnetic resonance; CD, circular dichroism; HIV-1, human immunodeficiency virus type 1; DQF-COSY, double-quantum-filtered correlated spectroscopy; ROESY, rotating-frame Overhauser enhancement spectroscopy; NOE, nuclear Overhauser enhancement; TPPI, time-proportional phase incrementation; single-letter amino acid code: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine; X, any natural amino acid or a non-critical amino acid.

E-mail addresses of the corresponding authors: jkscott@sfu.ca; kuhn@agua.bch.msu.edu

## Introduction

Antibody (Ab) responses against folded proteins include reactivities against linear and discontinuous epitopes. Alanine mutagenesis studies (Jin *et al.*, 1992) and other work have shown that the reactivities against linear epitopes represent a small fraction of the Ab response compared to those against discontinuous ones. Thus, the linear epitopes identified by mapping studies using synthetic peptides (Milton & Van Regenmortel, 1979; Atassi 1984; Geysen *et al.*, 1987) and phage-displayed peptide libraries (Folgori *et al.*, 1994; Yao *et al.*, 1995; Bonnycastle *et al.*, 1996), though minor, are significant in that they are the sites that typically cross-react well with peptides. Since the idea of peptide-based vaccines was first proposed (Lerner, 1982), much effort has focused on identifying peptides that will elicit Abs that cross-react with pre-specified sites on a target protein. Although a number of such peptides have been reported (Arnon *et al.*, 1971; Francis *et al.*, 1987; Hastings *et al.*, 1990; Javaherian *et al.*, 1990; Keller *et al.*, 1993; White-Scharf *et al.*, 1993; van der Werf *et al.*, 1994; Motti *et al.*, 1994; Robinson *et al.*, 1995; Lundkvist *et al.*, 1996), it remains difficult to predict the epitopes on folded proteins that will make effective targets for peptide-based vaccines. More often than not, Abs raised against peptides bind with high affinity to the peptide, but only weakly to the folded protein from which they were derived (Jemmerson, 1987).

The linear epitopes identified by cross-reactivity with short peptides are characterized by defined structures of relatively high flexibility (Tainer *et al.*, 1984; Westhof *et al.*, 1984), surface accessibility (Atassi & Lee, 1978; Milton & Van Regenmortel, 1979) and hydrophilicity (Hopp & Woods, 1981); these features usually correspond to turns and loops in folded proteins (Westhof *et al.*, 1984). These parameters have been used with some success to predict the regions of proteins that will cross-react with peptides (Thornton *et al.*, 1986; Pellequer *et al.*, 1991). Presumably, shape complementarity between an epitope and the antigen-binding site of an Ab is more readily induced in mobile accessible regions of a protein than in more rigid ones; however, as yet there are no structures of Abs in complex with linear epitopes available to support this. Evidence for an induced-fit binding mechanism for linear epitopes has been obtained by Getzoff *et al.* (1987), who identified the critical binding residues of several cross-reactive epitopes using peptides bearing single amino acid replacements. Several epitopes contained critical residues that were buried in the native protein, suggesting that the structure of these epitopes was altered upon binding Ab. Thus, epitope flexibility has functional significance in peptide cross-reactivity.

While epitope flexibility is clearly involved in peptide cross-reactivity, the role of conformational bias has also been recognized (Dyson *et al.*, 1988a). Although most short peptides possess considerable conformational flexibility in aqueous solution,

sequences have been identified that confer structural stability to a peptide, as assessed by nuclear magnetic resonance (NMR) and circular dichroism (CD) analyses (Dyson *et al.*, 1988b). Such sequences are expected to form largely the same structures in the context of a folded protein or a peptide, thereby allowing antigen cross-reactivity. In support of this, conformational preferences for turns and helices in solution have been demonstrated in a number of antigenic peptides (Dyson & Wright, 1995). Structural elements that are present in peptides in solution have also been demonstrated in the corresponding folded protein (Wien *et al.*, 1995), and in peptide-Ab complexes (Tsang *et al.*, 1992; Rini *et al.*, 1993; Ghiara *et al.*, 1994; Campbell *et al.*, 1997). The role of structural stability in peptide mimicry of protein epitopes is further illustrated by studies demonstrating improved cross-reactivity (Luzzago *et al.*, 1993; Hoess *et al.*, 1994) and the production of higher-affinity, protein-cross-reactive Abs (Arnon *et al.*, 1971; Satterthwait *et al.*, 1989; Zhong *et al.*, 1994; Cuniasso *et al.*, 1995) when disulfide and other covalent constraints are incorporated into peptides. Thus, both structural mimicry by a peptide and flexibility of the corresponding epitope contribute to cross-reactivity and the production of cross-reactive Abs.

The computational strategy, Sequery, was developed as a statistical screen of the Protein Data Bank (PDB; Bernstein *et al.*, 1977; Abola *et al.*, 1987), to determine whether an amino acid sequence, or sequence pattern, possesses a local conformational preference independent of its protein context. Sequery has been successfully applied to model protein-targeting motifs (Collawn *et al.*, 1991; Chang *et al.*, 1993) as confirmed by NMR (Bansal & Gierasch, 1991; Eberle *et al.*, 1991), and to model loop structures for crystallographic fitting (Parge *et al.*, 1993) and homology modeling (Fisher *et al.*, 1994). We employed Sequery, and an accessory program, Superpositional Structural Assignment (SSA), to identify conformational preferences in linear epitope sequences that are known to cross-react with peptides. Peptides bearing these conformationally preferred epitope sequences are expected to mimic the epitope structure, thus facilitating recognition by anti-protein Abs. The cross-reactive epitopes used for this analysis were identified by screening a panel of phage-displayed peptide libraries with polyclonal Abs against the model proteins hen eggwhite lysozyme (HEL) and worm myohemerythrin (MHR). For several of the epitope sequences, structural preferences emerged that matched the structures of the corresponding epitopes on HEL or MHR. An epitope sequence bearing a strong Sequery-predicted bias for a type II  $\beta$ -turn was shown to have the same well-defined turn preference when analyzed as a free peptide in solution by NMR and CD spectroscopy. Moreover, synthetic peptides predicted to bear structurally biased sequences were shown to bind Ab with sub-micromolar affinities. Thus, we have shown that short, cross-reactive sequences on linear epi-

topes have conformational preferences that allow peptides to functionally mimic epitope structures. The ability to identify, from a set of cross-reactive epitopes, those that have the highest structural propensities, should prove valuable in the design of peptide-based vaccines.

## Results

### Cross-reactive peptides reveal linear epitopes on native HEL and MHR

To identify peptides that bind Abs against folded proteins, we screened a panel of linear and conformationally biased peptide libraries with polyclonal anti-HEL and anti-MHR Abs. The random peptides in the majority of the libraries were fused to the N terminus of the major coat protein pVIII of filamentous phage, whereas the peptides in two of the libraries were fused to the minor coat protein pIII. Three to four rounds of screening were performed on the library panel, and five to 20 clones were analyzed from each of the phage pools showing high enrichment and/or strong ELISA signals. As shown in Table 1A and B, consensus sequences were revealed by aligning similar peptide sequences into consensus groups. Linear epitopes on HEL and MHR were subsequently identified by aligning consensus sequences to the amino acid sequence of the corresponding antigen. From this, five epitopes comprising the following sequences were identified on each antigen: DVQ, DITASV, NAWV, RTPGS, and TQATNR on HEL, and KYSEV, PEPYV, WDESF, GLSAPVDA, and APNLA on MHR. Two "sub-epitopes" were identified for the KYSEV epitope; KYSE from the pVIII library screening and YSEV from the pIII library screening. This probably reflects small differences in the anti-MHR response, as the pIII and pVIII libraries were screened with anti-MHR Abs from different rabbits (see Materials and Methods).

In a few cases, the alignment of a peptide to a single consensus group was not clear; either residue patterns common to two consensus groups were present, or the peptide had only two residues in common with a consensus group. Clone-against-clone competition ELISAs were performed to determine correct alignments; a peptide was assigned to a given consensus group if its corresponding phage clone could inhibit Ab binding to an immobilized clone whose displayed peptide aligned well to this group. The degree to which selected phage clones inhibited Ab-binding to representative immobilized phage is shown in Table 1A and B. This provided a functional assay for assigning peptides to a consensus group.

For peptides containing multiple cysteine residues, we investigated whether constraints imposed by disulfide bridging were required for Ab-binding. Reducing ELISAs were performed, in which binding to immobilized clones was assayed in the presence of 5 mM dithiothreitol (DTT), after reduction of disulfide bridges with 15 mM DTT.

The level of inhibition by DTT is shown in Table 1A and B. Peptides in the DVQ and KYSEV consensus groups were most dramatically affected by DTT reduction; binding of Ab was inhibited considerably for clones containing multiple cysteine residues. In both these groups, however, at least one strong-binding clone was derived from an unconstrained library, suggesting that the requirement for disulfide-bridging depends on the sequence of the peptide. Only the weak-binding clones in the DITASV group were affected, and little or no effect was seen on peptides in the PEPYV group. Thus, the presence of disulfide bonds can contribute to binding for many of the peptides, but is not an absolute requirement for cross-reactivity with any of the epitopes identified.

To ensure that the cross-reactive peptides were not isolated by minor Ab reactivities against denatured or proteolyzed forms of the HEL and MHR antigens, competition ELISAs were performed using HEL and MHR in solution. Both antigens were in their native, folded form, as each of their electrophoretic mobilities corresponded to single, discrete bands on a non-denaturing polyacrylamide gel. Table 1A and B shows that binding to the majority of clones was strongly inhibited ( $\geq 75\%$  inhibition) by their cognate antigen. Weak inhibition (less than 25%) was observed for a few clones, particularly those having relatively high affinities. In these cases, the Abs may have bound more tightly to the phage clone than to cognate antigen. The alternative, that these clones were selected by Abs specific for the unfolded protein, is unlikely, since other clones in the same consensus group (and in the same assays) were inhibited by native antigen.

### Binding of synthetic peptides to anti-HEL and anti-MHR Abs

To determine if free peptides behave similarly to the phage-displayed peptides, several peptides were synthesized and tested for their ability to bind Ab. Four biotinylated peptides, each derived from a tight-binding clone in a major consensus group, were synthesized as follows: Cys6-7 from the DVQ group; Cys5-1 from the DITASV group, X6-1 from the RTPGS group and X6-14 from the PEPYV group (Table 1A and B; see Materials and Methods for peptide sequences). All of the peptides were shown to bind strongly to Ab by direct ELISA and to inhibit binding of Ab to the corresponding clones in a dose-dependent manner (data not shown). Average equilibrium dissociation constants ( $K_d$  values) were obtained for each of the four peptides using the KinExA automated immunoassay instrument (Säpidyne Instruments, Boise ID). The KinExA instrument is a computer-controlled flow spectrofluorimeter designed to achieve the rapid separation and quantification of uncomplexed Ab present in reaction mixtures of Ab, antigen, and Ab-antigen complexes (Blake *et al.*, 1997). As shown in

Table 2, all four peptides bound Ab with sub-micromolar  $K_d$  values. Binding curves for each of the four peptides are shown in Figure S1 of Supplementary Material. Taken together, the data indicate that the Abs bind with moderate affinities to these peptides, whether they are displayed as pVIII-fusions on phage, or are free in solution.

### Assignment of critical epitopes on HEL and MHR

Each epitope on HEL and MHR was mapped by matching conserved residues in each consensus sequence group to a sequence within the cognate protein antigen. The relative importance of each residue for cross-reactivity was further deduced from its pattern of amino acid replaceability within the consensus sequence group. Residues that were irreplaceable, or could only be conservatively replaced, were considered "critical binding resi-

dues". These residues are likely to contribute significantly to binding energy by interacting directly with the Ab combining site, and/or by contributing to the overall structure of the epitope (Getzoff *et al.*, 1987). Based on this analysis, the critical binding sequences, or "critical epitopes", for HEL and MHR are DVQ, DITXXV, NAW, RXPGS, KYSE (pVIII), YSEV (pIII), EPYV and WDXSF, in which X represents a non-critical residue in the sequence. Table 1A and B shows the critical sequences for the epitopes whose consensus groups contain at least four peptides.

Figure 1(a) and (b) shows the locations of the linear epitopes on the HEL and MHR structures, respectively. Each of the epitopes is well exposed on the surface of the protein, with the exception of NAW, which lies on the surface of HEL, but at the opening of the substrate-binding cleft. Most of the epitopes comprise turns or loops between regions of regular secondary structure. In HEL, the RTPGS

**Table 1.** Consensus groups and alignments

HEL epitope and aligning phage clones <sup>a</sup>	Peptide sequence and alignment with HEL	Direct ELISA <sup>b</sup>	Inhibition ELISAs-level of inhibition <sup>c</sup> in the presence of:		
			In-solution phage clone <sup>d</sup>	5 mM DTT	5 $\mu$ M HEL
<b>A. Consensus groups, sequence alignment with HEL, direct ELISA and inhibition ELISA data</b>					
HEL (113-127)	****NRCKG <b>TDVQAW</b> IRGC**** <sup>e</sup>			Cys6-5	
X15-5	HD <b>VQ</b> RELIQLQRYMP	0.649	+	-	-
Cys6-7	VNIACNP <b>TDIQ</b> CLIRL	0.602	++	+++	+
Cys6-6	GPYTCTPN <b>DIQ</b> CRVRI	0.543	++	+++	+
Cys6-5	ND <b>VQ</b> CRLRSHQCETMR	0.274	++	++	+
Cys4-5	ERMSCDS <b>QRQ</b> QKAM	0.225		+++	+++
Cys6-3	GNSHCDAN <b>DVQ</b> CNTHN	0.179		+++	++
LX6-5	DCFD <b>DVQ</b> RCV	0.174		+++	++
X30-6	QPQSSSGRPQ <b>DVQ</b> SIARE ( $Z_{13}$ ) <sup>f</sup>	0.186			++
X30-1	ER <b>DVQ</b> AHMRPH ( $Z_{19}$ )	0.131			++
Critical sequence	<b>DVQ</b>				
HEL (83-95)	****LLSS <b>DITASV</b> NC****			Cys6-1	
Cys5-1	NQPS <b>CDITAC</b> LTPQ	0.530	+++	+	+++
LX6-2	QCHAD <b>ITKCL</b>	0.479	+++	+	+++
LX6-3	ACNAD <b>IT</b> SCL	0.447		+	+++
LX8-2	QCAP <b>DVTNS</b> ACF	0.377	+++	+++	+++
LX6-1	TCP <b>DITRCL</b>	0.349		-	+++
X8CX8-6	TDCN <b>DITAC</b> FGDHFQIM	0.292		+++	+++
Cys6-1	MNDN <b>CLGGDITCF</b> SQ	0.264	++	+++	+++
Cys5-2	VEPTCAD <b>ATA</b> CLSR	0.216		+++	+++
X30-2	( $Z_{17}$ ) DVAD <b>ITTF</b> LHPGS	0.102	++		+++
Critical sequence	<b>DITXXV</b>				
HEL (104-111)	****GM <b>NAW</b> AW****			X8CX8-1	
X30-7	<b>NAW</b> ELTFGPYHIPSDC	0.337	+++		+++
X30-3	<b>NSWA</b> ( $Z_{13}$ ) HAP <b>GS</b> FF ( $Z_6$ )	0.254	+++		++
X6-5	<b>NAW</b> YLR	0.203			+++
X6-2	<b>NAW</b> ATH	0.200			+++
X8CX8-1	<b>NSW</b> HDHVSCVAMCLRPP	0.111	++	-	++
Critical sequence	<b>NAW</b>				
HEL (65-74)	****ND <b>GRT</b> PGSRN****			X6-4	
X6-1	<b>SKTPGA</b>	0.254	+++		++
X6-4	<b>GRFPGS</b>	0.197	++		++
Critical sequence	<b>RXPGS</b>				
HEL (38-50)	****FN <b>TQATNR</b> NTDGS****			X8CX8-2	
LX8-1	SCV <b>TQASNR</b> GCCL	0.318	+++	++	+++
Cys4-1	MTEPCMG <b>QYCN</b> RFR	0.208	-	+++	+++
X8CX8-2	<b>MATNR</b> VYGCQHSMYARS	0.133	+++	-	+++
Cys3-4	SRNDR <b>CTD</b> LCSAVK	0.100			++

Table 1 (continued)

MHR epitope and aligning phage clones <sup>a</sup>	Peptide sequence and alignment with MHR	Direct ELISA <sup>b</sup>	Inhibition ELISAs-level of inhibition <sup>c</sup> in the presence of:		
			In-solution phage clone <sup>d</sup>	5 mM DTT	5 μM MHR
B. Consensus groups, sequence alignment with MHR, direct ELISA and inhibition ELISA data					
MHR (63-74)	****DA <b>AKYSEV</b> PHK**** <sup>e</sup>		X15-9		
X15-5	CPTSSLWMD <b>AKSSDL</b>	0.602		+	++
X8CX8-1	QY <b>QYSE</b> TCIQSATAAH	0.317		+++	-
X15-9	TSGMAHD <b>KYSD</b> SPIW	0.270	++	-	+++
X8CX8-13	TF <b>AKYSE</b> TCVPVTMSTL	0.236		+	++
LX6-12	DC <b>KYSG</b> SPCV	0.175	++	+++	+
X15-3	ETS <b>KYST</b> PPIVRBBBT <sup>f</sup>	0.134			+++
X8CX8-9	APSGDSDCG <b>PSGK</b> WSC	0.131		++	+
Cys4-4	EKN <b>NCLK</b> W <b>SCA</b> APM	0.101			++
Critical sequence	<b>KYSE</b>				
pIII-X6:1	<b>FRYSEV</b>	0.404			
pIII-X6:8	<b>WKYSEV</b>	0.378			
pIII-X15:28	GARIM <b>YSEV</b> PGYLAV	0.342			
pIII-X6:3	<b>FMYSEV</b>	0.247			
pIII-X6:16	<b>PFYSEV</b>	0.181			
pIII-X15:26	PGFA <b>YSEAH</b> PFASLP	0.135			
pIII-X6:4	<b>YSEFMY</b>	0.107			
Critical sequence	<b>YSEV</b>				
MHR (2-11)	****WEI <b>PEPYV</b> WD****		X6-8		
X6-14	<b>QEPYV</b> L	0.906	+++	-	+++
LX8-2	DCPVGA <b>EPYV</b> CL	0.739		-	+++
X6-8	SDV <b>EPY</b>	0.714	+++	-	+++
X15-11	FENY <b>TQLGIPEPYV</b> L	0.706	+++	-	+++
X15-13	SGQSS <b>PELY</b> VARAEY	0.681	++	-	+++
LX6-7	TC <b>EPYI</b> IKCT	0.661		-	+++
LX8-1	HCLPSQ <b>EMYT</b> CM	0.629		+	++
LX6-1	FC <b>EPYV</b> GNCL	0.618		-	+++
Cys6-1	MKTACTG <b>ERK</b> YTCVPTI	0.522	+++	+	++
αCT-2	SCCTQP <b>CA</b> YVTC	0.490		+	++
X6-3	<b>VPEPYA</b>	0.447		-	+++
X6-10	<b>VPEKYV</b>	0.446			+++
pIII-X6:7	<b>PEAYVI</b>	0.148			
Critical sequence	<b>EPYV</b>				
MHR (8-18)	****YV <b>WDES</b> FRV <b>FY</b> ****				
pIII-X15:30	GSST <b>FDAS</b> FWPCCD	0.564		-	+++
pIII-X15:36	GK <b>YDLS</b> FTMVVSHS	0.422			
pIII-X15:38	LA <b>FDAS</b> FTVSS	0.347			
pIII-X15:23	PP <b>WDIS</b> FSV <b>LG</b> SDLL	0.111			
Critical sequence	<b>WDXSF</b>				
MHR (84-94)	****IG <b>GLS</b> APVDA <b>AKN</b> ****				
pIII-X15:27	DLS <b>GLW</b> APGGTDHLN	0.408		-	+++
pIII-X6:4	PGLLS <b>ALC</b> AP <b>GLA</b>	0.204			
MHR (35-47)	****CIRDNS <b>APNLA</b> T <b>L</b> ****				
pIII-X15:24	PGLLS <b>ALC</b> AP <b>GLA</b>	0.204			

<sup>a</sup> Phage clones from the pIII-displayed peptide libraries are indicated; all others are from the pVIII libraries. The pIII sequences are preceded by the N-terminal sequence ADGA.

<sup>b</sup> ELISA signals are reported in absorbance units,  $A_{405}-A_{490}$  minus the background signal from f88 or fd-tet control phage.

<sup>c</sup> Level of inhibition of Ab binding to immobilized clones produced by in-solution phage clones, free HEL or MHR, or 5 mM DTT: + + +, 75 to 100% inhibition; + +, 50 to 74%; +, 25 to 49%; -, <24%.

<sup>d</sup> Level of inhibition of binding by  $2 \times 10^{10}$  phage clones in solution.

<sup>e</sup> Underlined and bold-faced amino acids, critical binding residues; boldfaced (but not underlined) amino acids align with the corresponding HEL or MHR sequence or with other peptides in the consensus group, but are not critical, as determined by incomplete conservation within the consensus group.

<sup>f</sup> Z, Non-aligning residues in peptides from the X30 libraries.

<sup>g</sup> B, Residues not determined.

epitope forms a type II  $\beta$ -turn, DVQ and DITASV both occur at the beginning of  $\alpha$ -helices, and NAW forms part of a loop. In MHR, EPYV and WDESF comprise type VI and type I  $\beta$ -turns, respectively, in the N-terminal loop; KYSEV and GLSAPVDA

are located in loops connecting helices, where GLSAPVDA contains a type VI  $\beta$ -turn, and APNLA is at the beginning of a helix. Of all of these epitopes, only the TQATNR epitope in HEL is extended, forming a  $\beta$ -strand. Most of the epi-

**Table 2.** Average binding affinities and error limits for synthetic peptides

Antibody <sup>a</sup>	Peptide	$K_d$ (nM)	Error bounds <sup>b</sup>	
			$K_d$ low (nM)	$K_d$ high (nM)
Anti-HEL	Cys5-1	38.1	23.7	47.0
Anti-HEL	Cys6-7	64.5	36.4	104
Anti-HEL	X6-1	234	120	455
Anti-MHr	X6-14	780	586	1038

<sup>a</sup> Initial Ab concentrations were as follows: 20 nM anti-HEL for Cys5, 5 nM anti-HEL for Cys6-7, 50 nM anti-HEL for X6-1, and 100 nM anti-MHr for X6-14.

<sup>b</sup> The error bounds represent a 95% confidence interval (similar to  $\pm 2$  standard deviations) in the  $K_d$  estimate, and are a function of best fit error of the theoretical binding curve to the data.

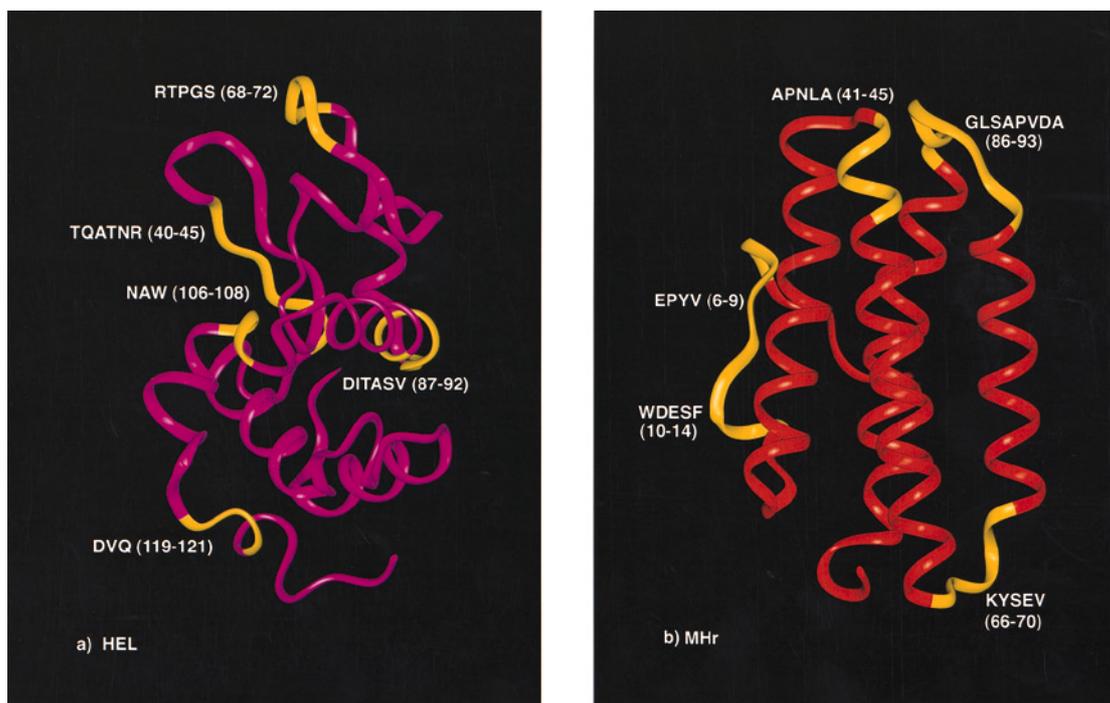
topes reside in regions of relatively high mobility and surface accessibility (data not shown; see Thornton *et al.*, 1986). Flexibility and accessibility are features typical of protein epitopes that cross-react with peptides (Westhof *et al.*, 1984; Tainer *et al.*, 1984; Thornton *et al.*, 1986; Geysen *et al.*, 1987).

#### Evidence for structural mimicry of protein epitopes by peptides

The sequences of critical epitopes were analyzed in three phases to assess the structural basis for mimicry by the peptides. In the first phase, the amino acid sequences of 493 non-redundant protein structures in the PDB (Hobohm *et al.*, 1992; see Materials and Methods) were scanned for occurrences of all or part of each critical epitope sequence. Four-residue sequence patterns were

used, each comprising three critical binding residues and one variable residue, since exact matches to more than three residues are rarely found in the PDB. Thus, for example, the critical epitope RXPGS was analyzed using the sequence patterns [R,K,H]XPG, XPGS and PGSX. Sequery provided a list of all sequence matches, or analogs, for each sequence pattern. The epitopes TQATNR on HEL and GLSAPVDA and APNLA on MHr were not included in the Sequery analysis, since only a few peptides aligned to each of these. At least one sequence pattern for each of the other epitopes yielded a substantial number of sequence analogs from the Sequery search. Sequence patterns yielding less than ten analogs were not considered to have given a sufficient statistical sample, and thus were not analyzed further.

In the second phase of the analysis, the structure of each sequence analog was compared with that



**Figure 1.** Ribbon structures of HEL and MHr showing the linear epitopes identified by peptide cross-reactivity. (a) HEL, PDB code 193l (M. C. Vaney, S. Maignan, M. Ries-Kautt & A. Ducruix, unpublished results). (b) MHr, PDB code 2mhr (Sheriff *et al.*, 1987). Epitopes are colored yellow.

**Table 3.** Summary of structural features for the HEL and MHR epitopes and the corresponding sequence patterns

Critical epitope on HEL or MHR	Frequency of analogs matching the epitope <sup>a</sup>	Overall conformational preference <sup>b</sup>	Epitope structure	Hydrogen-bonding pattern within the critical epitope	Surface accessibility of critical residues <sup>c</sup>
DVQ HEL: 119-121	DVQX 45% (9/20)	DVQX: helix 45% (9/20)	Beginning of helix (irreg.)	None	Exposed: D119, Q121 Buried: V120
RXPGS HEL: 68-72	XPGS 39% (13/33)	XPGS: type II $\beta$ -turn 39% (13/33)	TPGS: type II $\beta$ -turn	S72:O <sup><math>\gamma</math></sup> $\rightarrow$ T69:O	Exposed: R68, P70, G71 Buried: S72
DITXXV HEL: 87-92	DITX 25% (5/20)	DITX: irreg. 70% (14/20)	Beginning of helix (irreg.)	T89:N $\rightarrow$ D87:O <sup><math>\delta 1</math></sup> T89:O <sup><math>\gamma</math></sup> $\rightarrow$ D87:O <sup><math>\delta 1</math></sup>	Exposed: D87 Intermediate: T89 Buried: I88, V91
WDXSF MHR: 10-14	DXSF 29% (5/17)	DXSF: helix 29% (5/17)	DESF: type I $\beta$ -turn	S13:N $\rightarrow$ D11:O S13:N $\rightarrow$ D11:O <sup><math>\delta 1</math></sup> S13:O <sup><math>\gamma</math></sup> $\rightarrow$ D11:O <sup><math>\delta 1</math></sup> F14:N $\rightarrow$ D11:O	Intermediate: D11, S13 Buried: W10, F14
NAW HEL: 106-108	NA/SWX 0/20	NA/SWX: helix 50% (10/20)	NAWA: irreg. loop	None	Intermediate: N106, A107 Buried: W108
EPYV MHR: 6-9	EXYV 0/12	EXYV: helix 58% (7/12)	EPY: positions 2, 3 and 4 of a type VI $\beta$ -turn	V9:N $\rightarrow$ P7:O	Exposed: E6 Intermediate: P7, V9 Buried: Y8

<sup>a</sup> Number of sequence analogs with structures that superimpose on the corresponding HEL or MHR epitope with an RMSD  $\leq 0.75$  Å divided by the total number of sequence analogs for a given sequence pattern. Sequence analogs from proteins whose crystal structures were not well-resolved ( $\geq 2.6$  Å resolution) or whose backbones could not be superimposed due to missing atoms in the crystal structures, were not included in the analysis.

<sup>b</sup> An overall conformational preference was assigned to the sequence pattern if  $\geq 25\%$  of the analogs shared the same secondary structure. When this did not occur, the percentage of irregular structure is indicated.

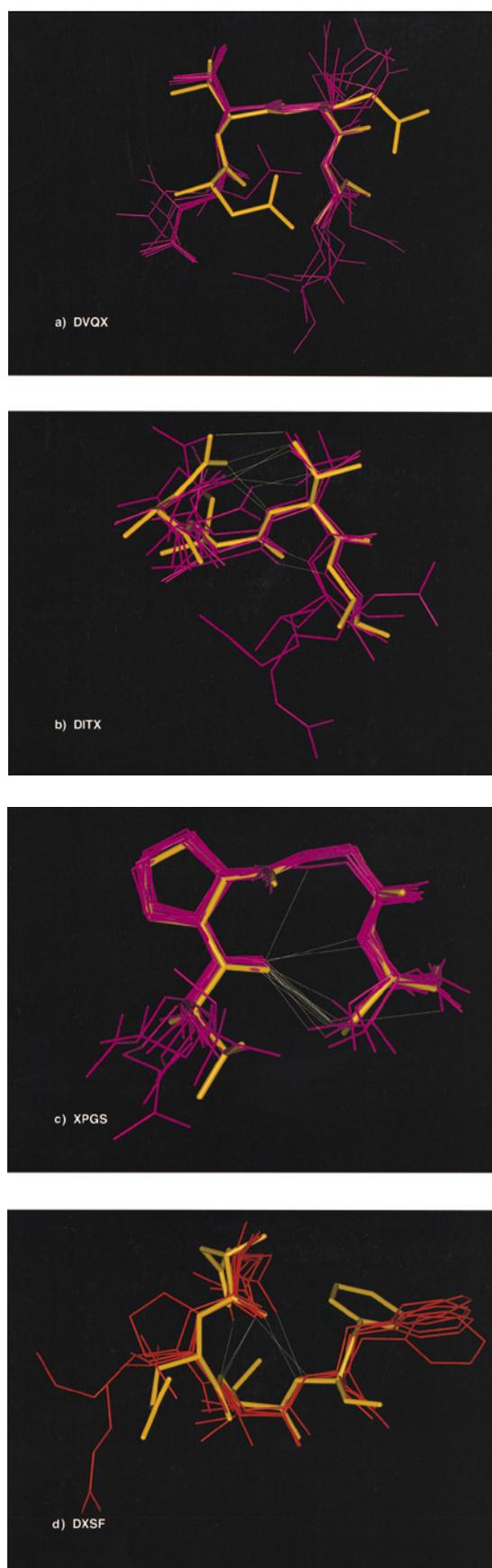
<sup>c</sup> Surface accessibility (Kabsch & Sander, 1983): buried, 0 to 40 Å<sup>2</sup>; intermediate, 41 to 80 Å<sup>2</sup>; exposed,  $>80$  Å<sup>2</sup>.

of the corresponding four-residue sequence within the appropriate epitope on HEL or MHR. The program SSA compared these structures by superimposing their backbone atoms and calculating the root-mean-square deviation (RMSD) between the positions of corresponding atoms in the two backbones. The structure of a given analog was considered to closely resemble an epitope if the superpositional RMSD between the two backbones was  $\leq 0.75$  Å. A sequence pattern was considered to have a conformational preference for the epitope structure if at least 25% of the analogs superimposed well on the epitope.

Based on this analysis, conformational preferences matching the corresponding epitope structures were detected for the sequence patterns, DVQX, DITX, XPGS, and DXSF as shown in Table S1 in Supplementary Material. The frequency of analogs matching the structure of the corresponding epitopes is indicated in Table 3; those with the highest frequency of superposition are assumed to have the strongest preference for the epitope conformation. These analogs are shown superimposed on their corresponding epitopes in Figure 2(a) to (d). In each case, conformational preference was highly dependent on the position of the variable residue, X, within the sequence pattern. For instance, nine of the 20 DVQX analogs superimposed well on the DVQA epitope in HEL, yet none of the XDVQ analogs superimposed well on the TDVQ site. This indicates that the conformation imposed by the critical binding sequence DVQ involves a fourth residue at its C terminus and not its N terminus. The fact that each critical

sequence occurs in the same structure in a number of non-homologous proteins suggests that local conformation is defined by local interactions within these sequences, rather than by tertiary interactions with the rest of the protein. Cross-reactive peptides bearing these sequences are likely to mimic the structures of the corresponding protein epitopes. Sequence patterns for the remaining epitopes did not yield analogs that superimposed well on the epitope structures.

In the third phase of the structural analysis, SSA was used to classify the structures of each epitope segment and the corresponding sequence analogs as turn, helix, extended ( $\beta$ -strand) or irregular, based on their superposition with model tetrapeptides having ideal secondary structures. An overall conformational preference for each sequence pattern was assigned if at least 25% of the analogs possessed the same structure. These data are summarized in Table 3 for the DVQ, RXPGS, WDXSF and DITXXV epitopes. Both the DVQX and DXSF patterns have helical conformational preferences, and the XPGS pattern has a type II  $\beta$ -turn preference. The conformational preference of the DITX pattern was defined as irregular, since neither the DITASV epitope, nor any of its analogs, superimposed well on any of the model tetrapeptides. Yet, five of the 20 DITX analogs superimpose well on DITA in HEL, and in all of these structures, the backbone of the aspartic acid is in an extended conformation, with the remainder of the sequence forming part of a turn (Figure 2(b)). Hence, the conformational preference of all four sequence patterns is considered turn-like. Helical conformation-



**Figure 2.** Superposition of sequence analogs on the corresponding epitopes. Only those analogs that superimpose on the epitope with an RMSD  $\leq 0.75$  Å are shown.

al preferences were also observed for the sequence patterns NAWX and EXYV (see Table 3). Yet these analogs did not superimpose well on their corresponding epitopes, and hence, were not investigated further as candidates for structural mimicry.

All of the epitope sequences that we analyzed were identified by cross-reactivity with peptides; a number of these showed a structural propensity. As a comparison, we wanted to determine the extent to which randomly chosen sequences would also show a structural propensity. A set of sequence patterns was randomly generated, and Sequery was used to identify sequence analogs from the list of non-redundant protein structures in the PDB. Ten of the sequence patterns yielded analog sets containing ten or more matches, as shown in Table S2 of Supplementary Material. Next, a pseudo-epitope was randomly chosen from each analog set, with the only requirement being that it be surface-exposed within its protein structure. SSA was used to compare the structure of each pseudo-epitope with its corresponding sequence analogs, and to assign secondary structures, as was done with the Ab-selected epitope sequences shown in Table 3. As with the Ab-selected sequence patterns, several of the randomly generated sequence patterns showed a clear propensity for a regular secondary structure. Furthermore, if the structure of a pseudo-epitope matched that of the overall structural propensity of the sequence pattern, a high proportion of the analogs also superimposed well on the pseudo-epitope. Thus, this analysis showed that the structural propensities of the epitopes we studied are similar to those of randomly chosen surface-exposed sites on proteins in general, and are not unique to peptide-cross-reactive sequences. Importantly, analysis by Sequery in combination with SSA has allowed us to target, from a set of peptide-cross-reactive epitopes, those whose structures are more likely to be mimicked by peptides.

### The molecular basis of conformational preferences

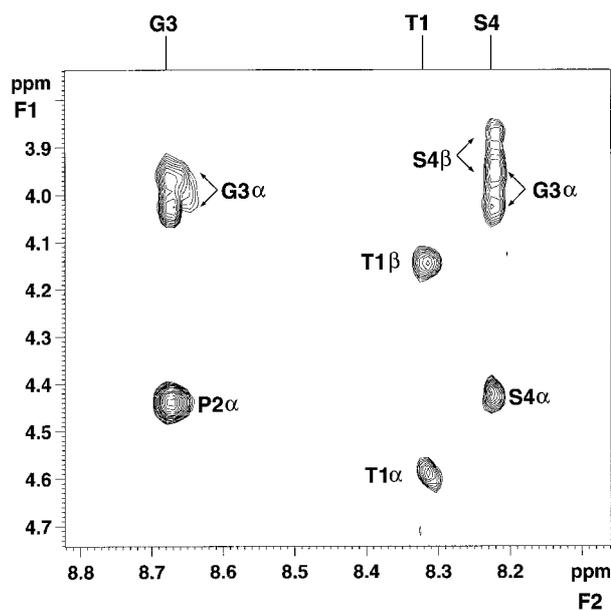
To understand the basis of locally induced structural preferences, we analyzed the inter-residue

The epitope tetramers are shown in yellow, the sequence analogs are shown in magenta for the HEL epitopes, and in red for the MHR epitope. Hydrogen bonds are indicated in white. (a) DVQX analogs superimposed on the DVQA epitope; PDB codes for the proteins bearing the sequence analogs are 1pox, 1pea, 2dnj, 7rsa, 1cns, 1smn, 1ash, 1ukz and 1qpg. (b) DITX analogs superimposed on the DITA epitope; PDB codes, 3cla, 1dpc, 8cat, 1rcb and 1cpt. (c) XPGS analogs superimposed on the TPGS epitope; PDB codes, 1slt, 1arv, 1mhl, 1cdo, 1fuj, 1pya, 1qpg, 1hxn, 1abr, 1eft, 2dnj, 8acn and 1arb. (d) DXSF analogs superimposed on the DESF epitope; PDB codes, 1pvd, 1gsa, 1gpc, 2ctc and 3pmg.

interactions within each epitope and within the corresponding structural matches. These interactions are summarized in Table 3, along with other relevant structural features. The epitopes and their analogs comprise turn-like conformations, which allow more stabilizing interactions within the tetramer than would occur in extended conformations. Specific hydrogen bonds appear to be important in stabilizing the structures of the DITASV, RTPGS, and WDESf epitopes and their analogs (Table 3 and Figure 2(b), (c) and (d)). The DITASV epitope and all five of its structural matches have a hydrogen bond between the O' of threonine and the O<sup>δ1</sup> of aspartic acid, as well as other hydrogen bonds; these interactions probably stabilize the helical-turn structure. Several of the hydrogen bonds present in the WDESf epitope are also seen in its structural matches. Finally, the hydrogen bond between the O' of serine and the carbonyl oxygen of threonine is present in the RTPGS epitope and in ten of the 13 XPGS analogs. Commonly,  $\beta$ -turns are stabilized by a hydrogen bond between the carbonyl oxygen at position 1 and the amide at position 4 of the turn (Venkatachalam, 1968). Atypical hydrogen-bonding in the XPGS analogs resulted in several of them being classified as irregular turns, since their structures are more open than classical type II  $\beta$ -turns. In addition to stabilization from hydrogen bonds, the type II  $\beta$ -turn preference suggested for the XPGS sequence pattern is promoted by proline and glycine, which often occur at positions 2 and 3, respectively, of type II turns (Wilmot & Thornton, 1988; Richardson & Richardson, 1990). Proline produces a kink in the polypeptide chain at position 2, due to its ring closure, and glycine is favored in position 3 because it lacks a  $\beta$ -carbon and therefore does not sterically interfere with the carbonyl oxygen of proline. These interactions likely contribute to the stability of the protein epitopes and to preferred structures for the cross-reactive peptides.

### TPGS has a type II turn conformational preference in solution

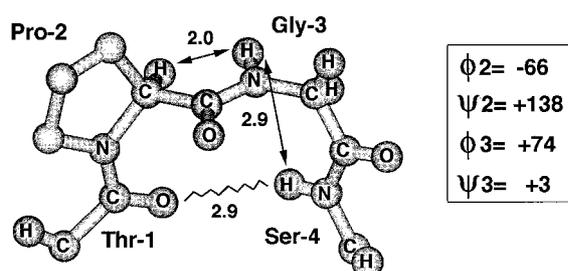
We predicted that peptides bearing conformationally biased sequences would have preferred structures in solution. To test this, the peptide Ac-TPGS-NH<sub>2</sub> was synthesized, and its structure was analyzed using NMR and CD spectroscopy. The TPGS sequence was selected for structural analysis over DVQA, DITA and DESf for several reasons: (i) the type II  $\beta$ -turn appeared to be a strong conformational propensity, with 39% of the XPGS sequence analogs possessing this structure; (ii) the TPGS turn in HEL and all of the XPGS analogs are stabilized by a hydrogen bond; and (iii) the peptides in the RXPGS consensus group were short and did not contain cysteine residues, suggesting that flanking residues were not necessary for stabilization of structure. The TPGS sequence was syn-



**Figure 3.** The H<sup>N</sup>-H<sup>α,β</sup> region of the <sup>1</sup>H ROESY spectrum of Ac-TPGS-NH<sub>2</sub> at 283 K showing intra- and inter-residue cross-peaks. The labels at the top of the Figure refer to the H<sup>N</sup> resonances, whereas the labels near the cross-peaks indicate the H<sup>α</sup> or H<sup>β</sup> resonances.

thesized as a blocked peptide in order to neutralize the termini and better mimic the natural state of this sequence within a protein or a longer peptide.

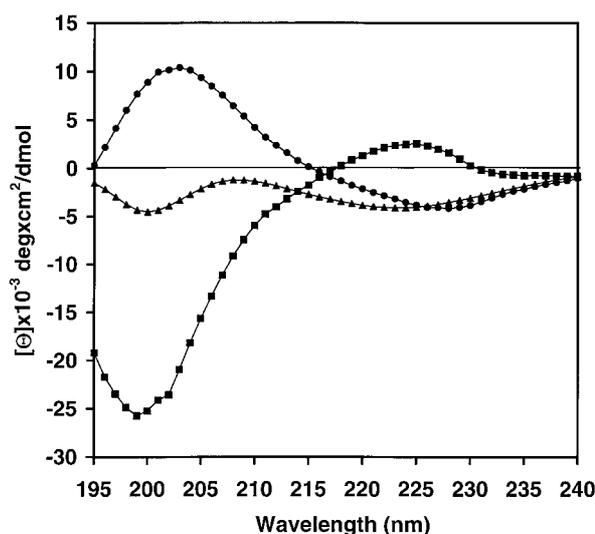
The structure of Ac-TPGS-NH<sub>2</sub> in aqueous solution was analyzed by <sup>1</sup>H NMR spectroscopy. Since the amino acid residues were unique, and sequential assignments followed unambiguously from spin system assignments, the proton resonances of Ac-TPGS-NH<sub>2</sub> were assigned directly from the DQF-COSY spectrum. Proton resonances are summarized in Table S3 of Supplementary Material. The *trans* conformation of the peptide was identified by the presence of strong cross-peaks between Thr1 H<sup>α</sup> and Pro2 H<sup>δ</sup> in the ROESY spectrum; no evidence for the *cis* conformation was found. The ROESY spectrum in Figure 3 shows a strong ROE between Pro2 H<sup>α</sup> and Gly3 H<sup>N</sup>. In addition, a medium ROE was observed between Gly3 H<sup>N</sup> and Ser4 H<sup>N</sup> (not shown), indicating a significant population of type II  $\beta$ -turn (Wagner *et al.*, 1986). The weak connectivity between Pro2 H<sup>α</sup> and Ser4 H<sup>N</sup>, which is also expected for type II  $\beta$ -turns, was not observed; this may be due to overlap of the H<sup>α</sup> resonances of Pro2 and Ser4. The temperature coefficient of Ser4 H<sup>N</sup> is -6.5 ppb/K, which is smaller than the temperature coefficient observed for Thr1 H<sup>N</sup> and Gly3 H<sup>N</sup>, -9.7 and -9.3 ppb/K, respectively; this suggests that the Ser4 backbone amide proton participates in an intramolecular hydrogen bond. Structural calculations were performed on the unblocked TPGS peptide. An ensemble of 50 peptide structures was calcu-



**Figure 4.** Calculated average structure of the TPGS peptide. Interproton distances (arrows) and the hydrogen bond distance (wavy line) between the serine amide and the threonine carbonyl oxygen are shown in Å. The backbone torsion angles of Pro2 and Gly3 are indicated in degrees.

lated using 17 ROE-based distance restraints and one hydrogen-bond distance restraint between Ser4 and Thr1, which was derived from the lowered Ser4  $H^N$  temperature coefficient (see Table S4, Supplementary Material). The distance restraint violations in the calculated structures were below 0.1 Å. The RMSD of all calculated structures to the average conformation was 0.61 Å for backbone atoms and 1.12 Å for all atoms. The backbone torsion angles of the average structure,  $\phi_2 = -66^\circ$ ,  $\psi_2 = +138^\circ$  and  $\phi_3 = +74^\circ$ ,  $\psi_3 = +3^\circ$  correlate well with those of an ideal type II  $\beta$ -turn, which has torsion angles of  $\phi_2 = -60^\circ$ ,  $\psi_2 = +120^\circ$  and  $\phi_3 = +90^\circ$ ,  $\psi_3 = 0$  (Richardson, 1981). The backbone RMSD for the superposition of the calculated TPGS average structure and the TPGS epitope in HEL (residues 69 to 72) was 0.44 Å, indicating close structural similarity for the two backbones. A ball-and-stick model of the calculated average conformation is shown in Figure 4, showing the interproton distances and the hydrogen bond between the serine amide and the threonine carbonyl oxygen. A hydrogen bond between the serine hydroxyl and the threonine carbonyl, as is seen in the TPGS epitope (Figure 2(c)), was not detected in the NMR spectra.

CD analysis of Ac-TPGS-NH<sub>2</sub> was performed to corroborate the NMR findings. The CD spectrum of Ac-TPGS-NH<sub>2</sub> in water, shown in Figure 5, is highlighted by two minima at 200 nm and 223 nm. Deconvolution of the CD spectrum by convex constraint analysis (Perczel *et al.*, 1991) yielded two pure component curves representing 55% (curve 1) and 35% (curve 2) of the peptide conformation. Curve 1 has a maximum at 203 nm characteristic of a type II  $\beta$ -turn conformation, whereas curve 2 has a minimum at 199 nm and was assigned as atypical or random secondary structure (Holloosi *et al.*, 1987; Perczel *et al.*, 1993). Thus, the NMR and CD results for the Ac-TPGS-NH<sub>2</sub> peptide are consistent with a type II  $\beta$ -turn, as predicted by Sequery/SSA and observed in the HEL epitope.



**Figure 5.** CD spectrum and pure component curves of Ac-TPGS-NH<sub>2</sub> in aqueous solution at 278 K. Observed CD spectrum (triangles); pure component curve 1 (circles) reflecting a type II  $\beta$ -turn; and pure component curve 2 (squares) reflecting an atypical or random secondary structural element.

## Discussion

### Peptide library screening in combination with computational and structural analyses identifies epitope sequences having inherent structural propensities

The goal of this work was to establish a means of identifying immunogenic sites on proteins that can be structurally mimicked by peptides. We characterized a set of linear epitopes on HEL and MHR based on their cross-reactivity with phage-displayed peptides. The critical binding sequences within the epitopes were defined from the patterns of conserved residues within the selected peptides, and were assessed for their degree of intrinsic structural bias using the Sequery and SSA computer algorithms. Turn-like structures for the sequences DVQX, XPGS, DITX, and DXSF were conserved between each epitope and a number of non-homologous proteins, suggesting that these linear epitopes have inherent conformational preferences. Analysis of the structures of the epitopes and the matching sequence analogs showed similar hydrogen-bonding patterns that could potentially stabilize these conformations. A type-II  $\beta$ -turn preference was confirmed for the RTPGS epitope by NMR and CD analysis of the synthetic peptide Ac-TPGS-NH<sub>2</sub>. Furthermore, synthetic peptides were made from selected clones from three of the consensus sequence groups predicted to have conformational preferences; these bound to Ab with sub-micromolar affinities, and inhibited Ab binding to immobilized phage bearing the same sequence. Thus, we have identified, from a set of linear epi-

topes, those whose structures and binding activities are mimicked by peptides.

Sequery, in combination with SSA, provides a means of identifying short sequences that have structural propensities, based on the structural similarity of a given sequence in the context of a number of non-homologous proteins. The molecular basis of these propensities could be understood by analyzing the interactions within the sequences in each protein structure. Thus, in our analysis of the peptide-cross-reactive epitopes in HEL and MHR, the presence of stabilizing features, such as intra-epitope hydrogen bonds that were also present in the analog set, further strengthened our confidence in a given epitope sequence having a structural propensity. Thus, the results of the Sequery, SSA and structural analyses were viewed together in determining epitope sequences that are likely to adopt defined conformations in the context of a peptide. We expect that these peptides will be more effective at eliciting Abs capable of recognizing the cognate antigen than peptides that do not mimic their epitope structures, and are in the process of testing this prediction.

Critical binding residues play a dual role in cross-reactivity; they impart structural stability, which allows recognition by Abs, and they provide energetically favorable binding contacts. Most likely, all peptides that cross-react with linear epitopes mimic the corresponding epitope structures to some degree. The structural biases we identified appear to be conferred mainly by hydrogen bonding and, in one case, a proline-induced backbone constraint (Table 3). Yet, the residues involved in promoting a given structure are not necessarily sufficient for recognition by Abs. In the case of the RTPGS epitope, although the TPGS sequence was important for structural mimicry, it was insufficient for Ab recognition, as the synthetic peptide Ac-TPGS-NH<sub>2</sub> did not inhibit Ab binding to phage from the TPGS consensus group, even at a concentration of 50 mM free peptide (L.C. & J.K.S., unpublished data). Thus, a basic residue N-terminal to the TPGS sequence appears to be required for binding (as seen by the phage ELISAs in Table 1A), but not for structure. In some cases, structural bias may require, or benefit from, further stabilization by other residues, such as cysteine. This was apparent from the variation in requirements for disulfide bridging within the different consensus groups (Table 1A and B). Additionally, residues other than critical binding residues can contribute to affinity through increased contacts with Abs.

### **Polyclonal anti-protein Abs favor peptide-mimics of linear epitopes**

All of the peptides selected in our polyclonal Ab screenings aligned to linear epitopes on HEL and MHR. This was surprising, since antibodies against discontinuous epitopes predominate in anti-protein responses (Benjamin *et al.*, 1984; Jin *et al.*, 1992),

and peptide-mimics of discontinuous epitopes, or "mimotopes", have been selected by others using polyclonal serum Abs (Folgori *et al.*, 1994; Meola *et al.*, 1995), and monoclonal Abs against discontinuous epitopes (Balass *et al.*, 1993; Felici *et al.*, 1993; Luzzago *et al.*, 1993; Stephen *et al.*, 1995; Bonnycastle *et al.*, 1996; L.C., J. Shen & J.K.S., unpublished data). The critical residues in mimotopes appear to be larger in number and more dispersed than in linear epitope mimics, and hence, they are likely to be present in smaller numbers in the libraries. This may explain why the clones that we select with monoclonal Abs against discontinuous epitopes are often weak binders (Bonnycastle *et al.*, 1996), which require optimization of residues flanking the consensus sequences to obtain moderate binding affinities (K. L. Brown, E. Leong & J.K.S., unpublished data). Our ultimate goal is to select peptides that not only cross-react with anti-protein Abs, but that also elicit protein-cross-reactive Abs; this is most likely to be successful if the affinities of the cross-reactive peptides are comparable to those of the proteins. Our results indicate that peptide-mimics of linear epitopes dominate in affinity-selections; thus, they may prove viable targets for peptide-based vaccine strategies.

### **Critical binding residues provide insight into the mechanism of Ab binding**

Four of the five MHR epitopes found in our study were identified by Geysen *et al.* (1987), who tested the reactivity of polyclonal anti-MHR sera against a panel of short, overlapping synthetic peptides representing the entire sequence of the antigen. Several of the most reactive peptides were remarkably similar to the linear epitopes found in our peptide-library screenings. Although the critical residues defined by Geysen *et al.* (1987) differed somewhat from our assignments, both studies revealed a requirement for aromatic residues in the cross-reactive peptides. Aromatic residues that could not be replaced by chemically similar amino acids were assumed to be directly involved in Ab binding (Getzoff *et al.*, 1987). The side-chains of these critical aromatic residues are buried in the protein epitope and thus are locally inaccessible (see Table 3 and Getzoff *et al.*, 1987). To explain how Ab might interact with these inaccessible residues in the protein epitope, Getzoff *et al.* (1987) proposed an induced-fit binding mechanism, in which Ab initially contacts the solvent-accessible residues on the epitope, then induces a conformational change exposing the hydrophobic side-chain for binding. Such changes may be limited to local movements exposing a buried side-chain, or may involve more significant backbone rearrangements.

Conformational changes of epitopes induced upon Ab binding are likely facilitated by their locations in flexible regions of the protein. However, even those epitopes with high mobility values have well-defined structures, since they can be

resolved crystallographically. Structural stability has been demonstrated for peptides that cross-react with linear epitopes, as shown here and by others (Dyson & Wright, 1995). These findings are consistent with a two-step binding mechanism, in which Abs initially recognize gross structural elements, followed by conformational adjustments that improve complementarity at the Ab/antigen interface (Tainer *et al.*, 1984; Friedman *et al.*, 1994). Thus, structural bias can play a role in cross-reactivity with peptides, even when an induced-fit binding mechanism is involved.

### **Cross-reactivity can occur with a minimal number of shared residues**

We and others have shown that peptides having only a few residues in common with a protein epitope can cross-react with that epitope. The idea that short peptides can cross-react with folded proteins has been challenged on the basis of crystallographic data (Laver *et al.*, 1990). X-ray crystal structures of Ab-protein complexes show 15 to 22 protein residues in contact with the combining site of an Ab (Davies *et al.*, 1988; Wilson & Stanfield, 1993; Braden & Poljak, 1995), and thus short peptides seem unlikely to provide sufficient contacts for binding. However, thermodynamic studies of Ab-protein interactions indicate that only five or six of the large number of protein residues contacting the antigen combining site actually contribute to the energetics of binding (Novotny, 1991). Moreover, the only Ab-protein complexes whose crystal structures have been solved involve discontinuous epitopes, in which the critical binding residues are spread over two or more segments of the protein. Peptides generally cross-react with linear epitopes on folded proteins, where the critical residues are located on a continuous segment of the protein (although additional residues on neighboring segments may make non-critical contacts). Thus, although the critical binding residues identified for the HEL and MHr epitopes probably represent only a subset of the contact residues in the protein epitope, they make major contributions to the binding energetics.

### **Cross-reactivity requires that the Ab combining site has shape complementarity with both the linear epitope and the peptide**

It has been suggested that all epitopes are discontinuous to some extent, and that short peptides that cross-react with native antigen only mimic part of the epitope (Barlow *et al.*, 1986). For this to occur, the critical binding residues must be located on a single segment of the protein that protrudes somewhat, so that these residues will make extensive contacts with the Ab combining site. The locations of most of the linear epitopes identified here correspond to regions of high mobility and accessibility. Such regions tend to have high protrusion indices (Barlow

*et al.*, 1986; Thornton *et al.*, 1986), which allow Ab to make central, critical contacts with a continuous segment of the protein. Abs that bind to such sites, as well as to cross-reactive peptides, must have unique binding sites that will accommodate the epitopes on both antigens.

The structure of the Ab combining site is determined to a large degree by the surface formed from its six hypervariable loops (L1-L3, H1-H3). Chothia & Lesk (1987) identified a set of backbone frameworks, or canonical structures, which describe much of the structural variation in five of the six hypervariable loops (all except H3). The combinations of canonical structures used by Abs against a variety of antigens including proteins, peptides, carbohydrates and haptens has been analyzed (Vargas-Madrado *et al.*, 1995). This study showed that anti-peptide Abs use a restricted set of canonical structure combinations as compared to anti-protein Abs; this restriction is reflected in variable-gene usage (Lara-Ochoa *et al.*, 1996). This group and others (Wilson & Stanfield, 1993; Wilson *et al.*, 1995; MacCallum *et al.*, 1996) have characterized the combining sites of anti-peptide Abs as grooved; this shape allows extensive interaction with a small number of peptide residues. In contrast, the antigen binding site of Abs specific for discontinuous epitopes is flattened, to accommodate a larger surface area, and the residues of these epitopes are, on average, less buried within the Ab combining site than those of bound peptides.

The cross-reactivity observed here between peptides and linear epitopes is likely due to Abs whose combining sites resemble those of anti-peptide Abs. Evidence supporting this comes from crystallographic studies of two monoclonal Abs in complex with peptides (Tormo *et al.*, 1994; Wien *et al.*, 1995). Both Abs were raised against native or heat-inactivated virions, and recognize intact viral particles; they also cross-react with peptides, and their combining sites resemble those of anti-peptide Abs. Moreover, the linear epitopes they recognize protrude (Wien *et al.*, 1995), or are predicted to protrude (Tormo *et al.*, 1994) from the viral coat, which would allow them to be buried in the Ab combining site much like the peptides. Thus, in these cases, cross-reactivity is associated with a grooved, "anti-peptide-Ab-like" combining site.

### **Future studies**

An understanding of the molecular basis of protein-peptide cross-reactivity is critical in the design of vaccines that use peptides to target epitopes on folded proteins. We describe a means of identifying immunodominant, linear epitopes on protein antigens that are likely to be structurally mimicked by peptides. These peptides, when used as immunogens in conventional immunization strategies, may not elicit strong reactivity against a targeted site, since the combining-site structure of the anti-

peptide Abs elicited will typically be groove-like. As mentioned, even though the combining site of such Abs may be centrally configured to make favorable contacts with a protruding linear epitope, their ridged periphery may not fit with regions surrounding the epitope. Abs having a different combining-site structure from those elicited against the peptide may be required for cross-reactivity with the targeted site. Several prime-and-boost strategies, involving immunizations with whole protein and peptide, have successfully enhanced protein cross-reactivity (Emini *et al.*, 1983; Girard *et al.*, 1995; Davis *et al.*, 1997). These strategies involve priming with one antigen and boosting with the other, so as to enhance the production of cross-reactive Abs. These immunization strategies, using peptides identified in this study, should provide a substantial test of the viability of epitope-targeted vaccines. Ultimately, our ability to target Ab responses to pre-selected epitopes will depend on our understanding of all aspects of cross-reactivity, including the structures of the peptide, the targeted site and the Abs involved in the desired cross-reaction. This knowledge should allow us to develop a rational means for producing *in vivo* Abs having predetermined antigen specificity.

## Materials and Methods

### Materials

All of the IgGs and antisera used were derived from rabbits. MHR protein and anti-MHR Abs that had been affinity purified on MHR (Getzoff *et al.*, 1987) were provided by H. Alexander (University of Missouri-Columbia); anti-MHR serum (Getzoff *et al.*, 1987) was a gift from S. J. Rodda (Chiron Corporation, Melbourne). Anti-HEL immune sera and IgG that had been affinity purified on HEL, as well as purified HEL were provided by I. Kumagai (Tohoku University, Sendai). The pIII-displayed X6 library (Scott & Smith, 1990) and the Cys2-6 pVIII libraries were provided by G.P. Smith (University of Missouri-Columbia); the 15mer pVIII library (Nishi *et al.*, 1993) was provided by H. Saya (University of Kumamoto, School of Medicine). The remaining pVIII libraries, with the exception of AH (see below), are described by Bonnycastle *et al.* (1996). Biotinylation of Abs, isolation of IgG from immune sera, preparation of the anti-phage antisera, and the sources of other reagents are as described (Bonnycastle *et al.*, 1996).

The TPGS peptide (Ac-TPGS-NH<sub>2</sub>) was synthesized by UBC Biotechnology, NAPS Unit (Vancouver). The biotinylated peptides Cys6-7, Cys5-1, X6-1 and X6-14 were synthesized and purified by reversed phase HPLC by The Alberta Peptide Institute (Edmonton). The Cys6-7 and Cys5-1 peptides were subsequently oxidized and re-purified to isolate only disulfide-bridged peptides. The sequences of the peptides are as follows: Cys6-7, NH<sub>2</sub>-VNIACNPTDIQCLIRLPAE-bioK-norL-K-CONH<sub>2</sub>; Cys5-1, NH<sub>2</sub>-NQPSCQDITACLTPQ-bioK-norL-K-CONH<sub>2</sub>; X6-1, NH<sub>2</sub>-SKTPGAAAE-bioK-norL-C-CONH<sub>2</sub>; and X6-14, NH<sub>2</sub>-QEPYVLAEGD-bioK-norL-K-CONH<sub>2</sub>, in which bioK is biotinylated lysine, and norL is norleucine. Cy3-conjugated AffiPure goat-anti-rabbit IgG was purchased from Jackson ImmunoResearch (West Grove, PA). Bio-

tin-labeled bovine serum albumin (BSA) was purchased from Sigma Chemical Co. (St. Louis, MO). Polymethylmethacrylate (PMMA) beads (100 μm) were from Säpdyne Instruments Inc. (Boise, ID).

### The panel of epitope libraries

The peptides in phage-display libraries are fused to the N terminus of either the minor coat protein pIII or the major coat protein pVIII. The panel of peptide libraries comprised two pIII-displayed libraries, X6 and X15, and 13 pVIII-displayed libraries, X5CX2CX5 (Cys2), X5CX3CX4 (Cys3), X4CX4CX4 (Cys4), X4CX5CX4 (Cys5), X4CX6CX4 (Cys6), XCX6CX (LX6), XCX8CX (LX8), X8CX8, XCCX3CX5C (αCT), X6, X15, X30 and AH (see below); X<sub>n</sub> represents *n* randomized residues (X), and C represents a fixed cysteine. The random peptides of the pIII libraries are preceded by the N-terminal sequence ADGA and followed by GGAGAE. The random peptides in all of the pVIII-display libraries begin at the N terminus and are followed by a variety of sequences: PAEGDD (Cys libraries); GGIEGRGP (αCT), GGPAEG (one of the two LX6 libraries), AAEGDD (the remaining libraries). The AH library was designed by T. J. Mason (Melbourne), J. A. Tainer (The Scripps Research Institute, La Jolla) and H. M. Geysen (Glaxo-Wellcome, Research Triangle Park), and constructed by L. L. C. Bonnycastle (Simon Fraser University) as described (Bonnycastle *et al.*, 1996). The peptide insert has the sequence EQLAKSEEQLAKXXEQXXKXXXKXEQLAKSEEQLAX, which is predicted to have α-helical propensity.

### Screening the panel of libraries with anti-HEL and anti-MHR Abs

The screening of pVIII libraries with anti-HEL and anti-MHR Abs (screening 1) has been described (Bonnycastle *et al.*, 1996). Caprylic acid-purified Abs were used throughout four rounds of anti-MHR selection, and for rounds 1, 3 and 4 of the anti-HEL selection, with affinity-purified anti-HEL Abs being used in round 2. The pIII and the AH libraries were screened with anti-MHR in two separate experiments; affinity-purified anti-MHR Abs were used to screen the two pIII libraries (screening 2), and caprylic acid-purified, anti-MHR Abs were used to screen the AH library (screening 3) essentially as described (Bonnycastle *et al.*, 1996). Briefly, ~10<sup>10</sup> to 10<sup>12</sup> virions from each library were affinity-selected over three or four rounds of panning on biotinylated Abs that had been immobilized on avidin or streptavidin-coated microwells or Petri plates; enrichment for Ab-binding phage was assessed by titrating the phage output after each round of panning. Phage were amplified in *Escherichia coli* cells, and amplified phage pools from the third and fourth rounds were tested for Ab binding by ELISA. Individual clones were isolated from pools showing positive enrichment and/or ELISA signals. These clones were sequenced in the region of the peptide insert, and tested for binding by ELISA. Clones were sequenced using the method of either Haas & Smith (1993) or Sanger *et al.* (1977). Pooled phage from the Cys2, Cys3 and AH libraries showed only weak enrichment and ELISA signals after three rounds, and were not studied further.

## ELISAs

Following a previously described protocol (Bonnycastle *et al.*, 1996),  $\sim 2 \times 10^{10}$  virions were immobilized either by direct adsorption to microwells by overnight incubation at 4°C, or by incubation for one hour at room temperature in wells that had been coated overnight with 1 µg of anti-phage Ab. Wells containing bound phage were blocked with BSA or non-fat milk, then 60 nM biotinylated anti-HEL Ab or 100 nM biotinylated anti-MHr Ab was added for 3.5 hours at room temperature. The plates were washed after each step using an automated microplate washer. Bound Ab was detected using horseradish peroxidase-avidin complexes and 2,2'-azino-bis(3-ethylbenzthiazoline-6) sulfonic acid as substrate. Absorbance values were measured at 405 nm minus 490 nm. The ELISA signal for fd-tet (pIII) or f88.4 (pVIII) phage, without a peptide insert was subtracted from those of the isolated phage clones (f88, 0.017 to 0.042 absorbance unit; fd-tet, 0.046 to 0.087 absorbance unit). The results are reported only for clones having an  $A_{405-490} \geq 0.100$  absorbance unit above that of the fd-tet or f88 phage. For competition ELISAs, biotinylated Abs (100 nM) were pre-incubated with 5 µM native HEL or MHr, or with  $\sim 2 \times 10^{10}$  virions, for two hours at room temperature prior to incubation with immobilized phage. Only a few of the clones from the pIII libraries were tested for competition with MHr, because the Abs used to screen the pIII libraries had been affinity-purified on native MHr.

## KinExA immunoassay

PMMA beads were coated with peptide as follows: 200 mg of beads were suspended in Tris-buffered saline (TBS, 50 mM Tris-HCl (pH 7.4), 150 mM NaCl) containing 0.1 mg of biotinylated BSA and gently rotated for one hour at 37°C. The beads were washed five times with TBS, then rotated with 0.1 mg of streptavidin in TBS for one hour at 37°C. The beads were washed ten times with TBS, then rotated with 10 µg of biotinylated peptide in TBS for one hour at 37°C, and stored at 4°C. Prior to use, 10 µg of biotin was added to the coated beads, which were then rotated for one hour at 37°C and suspended in 30 ml of TBS/10 mg/ml BSA.

The KinExA assay for  $K_d$  determination is described by detail by Blake *et al.* (1997), and is described here briefly. For each peptide, the optimal Ab concentration was chosen to give a fluorescent signal of approximately one volt when 100 µl was passed through a KinExA flow cell containing peptide-coated beads. Ab at this concentration was incubated with varying concentrations of biotinylated peptide in TBS/10 mg/ml BSA/0.1% (v/v) Tween and allowed to come to equilibrium at room temperature for a minimum of two hours. The amount of uncomplexed Ab in the equilibrium reaction was quantified by capture on peptide-coated beads, followed by detection using fluorescently labeled anti-rabbit Abs. A fraction of the uncomplexed Ab in each equilibrium reaction was captured by flowing 100 µl of the reaction over peptide-coated beads, followed by a wash with 350 µl of TBS/0.1% Tween. The amount of primary Ab bound to the beads was quantified by flowing 1 ml of Cy3-conjugated goat anti-rabbit Ab (1 µg/ml) over the beads, followed by a wash (1.625 ml). All reagents were passed over the beads in the flow cell at a rate of 0.250 ml/min. Fluorescence of the bead-bound Cy3-conjugate in the flow cell was measured over the course of

the reaction, and the difference in the signal before and after addition of Cy3 was taken as a measure of bead-bound primary Ab. This signal should be directly proportional to the amount of uncomplexed Ab in the equilibrium reaction mixture. Reactions were run in duplicate, and fluorescent signals were plotted as a function of peptide concentration.

Data analysis was performed using the software provided by the manufacturer (Säpidyne Instruments, Inc.). The fluorescent signals were entered along with the corresponding antigen concentrations, and the best fit values for the  $K_d$  and the concentration of antibody active sites were calculated, as well as the 95% confidence interval for these values. The software performed the calculations in the following manner: binding was assumed to follow the equation  $K_d = [Ab][Ag]/[AbAg]$ , in which [Ab] and [Ag] represent the concentrations of free Ab and antigen, respectively. This equation can be rearranged to yield an equation for the free Ab at equilibrium:  $[Ab] = K_d[AbAg]/[Ag]$ . Equations stating that the total Ab concentration ( $Ab_t = [Ab] + [AbAg]$ ), and likewise, for the total antigen ( $Ag_t = [Ag] + [AbAg]$ ), were substituted into the free Ab equation to get free Ab as a function of  $K_d$ ,  $Ab_t$ , and  $Ag_t$ .

The fluorescent signal in the KinExA instrument is linear with [Ab] (Blake *et al.*, 1997), and is defined by the signals in the absence of free Ab (Sig0%), and when 100% of the Ab is free (Sig100%). Thus, the calculation of intermediate signals is a simple linear interpolation:

$$\text{Signal} = (\text{Sig100\%} - \text{Sig0\%})([Ab]/Ab_t) + \text{Sig0\%}$$

Combining this equation with the equation for free antibody at equilibrium resulted in a theory of signal as a function of  $K_d$ ,  $Ab_t$ ,  $Ag_t$ , Sig100% and Sig0%. The software used this theory to find the values of  $K_d$ ,  $Ab_t$ , Sig100%, and Sig0% which minimized the squared error between the theory and the measured signals. The 95% confidence interval was found by varying the  $K_d$  parameter and re-optimizing the other variables at each point. Thus, the analysis finds the range over which the  $K_d$  or  $Ab_t$  maintain fit to the data points.

## Seqery

A set of amino acid sequence patterns was determined for each epitope, based on the critical binding residues identified by alignment of peptides with the HEL and MHr epitopes. The computer program Seqery was developed by L.A.K., M. É. Pique, E. D. Getzoff & J. A. Tainer at The Scripps Research Institute (as a successor to Searchwild, which is described by Collawn *et al.*, 1990). Seqery was used to search the largest non-redundant set of PDB sequences (493 with no more than a 25% sequence identity; Hobohm *et al.*, 1992; PDB\_Select database: <http://swift.embl-heidelberg.de/pdbsel>) for sequences (analogs) matching a given pattern. Each sequence pattern comprised four residues, with one designated as fully variable (indicated by an X). The sequence patterns used in the Seqery searches were XDVQ, DVQX, DXTA, XDIT, DITX, XN[A,S]W, N[A,S]WX, [R,K,H]XPG, XPGS and PGSX for HEL; and XKYS, K[W,Y]SX, XYSE, YSEX, PEXY, EPYX, EXYV, DXSF and WDXF for MHr. Seqery provided a list of all analogs for each sequence pattern, including the names and PDB codes of the proteins in which the analogs

were found, and the sequence and position of the analogs and residues flanking the analogs. The structure of each sequence analog was then assessed using SSA (developed by P.C.S. & L.A.K.).

## SSA

The SSA algorithm was used to (i) assess the degree of structural similarity between epitopes on HEL and MHR and the corresponding sequence analogs found by Sequery, and (ii) assign secondary structures to the sequence analogs. SSA uses a least-squares fitting algorithm to superimpose the backbone atoms of an analog in question onto a template, then calculates the backbone RMSD as a measure of structural similarity of the two. For the analysis of the HEL and MHR epitopes, tetrapeptide templates were derived from the structural coordinates of each epitope identified by peptide mapping. The sequence analogs were then superimposed on the corresponding epitope template; those having a RMSD of  $\leq 0.75$  Å were considered to match the epitope structure.

For the assignment of secondary structures to the sequence analogs, each analog was compared with a set of tetrapeptides comprising ideal structures of an  $\alpha$ -helix, a  $\beta$ -strand, and the most prevalent  $\beta$ -turns (type I, type II, type II', type III, and type VIII). These ideal structures were created using the Biopolymer module in InsightII (Molecular Simulations, Inc., San Diego). InsightII values for the backbone  $\phi$  and  $\psi$  angles were used for the  $\alpha$ -helix and  $\beta$ -strand constructs, and the average  $\phi$  and  $\psi$  values from a statistical analysis of  $\beta$ -turns, as reported by Hutchinson & Thornton (1994), were used to construct each turn segment. Sequence analogs that superimposed onto a model secondary structure with a RMSD of  $\leq 0.75$  Å were assigned to that structure; however, some analogs matched well with both the helical and  $\beta$ -turn models. To distinguish between these, the RMSD between the helical model and the analog was also measured by shifting one residue forward or backward on the protein structure from which the analog was derived; if either of these was also  $\leq 0.75$  Å, the analog was assigned as a helix. Analogous were assigned as "irregular" if none of the model peptides superimposed with an RMSD  $\leq 0.75$  Å.

Four-residue peptides were used in the secondary-structure assignments for several reasons: (i)  $\beta$ -turns are inherently four residues long; (ii) the periodicity of  $\alpha$ -helices is approximately four residues; (iii)  $\beta$ -strands alternate twice within four residues; and (iv) most of the work with Sequery uses tetrapeptides, since exact matches to longer sequences are not likely to be found in the PDB. Prior to this study, SSA was validated by comparing its automated assignments to those made by detailed visual molecular graphics analyses for 158 tetrapeptide segments from a broad range of proteins (P.C.S. & L.A.K., unpublished data). This test resulted in a four-state ( $\alpha$ -helix,  $\beta$ -strand,  $\beta$ -turn and irregular) agreement between SSA and visual assignments of 81.6%. A second comparison was performed on 533 tetrapeptide segments from two proteins representing all  $\beta$ , and  $\alpha$  and  $\beta$  structural folds, and resulted in a four-state agreement of 85.9% over 533 segments. An upper bound of 0.75 Å for backbone atom RMSD was found to be optimal for automated assignment of secondary structure.

The Sequery and SSA software were written in the C programming language, and will be provided to other research groups upon request to L.A.K.

## NMR spectroscopy and the structure calculation of the TPGS peptide

The sample of Ac-TPGS-NH<sub>2</sub> (7 mM in 0.5 ml) was prepared by diluting a stock solution of the peptide with aqueous buffer containing 0.1 M sodium chloride, 0.04 M sodium phosphate, 1% (w/v) sodium azide and 5% <sup>2</sup>H<sub>2</sub>O (pH 5.5). NMR spectra were acquired on a Bruker AMX600 spectrometer (Bruker, Karlsruhe, Germany). Two-dimensional phase-sensitive DQF-COSY (Rance *et al.*, 1983) and ROESY (Bothner-By *et al.*, 1984; Bax & Davis, 1985) spectra were acquired at 283 K using TPPI (Redfield & Kunz, 1975). The sweep widths were 6666 Hz in both dimensions, and 512 *t*<sub>1</sub> increments were collected in 2048 complex data points with 32 transients. The spin-lock pulse for the ROESY spectrum had a strength of 6024 Hz and a duration of 300 ms. Water suppression was achieved using the WATERGATE technique (Piotto *et al.*, 1992; Sklenár *et al.*, 1993) in the ROESY spectrum, and selective irradiation of the water frequency during the recycling delay in the DQF-COSY spectrum. All NMR data were processed with the XWIN-NMR 1.2 program (Bruker). The data were apodized by 90° shifted quadratic sine-bell windows functions, zero-filled in both dimensions and transformed to 2000 × 2000 matrices with a post-acquisition treatment to suppress the residual water peak in the ROESY spectrum. An automatic base-line correction was applied to the ROESY spectrum before integrating the cross-peaks with XWIN-NMR. All chemical shifts were referenced to internal 4,4-dimethyl-4-silapentane-1-sulfonate.

Temperature coefficients for the amide resonances were obtained by recording one-dimensional spectra between 283 and 303 K in steps of 5 K. The amide chemical shift was plotted as a function of temperature and the plot was analyzed by linear regression to extract the slope.

The ROESY cross-peak volumes were converted to distance restraints and calibrated to Pro2 H<sup>β1</sup>-H<sup>β2</sup> (1.8 Å) using the procedure described by Hyberts *et al.* (1992). The upper distance bounds were adjusted for pseudo-atoms by adding 1.5 Å for any methyl group. For restraints involving only one of the non-stereospecifically assigned methylene protons, 1.7 Å were added to the upper distance limit. The structure calculations were performed with the unblocked TPGS peptide (i.e. the N-terminal acetyl group and the C-terminal amide were replaced with amide and carbonyl groups, respectively). The restraints to Thr1 H<sup>N</sup> were imposed on the pseudo-atom of the N-terminal amino group, H<sup>N\*</sup>, with no pseudo-atom correction. The structural ensemble of TPGS was calculated with the DGII program of Insight II 95.0 (Biosym/MSI, San Diego, CA).

## CD spectroscopy of the TPGS peptide

The CD spectrum of Ac-TPGS-NH<sub>2</sub> was obtained at 278 K on a Jasco J710 spectropolarimeter (Japan Spectroscopic Co., Ltd, Tokyo). The peptide was dissolved in water and diluted to a concentration of 1 mM. The measurements were performed in a quartz cell of 0.1 cm pathlength. The CD spectrum is the average of 30 consecutive scans from 180 to 260 nm, recorded with a bandwidth of 0.5 nm, a time constant of 0.25 second and a scan rate of 20 nm/min. Following baseline correction, the observed ellipticities were converted to molar ellipticities. The CD spectrum was deconvol-

luted using the convex constraint algorithm of Perczel *et al.* (1992).

## Acknowledgments

We appreciate generous gifts from H. Alexander, S. J. Rodda, I. Kumagai, H. Saya and G. P. Smith, and the helical library design of J. A. Tainer, T. J. Mason and H. M. Geysen. We thank M. Rashed, J. S. Mehroke and C. Hook for excellent technical assistance, and B. M. Pinto, M. B. Zwick and P. Blancafort for comments on the manuscript. We are grateful to B. M. Pinto and H. J. Dyson for their advice on the structural analysis of the synthetic peptide, and to R. J. Cushley and M. Tracy for their assistance and use of the Simon Fraser University NMR Facility. This work was supported by grants from the US Army Research Office (DAA L03-92-G-0178 to G. P. Smith, University of Missouri-Columbia), the President's Research Fund at Simon Fraser University (J.K.S.), the Natural Sciences and Engineering Research Council (J.K.S.), and the Michigan Research Excellence Fund Centers for Protein Structure, Function, and Design and for Academic Computing (L.A.K.). J.K.S. was supported, in part, by a fellowship from Terrapin Technologies and a scholarship from the B.C. Health Research Foundation. L.C. was supported by a GREAT award from the B.C. Science Council and a Graduate Research Fellowship from the Institute of Molecular Biology and Biochemistry at SFU. The development of Sequery by L.A.K., M. E. Pique, E. D. Getzoff and J. A. Tainer at The Scripps Research Institute was supported by National Science Foundation grant BIR 9631436 (M.E.P., E.D.G. and J.A.T.).

## References

- Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). Protein Data Bank. In *Crystallographic Databases – Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G. & Sievers, R., eds), pp. 107–132. Data Commission of the International Union of Crystallography, Bonn, Cambridge, Chester.
- Arnon, R., Maron, E., Sela, M. & Anfinsen, C. B. (1971). Antibodies reactive with native lysozyme elicited by a completely synthetic antigen. *Proc. Natl Acad. Sci. USA*, **68**, 1450–1455.
- Atassi, M. Z. (1984). Antigenic structures of proteins. *Eur. J. Biochem*, **145**, 1–20.
- Atassi, M. Z. & Lee, C. L. (1978). The precise and entire antigenic structure of native lysozyme. *Biochem. J.* **171**, 429–434.
- Balass, M., Heldman, Y., Cabilly, S., Givol, D., Katchalski-Katzir, E. & Fuchs, S. (1993). Identification of a hexapeptide that mimics a conformation-dependent binding site of acetylcholine receptor by use of a phage-epitope library. *Proc. Natl Acad. Sci. USA*, **90**, 10638–10642.
- Bansal, A. & Gierasch, L. M. (1991). The NPXY internalization signal of the LDL receptor adopts a reverse-turn conformation. *Cell*, **67**, 1195–1201.
- Barlow, D. J., Edwards, M. S. & Thornton, J. M. (1986). Continuous and discontinuous protein antigenic determinants. *Nature*, **322**, 747–748.
- Bax, A. & Davis, D. G. (1985). Practical aspects of two-dimensional transverse NOE-spectroscopy. *J. Magn. Res.* **63**, 207–213.
- Benjamin, D. C., Berzofsky, J. A., East, I. J., Gurd, F. R., Hannum, C., Leach, S. J., Margoliash, E., Michael, J. G., Miller, A., Prager, E. M. *et al.* (1984). The antigenic structure of proteins: a reappraisal. *Annu. Rev. Immunol.* **2**, 67–101.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
- Blake, D. A., Khosraviani, M., Pavlov, Ar. R. & Blake, R. C. (1997). Characterization of a metal-specific monoclonal antibody. *Immunochem. Technol. Environ. Applic.* **657**, 49–60.
- Bonnycastle, L. L. C., Mehroke, J. S., Rashed, M., Gong, X. & Scott, J. K. (1996). Probing the basis of antibody reactivity with a panel of constrained peptide libraries displayed by filamentous phage. *J. Mol. Biol.* **258**, 747–762.
- Bothner-By, A. A., Stephens, R. L., Lee, J.-M., Warren, C. D. & Jeanloz, R. W. (1984). Structure determination of a tetrasaccharide: transient nuclear overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811–813.
- Braden, B. C. & Poljak, R. J. (1995). Structural features of the reactions between antibodies and protein antigens. *FASEB J.* **9**, 9–16.
- Campbell, A. P., Wong, W. Y., Houston, M., Jr, Schweizer, F., Cachia, P. J., Irvin, R. T., Hindsgaul, O., Hodges, R. S. & Sykes, B. D. (1997). Interaction of the receptor binding domains of *Pseudomonas aeruginosa* pili strains PAK, PAO, KB7 and P1 to a cross-reactive antibody and receptor analog: implications for synthetic vaccine design. *J. Mol. Biol.* **267**, 382–402.
- Chang, C.-P., Lazar, C. S., Walsh, B. J., Komuro, M., Collawn, J. F., Kuhn, L. A., Tainer, J. A., Trowbridge, I. S., Farquhar, M. G., Rosenfeld, M. G., Wiley, H. S. & Bill, G. N. (1993). Ligand-induced internalization of the epidermal growth factor receptor is mediated by multiple endocytic codes analogous to the tyrosine motif found in constitutively internalized receptors. *J. Biol. Chem.* **268**, 19312–19320.
- Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.
- Collawn, J. F., Stangel, M., Kuhn, L. A., Esekogwu, V., Jing, S., Trowbridge, I. S. & Tainer, J. A. (1990). Transferrin receptor internalization sequence YXRF implicates a tight turn as the structural recognition motif for endocytosis. *Cell*, **63**, 1061–1072.
- Collawn, J. F., Kuhn, L. A., Sue, Liu L.-F., Tainer, J. A. & Trowbridge, I. S. (1991). Transplanted LDL and mannose-6-phosphate receptor internalization signals promote high-efficiency endocytosis of the transferrin receptor. *EMBO J.* **10**, 3247–3253.
- Cuniasse, P., Thomas, A., Smith, J. C., Thanh, H. L., Leonetti, M. & Menez, A. (1995). Structural basis of antibody cross-reactivity: solution conformation of an immunogenic peptide fragment containing both T and B epitopes. *Biochemistry*, **34**, 12782–12789.
- Davies, D. R., Sheriff, S. & Padlan, E. A. (1988). Antibody-antigen complexes. *J. Biol. Chem.* **263**, 10541–10544.

- Davis, D., Morien, B., Akerblom, L., van Gils, M. E., Bogers, W., Teeuwssen, V. & Heeney, J. L. (1997). A recombinant prime, peptide boost vaccination strategy can focus the immune response on to more than one epitope even though these may not be immunodominant in the complex immunogen. *Vaccine*, **15**, 1661–1669.
- Dyson, H. J. & Wright, P. E. (1995). Antigenic peptides. *FASEB J.* **9**, 37–42.
- Dyson, H. J., Lerner, R. A. & Wright, P. E. (1988a). The physical basis for induction of protein-reactive anti-peptide antibodies. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 305–324.
- Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988b). Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161–200.
- Eberle, W., Sander, C., Klaus, W., Schmidt, B., von Figura, K. & Peters, C. (1991). The essential tyrosine of the internalization signal in lysosomal acid phosphatase is part of a beta turn. *Cell*, **67**, 1203–1209.
- Emini, E. A., Jameson, B. A. & Wimmer, E. (1983). Priming for and induction of anti-polio virus neutralizing antibodies by synthetic peptides. *Science*, **304**, 699.
- Felici, F., Luzzago, A., Folgori, A. & Cortese, R. (1993). Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the *Bordetella pertussis* toxin from phage peptide libraries. *Gene*, **128**, 21–27.
- Fisher, C. L., Greengard, J. S. & Griffin, J. H. (1994). Models of the serine protease domain of the human antithrombotic plasma factor activated protein C and its zymogen. *Protein Sci.* **3**, 588–599.
- Folgori, A., Tafi, R., Meola, A., Felici, F., Galfre, G., Cortese, R., Monaci, P. & Nicosia, A. (1994). A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *EMBO J.* **13**, 2236–2243.
- Francis, M. J., Hastings, G. Z., Sangar, D. V., Clark, R. P., Syred, A., Clarke, B. E., Rowlands, D. J. & Brown, F. (1987). A synthetic peptide which elicits neutralizing antibody against human rhinovirus type 2. *J. Gen. Virol.* **68**, 2687–2691.
- Friedman, A. R., Roberts, V. A. & Tainer, J. A. (1994). Predicting molecular interactions and inducible complementarity: fragment docking of Fab-peptide complexes. *Proteins: Struct. Funct. Genet.* **20**, 15–24.
- Getzoff, E. D., Geysen, H. M., Rodda, S. J., Alexander, H., Tainer, J. A. & Lerner, R. A. (1987). Mechanisms of antibody binding to a protein. *Science*, **235**, 1191–1196.
- Geysen, H. M., Tainer, J. A., Rodda, S. J., Mason, T. J., Alexander, H., Getzoff, E. D. & Lerner, R. A. (1987). Chemistry of antibody binding to a protein. *Science*, **235**, 1184–1190.
- Ghiara, J. B., Stura, E. A., Stanfield, R. L., Profy, A. T. & Wilson, I. A. (1994). Crystal structure of the principal neutralization site of HIV-1. *Science*, **264**, 82–85.
- Girard, M., Meignier, B., Barre-Sinoussi, F., Kieny, M. P., Matthews, T., Muchmore, E., Nara, P. L., Wei, Q., Rimsky, L. & Weinhold, K., *et al.* (1995). Vaccine-induced protection of chimpanzees against infection by a heterologous human immunodeficiency virus type 1. *J. Virol.* **69**, 6239–6248.
- Haas, S. J. & Smith, G. P. (1993). Rapid sequencing of viral DNA from filamentous bacteriophage. *Biotechniques*, **15**, 422–431.
- Hastings, G. Z., Speller, S. A. & Francis, M. J. (1990). Neutralizing antibodies to human rhinovirus produced in laboratory animals and humans that recognize a linear sequence from VP2. *J. Gen. Virol.* **71**, 3055–3059.
- Hobohm, U., Scharf, M., Schneider, R. & Sander, C. (1992). Selection of representative protein data sets. *Protein Sci.* **1**, 409–417.
- Hoess, R. H., Mack, A. J., Walton, H. & Reilly, T. M. (1994). Identification of a structural epitope by using a peptide library displayed on filamentous bacteriophage. *J. Immunol.* **153**, 724–729.
- Hollosi, M., Kover, K. E., Holly, S., Radics, L. & Fasman, G. D. (1987). Beta-turns in bridged proline-containing cyclic peptide models. *Biopolymers*, **26**, 1555–1572.
- Hopp, T. P. & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl Acad. Sci. USA*, **78**, 3824–3828.
- Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3**, 2207–2216.
- Hyberts, S. G., Goldberg, M. S., Havel, T. F. & Wagner, G. (1992). The solution structure of eglin c based on measurements of many NOEs and coupling constants and its comparison with X-ray structures. *Protein Sci.* **1**, 736–751.
- Javaherian, K., Langlois, A. J., LaRosa, G. J., Profy, A. T., Bolognesi, D. P., Herlihy, W. C., Putney, S. D. & Matthews, T. J. (1990). Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. *Science*, **250**, 1590–1593.
- Jemmerson, R. (1987). Antigenicity and native structure of globular proteins: low frequency of peptide reactive antibodies. *Proc. Natl Acad. Sci. USA*, **84**, 9180–9184.
- Jin, L., Fendly, B. M. & Wells, J. A. (1992). High resolution functional analysis of antibody-antigen interactions. *J. Mol. Biol.* **226**, 851–865.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Keller, P. M., Arnold, B. A., Shaw, A. R., Tolman, R. L., Van M., F., Bondy, S., Rusiecki, V. K., Koenig, S., Zolla-Pazner, S. & Conard, P., *et al.* (1993). Identification of HIV vaccine candidate peptides by screening random phage epitope libraries. *Virology*, **193**, 709–716.
- Lara-Ochoa, F., Almagro, J. C., Vargas-Madrado, E. & Conrad, M. (1996). Antibody-antigen recognition: a canonical structure paradigm. *J. Mol. Evol.* **43**, 678–684.
- Laver, W. G., Air, G. M., Webster, R. G. & Smith-Gill, S. J. (1990). Epitopes on protein antigens: misconceptions and realities. *Cell*, **61**, 553–556.
- Lerner, R. A. (1982). Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature*, **299**, 593–596.
- Lundkvist, A., Kallio-Kokko, H., Sjolander, K. B., Lankinen, H., Niklasson, B., Vaheri, A. & Vapalahti, O. (1996). Characterization of Puumala virus nucleocapsid protein: identification of B-cell epitopes and domains involved in protective immunity. *Virology*, **216**, 397–406.

- Luzzago, A., Felici, F., Tramontano, A., Pessi, A. & Cortese, R. (1993). Mimicking of discontinuous epitopes by phage-displayed peptides. I. Epitope mapping of human H ferritin using a phage library of constrained peptides. *Gene*, **128**, 51–57.
- MacCallum, R. M., Martin, A. C. & Thornton, J. M. (1996). Antibody-antigen interactions: contact analysis and binding site topography. *J. Mol. Biol.* **262**, 732–745.
- Meola, A., Delmastro, P., Monaci, P., Luzzago, A., Nicosia, A., Felici, F., Cortese, R. & Galfre, G. (1995). Derivation of vaccines from mimotopes: immunological properties of HBsAg mimotopes displayed on filamentous phage. *J. Immunol.* **146**, 3162–3172.
- Milton, L. R. C. & Van Regenmortel, M. H. V. (1979). Immunochemical studies of tobacco mosaic virus-III. *Mol. Immunol.* **16**, 179–184.
- Motti, C., Nuzzo, M., Meola, A., Galfre, G., Felici, F., Cortese, R., Nicosia, A. & Monaci, P. (1994). Recognition by human sera and immunogenicity of HBsAg mimotopes selected from an M13 phage display library. *Gene*, **146**, 191–198.
- Nishi, T., Tsurii, H. & Saya, H. (1993). Construction of a phage random peptide library and its applications. *Exp. Med.* **11**, 1759–1764.
- Novotny, J. (1991). Protein antigenicity: a thermodynamic approach. *Mol. Immunol.* **28**, 201–207.
- Parge, H. E., Arvai, A. S., Murtari, D. J., Reed, S. I. & Tainer, J. A. (1993). Human CksHs2 atomic structure: a role for its hexameric assembly in cell cycle control. *Science*, **262**, 387–395.
- Pellequer, J. L., Westhof, E. & Van, R. M. (1991). Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol.* **203**, 176–201.
- Perczel, A., Hollosi, M., Tusnady, G. & Fasman, G. D. (1991). Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Eng.* **4**, 669–679.
- Perczel, A., Park, K. & Fasman, G. D. (1992). Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. *Anal. Biochem.* **203**, 83–93.
- Perczel, A., Kollat, E., Hollosi, M. & Fasman, G. D. (1993). Synthesis and conformational analysis of N-glycopeptides. II. CD, molecular dynamics, and NMR spectroscopic studies on linear N-glycopeptides. *Biopolymers*, **33**, 665–685.
- Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
- Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wuthrich, K. (1983). Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485.
- Redfield, A. G. & Kunz, S. D. (1975). Quadrature Fourier NMR detection: simple multiplex for dual detection and discussion. *J. Magn. Reson.* **19**, 250–254.
- Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.
- Richardson, J. S. & Richardson, D. C. (1990). Principles and patterns of protein conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, E. G., ed.), pp. 2–98, Plenum Press, New York.
- Rini, J. M., Stanfield, R. L., Stura, E. A., Salinas, P. A., Profy, A. T. & Wilson, I. A. (1993). Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. *Proc. Natl Acad. Sci. USA*, **90**, 6325–6329.
- Robinson, K., Mostratos, A. & Grecis, R. K. (1995). Generation of rubella virus-neutralising antibodies by vaccination with synthetic peptides. *FEMS Immunol. Med. Microbiol.* **10**, 191–198.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- Satterthwait, A. C., Arrhenius, T., Hagopian, R. A., Zavala, F., Nussenzweig, V. & Lerner, R. A. (1989). The conformational restriction of synthetic peptides, including a malaria peptide, for use as immunogens. *Phil. Trans. Roy. Soc. Lond. Biol.* **323**, 565–572.
- Scott, J. K. & Smith, G. P. (1990). Searching for peptide ligands with an epitope library. *Science*, **249**, 386–390.
- Sheriff, S., Silverton, E. W., Padlan, E. A., Cohen, G. H., Smith-Gill, S. J., Finzel, B. C. & Davies, D. R. (1987). Three-dimensional structure of an antibody-antigen complex. *Proc. Natl Acad. Sci. USA*, **84**, 8075–8079.
- Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. (1993). Gradient-tailored water suppression for <sup>1</sup>H-<sup>15</sup>N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson.* **A102**, 241–245.
- Stephen, C. W., Helminen, P. & Lane, D. P. (1995). Characterisation of epitopes on human p53 using phage-displayed peptide libraries: insights into antibody-peptide interactions. *J. Mol. Biol.* **248**, 58–78.
- Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A. & Hendrickson, W. A. (1984). The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature*, **312**, 127–134.
- Thornton, J. M., Edwards, M. S., Taylor, W. R. & Barlow, D. J. (1986). Location of 'continuous' antigenic determinants in the protruding regions of proteins. *EMBO J.* **5**, 409–413.
- Tormo, J., Blaas, D., Parry, N. R., Rowlands, D., Stuart, D. & Fita, I. (1994). Crystal structure of a human rhinovirus neutralizing antibody complexed with a peptide derived from viral capsid protein VP2. *EMBO J.* **13**, 2247–2256.
- Tsang, P., Rance, M., Fieser, T. M., Ostresh, J. M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1992). Conformation and dynamics of an Fab'-bound peptide by isotope-edited NMR spectroscopy. *Biochemistry*, **31**, 3862–3871.
- van der Werf, S., Briand, J. P., Plaue, S., Burckard, J., Girard, M. & Van Regenmortel, M. H. (1994). Ability of linear and cyclic peptides of neutralization antigenic site 1 of poliovirus type 1 to induce virus cross-reactive and neutralizing antibodies. *Res. Virol.* **145**, 349–359.
- Vargas-Madrado, E., Lara-Ochoa, F. & Almagro, J. C. (1995). Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. *J. Mol. Biol.* **254**, 497–504.
- Venkatachalam, M. (1968). Stereochemical criteria for polypeptides and proteins: conformation of a system of three linked peptide units. *Biopolymers*, **6**, 1425–1436.

- Wagner, G., Neuhaus, D., Worgotter, E., Vasak, M., Kagi, J. H. & Wuthrich, K. (1986). Nuclear magnetic resonance identification of "half-turn" and 3(10)-helix secondary structure in rabbit liver metallothionein-2. *J. Mol. Biol.* **187**, 131–135.
- Westhof, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A. & Van Regenmortel, M. H. (1984). Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature*, **311**, 123–126.
- White-Scharf, M. E., Potts, B. J., Smith, L. M., Sokolowski, K. A., Rusche, J. R. & Silver, S. (1993). Broadly neutralizing monoclonal antibodies to the V3 region of HIV-1 can be elicited by peptide immunization. *Virology*, **192**, 197–206.
- Wien, M. W., Filman, D. J., Stura, E. A., Guillot, S., Delpyroux, F., Crainic, R. & Hogle, J. M. (1995). Structure of the complex between the Fab fragment of a neutralizing antibody for type 1 poliovirus and its viral epitope. *Nature Struct. Biol.* **2**, 232–243.
- Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.* **203**, 221–232.
- Wilson, I. A. & Stanfield, R. L. (1993). Antibody-antigen interactions. *Curr. Opin. Struct. Biol.* **3**, 113–118.
- Wilson, I. A., Ghiara, J. B. & Stanfield, R. L. (1995). Structure of anti-peptide antibody complexes. *Res. Immun.* **145**, 73–78.
- Yao, Z. J., Kao, M. C. & Chung, M. C. (1995). Epitope identification by polyclonal antibody from phage-displayed random peptide library. *J. Protein Chem.* **14**, 161–166.
- Zhong, G., Smith, G. P., Berry, J. & Brunham, R. C. (1994). Conformational mimicry of a chlamydial neutralization epitope on filamentous phage. *J. Biol. Chem.* **269**, 24183–24188.

*Edited by J. A. Wells*

*(Received 29 May 1997; received in revised form 20 April 1998; accepted 28 April 1998)*



<http://www.hbuk.co.uk/jmb>

Supplementary material comprising four Tables and one Figure is available from JMB Online.