



The Human IgE-encoding Transcriptome to Assess Antibody Repertoires and Repertoire Evolution

Ulrika Andréasson¹, Sabine Flicker⁴, Malin Lindstedt¹, Rudolf Valenta⁴
Lennart Greiff², Magnus Korsgren³, Carl A. K. Borrebaeck¹
and Mats Ohlin^{1*}

¹Department of
Immunotechnology, Lund
University, Lund, Sweden

²Department of
Otorhinolaryngology, Lund
University, Lund, Sweden

³Department of Clinical
Pharmacology, Lund University
Lund, Sweden

⁴Department of
Pathophysiology, General
Hospital of Vienna, Vienna
Austria

Upon encounter with antigen, the B lymphocyte population responds by producing a diverse set of antigen-specific antibodies of various isotypes. The vast size of the responding populations makes it very difficult to study clonal evolution and repertoire composition occurring during these processes in humans. Here, we have explored an approach utilizing the H-EPSILON-encoding transcriptome to investigate aspects of repertoire diversity during the season of antigen exposure. We show through sequencing of randomly picked transcripts that the sizes of patients' repertoires are relatively small. This specific aspect of the transcriptome allows us to construct evolutionary trees pinpointing features of somatic hypermutation as it occurs in humans. Despite the small size of the repertoires, they are highly diverse with respect to VDJ gene usage, suggesting that the H-EPSILON-encoding transcriptome is a faithful mimic of other class-switched isotypes. Importantly, it is possible to use antibody library and selection technologies to define the specificity of clonotypes identified by random sequencing. The small size of the H-EPSILON-encoding transcriptome of peripheral blood B cells, the simple identification of clonally related sets of genes in this population, and the power of library and selection technologies ensure that this approach will allow us to investigate antibody evolution in human B lymphocytes of known specificity. As H-EPSILON repertoires show many of the hallmarks of repertoires encoding other isotypes, we suggest that studies of this type will have an impact on our understanding of human antibody evolution even beyond that occurring in the IgE-producing B cell population.

© 2006 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: allergy; antibody; B cell; phage display; repertoire

Introduction

Specific antibodies develop as a consequence of the encounter of B cells with antigens under

conditions providing appropriate signals including those from antigen-activated T cells. In this way, the nature of the antibody response can be regulated to produce, in many cases, a response appropriate for the particular antigen. In all, the antibody response of an individual is very complex involving such a large number of different and unrelated clones that it is impossible to grasp but a small fraction of the entire population. It is thus very difficult to get an overall picture of the composition and evolution of an entire antibody repertoire. Nevertheless, many studies have taken on to identify the overall features of such repertoires, including e.g. gene usage,^{1–4} mutational patterns^{5–10} and in some cases also the composition of antigen specific antibody responses.^{11–21} Such studies help us not only to define basic rules related to antibody development

Abbreviations used: AA, amino acid(s); D, diversity; ELISA, enzyme-linked immunosorbent assay; H, heavy; H-EPSILON, immunoglobulin heavy epsilon chain; H-GAMMA, immunoglobulin heavy gamma chain; J, joining; L, light; Nt, nucleotide(s); PBS, phosphate buffered saline; scFv, single chain antibody fragment; V, variable; VH, V heavy chain domain (V-D-J-region); VL, V light chain domain (V-J-region); V-kappa, V kappa domain; V-lambda, V lambda domain.

E-mail address of the corresponding author:
mats.ohlin@immun.lth.se

but also to understand e.g. why certain populations fail to respond properly to challenge with certain antigens (see e.g. Feeney *et al.*²²). The development of tools in genetic engineering and bioinformatics has greatly aided in these efforts. Nevertheless, the analysis of humoral immune repertoires in humans still suffers from a lack of data. Altogether the characterized repertoires still make up only a minuscule part of the entire immunoglobulin (Ig) response as it occurs *in vivo*. We therefore set out to explore a system based on a defined fraction of the antibody repertoire, more specifically IgE, which we considered to have hallmarks suggesting that it could be mastered experimentally. This isotype is a very minor component of the humoral immune response but one that is of extensive medical importance as it is the major mediator of type I hypersensitivity. We believe in the first place that studies of such repertoires will improve our knowledge of these diseases, as they will tell us how diverse the repertoires are with respect to composition and target recognition. Importantly for this investigation, the low amounts of IgE found in serum suggest that relatively few B lymphocytes populate the pool of cells that actually produce these antibodies. We therefore hypothesised that they represent a set of clones of manageable size for the study of antibody repertoire composition and evolution. Developments in the toolbox that comprises the area of molecular biology now allow us to approach these issues in a rational manner. In particular PCR technology and the complete sequencing of the germline immunoglobulin gene loci have facilitated the analysis of the origin of antibodies in general and as a consequence also of IgE encoded by B lymphocytes. Furthermore, the development of combinatorial library approaches allows us to link given members of a repertoire with antigen specificity. Many studies^{23–35} have addressed aspects of one or the other of these approaches in relation to IgE repertoires but not tried to combine them into a unit that investigates both aspects. We have now validated an approach comprised of methods to carry out random sequencing of the immunoglobulin heavy epsilon chain (H-EPSILON)-encoding transcriptome and isolation of allergen-specific sequences by combinatorial antibody library technology. We demonstrate the appropriateness of using the IgE repertoire to study antibody repertoire composition. We also confirm our hypothesis that the IgE repertoire at a given point in time is populated by a small set of clonotypes, which nevertheless has many of the characteristics of other, complex immunoglobulin repertoires. Furthermore, we demonstrate how the selection of antigen-specific antibody fragments from a library made from genes encoding the H-EPSILON can be used not only to identify allergen-specific antibody fragments, but importantly also that selected clonotypes can be placed into the context of the overall H-EPSILON-encoding repertoire. We thus conclude that the human IgE repertoire represents a population of molecules

that can be explored to investigate antibody repertoires and repertoire development as it occurs *in vivo* and we suggest that this combination of tools will help us to understand processes involving immune responses in the future.

Results

Basis and validity of the sequencing process

To assess the composition of the H-EPSILON-encoding transcriptome in allergic patients, we obtained mRNA from approximately 2×10^8 mononuclear cells from a number of allergic donors. To ensure specific amplification of variable heavy chain domain (V-D-J-REGION) (VH) genes encoding the H-EPSILON repertoire, cDNA was obtained by reversed transcription using primers specific for the IGHE constant gene.³⁶ Specificity was further enhanced in the subsequent PCR process as we used primers specific for the IGHE CH1 sequence in addition to primers specific for all IGHV gene subgroups. Several such cDNAs repeatedly produced PCR products of the estimated size, and samples obtained from two donors, A3 and A4, that were a part of a separate study of the immunology of allergy,³⁷ were used for further sequence analysis. An assessment of sequences at the 3'-end of the amplification products confirmed that these sequences indeed originated from an H-EPSILON-encoding transcript, demonstrating the validity of the procedure and ruled out the possibility that the products were encoding isotypes other than IgE. Only 3/114 sequences (A4-30, A4-63 and A4-154) carried stop codons, that may have originated not only from an aberrant gene rearrangement or mutational process occurring *in vivo*, but also as an *in vitro*-artefact by the introduction of rare mutations during the PCR process. Nevertheless, the rarity of sequences not carrying an open reading frame suggested that the amplified sequences largely encoded functional IgE, i.e. they had the required origin, as expected from the experimental set-up.

A critical feature of our approach is whether our system behaves predictably in comparison to other previously described methodologies, or not. To assess this issue, we sequenced a set of gene products amplified from the immunoglobulin heavy gamma chain (H-GAMMA)-encoding transcriptome as well. Since much information is available in the literature^{2,3,11} with regards to its composition we could assess to what extent our procedure produced similar products. Among the sequenced genes ($n=21$), we got an appropriate set originating from 17 different germline genes representing diverse IGHV subgroups (Table 1). Even in this small set of products, all IGHJ subgroups and all IGHD subgroups apart from the rarely used IGHD7 subgroup^{38,39} were identified. Thus, our amplification and cloning approach allowed us to identify sets of genes that compared well with what is

Table 1. Frequency of IGHV subgroups and genes used in H-EPSILON repertoires as calculated on data of clonally unrelated sequences in comparison to other immunoglobulin repertoires

IGHV subgroup	Donor A3 (n=33) (%)	Donor A4 (n=39) (%)	Overall frequency (n=72) (%)	Germline genes used in these IgE repertoires (IGHV)	Frequency in IgG repertoire (n=21) (%)	Previously reported repertoire frequencies (%)			
						Mononuclear cells ¹	Antibodies specific for proteins ¹¹	IgG-producing B cells ²	IgG-producing B cells ³
1	24	41	33	1-02, 1-18, 1-46, 1-58, 1-69	33	11	26	14.9	12
2	0	0	0	None	0	3	4	1.3	1
3	52	49	50	3-07, 3-11, 3-15, 3-21, 3-23, 3-30, 3-33, 3-43, 3-48, 3-53	48	52	40	54.5	62
4	9	3	6	4-31, 4-39	5	19	22	21.1	18
5	9	8	8	5-51, 5-a	14	10	8	7.7	6
6	0	0	0	None	0	5	0	0.3	2
7	6	0	3	7-4.1	0	Not reported	Not reported	0.3	Not reported

known regarding the composition of antibody repertoires. Altogether, we demonstrated that the amplification and cloning methodology yielded products relevant for the study of antibody repertoire composition.

Composition of the H-EPSILON-encoding repertoire

A total of 114 sequences (representing 72 independent clonotypes, see below) originating from the H-EPSILON-encoding transcriptome were analysed by sequencing (GenBank accession nos. DQ065615–DQ065728). Sequences belonging to the major IGHV subgroups 1, 3, 4 and 5 were found in both donors and subgroup 7 was also found in donor A3. Thus, of the seven IGHV subgroups,³⁶ all subgroups but IGHV2 and IGHV6, which are used only rarely in immunoglobulin repertoires,^{1,2} could be identified among the sequenced genes. No striking overrepresentation of the IGHV5 subgroup, that has been reported in some H-EPSILON repertoires,^{25,26} was seen among peripheral blood H-EPSILON-producing B cells. A slight apparent over-representation of subgroup 1 and an under-representation of subgroup 4 were, however, noted (Table 1). The latter finding correlates with the fact that IGHV4 subgroup although frequently used in immunoglobulin heavy mu chain (H-MU) repertoires, is more rarely recruited into class-switched repertoires, like H-GAMMA.³ The fact that the small population of H-GAMMA-encoding repertoire showed similar frequencies of IGHV1 and IGHV4 as the H-EPSILON-encoding repertoire, however, suggests that there is a slight bias in the PCR amplification process affecting the observed frequency of these transcripts. In general though, several IGHV genes often found in other immunoglobulin repertoires were also identified as being frequently used in these H-EPSILON-encoding repertoires. For instance, among the ten most commonly used IGHV genes in H-GAMMA repertoires,² eight were found in the herein studied H-EPSILON-repertoire accounting for 57% of all independent clones. The same ten genes accounted

for 58% of gene usage in H-GAMMA repertoires. In contrast, among the 30 least commonly used genes in the H-GAMMA repertoire, only eight were used in the assessed H-EPSILON repertoire accounting for 31% of independent clones. These very same 30 genes encoded 30% of the H-GAMMA repertoire.² Thus, H-EPSILON-encoding gene usage had extensive similarities with H-GAMMA-encoding gene usage. Similarly, the length of the H chain third complementarity determining region CDR3-IMGT (14.8 residues (range 7–26) as assessed on the 72 independent clonotypes) was close to that reported for other human immunoglobulin repertoires (15.1–15.3 residues^{4,40}). J gene usage (Table 2) also showed distribution mostly resembling the distribution of gene usage in other immunoglobulin populations, with IGHJ4 being the most frequently used gene. Only IGHJ5 was less frequently used in the herein investigated H-EPSILON repertoire. Similarly, IGHD genes used in H-EPSILON-encoding repertoires distributed essentially as expected from studies of other immunoglobulin repertoires^{38,39} between different subgroups (Table 2) and included at a high frequency e.g. IGHD2-2, IGHD3-3, IGHD3-22 and IGHD6-19. In fact, IGHD-gene usage was indeed very diverse as not less than 18/27 IGHD genes could be identified among the 72 clonotypes. Furthermore, preferences for IGHD gene reading frames for the most frequently used genes (Table 3) was similar to that reported for other antibody repertoires.³⁸ We thus conclude that peripheral H-EPSILON repertoires are diverse with respect to gene usage and represent a set of sequences with features of IGHV, IGHD and IGHJ gene usage similar to other immunoglobulin populations. Studies of H-EPSILON-encoding transcriptomes are thus likely to provide information on antibodies representative not only of itself but also of other isotypes participating in adaptive immune responses, as well.

Mutations in the H-EPSILON repertoire

An analysis of the H-EPSILON-encoding sequences demonstrated that they were largely somatically hypermutated as expected from genes

Table 2. Frequency of usage of IGHD and IGHJ genes in H-EPSILON transcripts in comparison to previously published studies

Locus	Subgroup	Frequency in donor A3 (<i>n</i> =33) (%)	Frequency in donor A4 (<i>n</i> =39) (%)	Compiled data (%)	Other antibody repertoires (%)		
					Corbett <i>et al.</i> ³⁸	Souto-Carneiro <i>et al.</i> ³⁹	Yamada <i>et al.</i> ⁴
D	IGHD1	9	15	13	3.4	13	
	IGHD2	15	31	24	8.1	12	
	IGHD3	52	31	40	19.8	36	
	IGHD4	3	5	4	4.1	8	
	IGHD5	0	3	1	4.7	7	
	IGHD6	15	10	13	9.5	17	
	IGHD7	0	0	0	0.9	1	
	Unidentified/others	6	5	6	49.5	5	
J	IGHJ1	3	3	3			1
	IGHJ2	3	0	1			0
	IGHJ3	6	11	8			9
	IGHJ4	67	45	55			53
	IGHJ5	3	3	3			15
	IGHJ6	18	39	30			22

that have undergone the evolutionary processes occurring in germinal centers. Indeed, independent clonotypes had mutated more than 2% of their bases. Individual clones carried 3–52 (median value: 14; *n*=114) mutations in the part of the V-REGION between the 5'-PCR primer and the start of CDR3-IMGT (codons 8–104, nucleotides 24–312, according to the IMGT unique numbering for V-DOMAIN³⁶). An analysis of the composition of such mutations (Table 4) indicated a preference for C mutating to T, A mutating to G, G mutating to A and C, while T mutated mostly to C or G. The most frequently mutated residues were C and G, probably reflecting the targeting of these bases in the major hot spot motif⁸ by activation-induced cytidine deaminase.⁴¹ The impact of mutations originating from Taq polymerase-induced sequence errors was thus not strong as there was no extreme overrepresentation of T→C and A→G mutations, that are typical hallmarks of such errors,⁴² in comparison to mutations found in other immunoglobulin repertoires.⁴³ Importantly, G bases were much more readily mutated if they were present in typical hot-spots of mutation whether represented by the classical RGYW⁹ or alternative DGYW⁸ sequence motives (Table 5). We thus conclude that the pattern of mutations in these H-EPSILON repertoires was not an artefact of the PCR amplification procedure but rather that it is a true reflection of the situation in the H-EPSILON-encoding gene population. Clonal evo-

lution patterns and mutations in members of the identified repertoires, be them randomly picked or identified after selection on allergen, are thus likely to represent sequences actually found *in vivo*.

A very limited size of the H-EPSILON-encoding repertoire

With access to an immunoglobulin-encoding transcriptome faithfully copied into DNA, we were able to address matters related to its composition. To validate or abandon our initial hypothesis, i.e. the existence of a very limited size of the H-EPSILON-encoding transcriptome in a given individual, we looked for evidence in favour or against this notion. Although the H-EPSILON-encoding repertoires of the two individuals were, as we have seen above, quite diverse with respect to gene usage, several observations suggested that they were quite limited in size. In fact, among the 63 and 51 sequences derived from donors A3 and A4, we repeatedly identified clonally related sequences. This set of 114 sequences contained only 72 unique clonotypes as 23 sets of clonally related sequences, as evidenced by the occurrence of sequences sharing an origin in the same V gene and harbouring highly homologous CDR3 sequences, were found. These sets were populated by two to six members each. Thus, a single clonotype could represent up to approximately one tenth of all

Table 3. IGHD reading frame in H-EPSILON-transcripts and in a large set of other, rearranged human H chain genes

IGHD gene	H-EPSILON repertoire				Corbett <i>et al.</i> ³⁸			
	Number of sequences	Frequency of reading frames ^a (%)			Number of sequences	Frequency of reading frames (%)		
		Stop	Hydrophilic	Hydrophobic		Stop	Hydrophilic	Hydrophobic
2-2	7	14	71	14	33	6	67	27
3-3	12	25	58	17	43	19	58	23
3-22	7	0	100	0	34	3	88	9
6-19	7	14	43	43	42	12	43	45

^a As defined by Corbett *et al.*³⁸

Table 4. Frequency of mutations in clonally unrelated genes that originated from IGHV3-23 ($n=8$)

From	To				Number of mutations	Frequency of mutated bases (%)
	T	C	A	G		
T	–	0.37	0.21	0.42	19	4.2
C	0.60	–	0.12	0.28	38	7.0
A	0.28	0.08	–	0.65	28	6.0
G	0.10	0.44	0.46	–	57	8.7

sequences derived from the transcriptome of a single donor. Importantly, members of each clonotype usually differed in sequence. Thus, they did not represent repeated cloning of cDNA derived from a few cells with a high mRNA content, but rather represented the transcripts found in multiple cells with a common origin that to different degrees had undergone somatic hypermutation. The sets of sequences of H-EPSILON-encoding genes found in this study that were represented several times had their origin in all five IGHV subgroups (1, 3, 4, 5 and 7). Together, these results demonstrate that the pool of lymphocytes that populates the peripheral H-EPSILON-producing population is very restricted in its size but that diversification of these clonotypes as expected enhances the diversity of the produced antibodies. As the H-EPSILON-encoding repertoire is small it can be handled experimentally. This transcriptome thus has features appropriate for the analysis of antibody evolution as it occurs in humans.

Occurrence of similar but independent clonotypes in H-EPSILON repertoires

Further analysis of the amplified sequences indicated that the rearrangements were largely different in-between individuals. Similarly, these rearrangements were largely different from those found in the H-EPSILON-encoding repertoire of three individuals investigated in another study.⁷ Thus, chance events and/or differences in antigens and epitopes recognised by the H-EPSILON repertoire of different individuals results in differences in repertoire makeup in-between individuals. This is to be expected from polyclonal responses targeting a multitude of antigens (in this case: allergens). Nevertheless, the small sizes of the H-EPSILON transcriptomes aided in the identifications of important exceptions to the general rule. This is best exemplified by the occurrence of a similar rearrangement based on the IGHV4-39 germline gene in donors A3 (clones A3-EQCA17 and A3-EE24) and A4 (clone A4-48) (Figure 1). Thus, even when sequencing a relatively small number of sequences (63 and 51, in donors A3 and A4, respectively), it was possible to identify a definitive overlap between repertoires recruited into the IgE-producing B cell population of different individuals. BLAST analysis against GenBank did not identify this type of rearrangement in

the database, suggesting that it is not a common rearrangement found in immunoglobulin populations in general. Rather these sequences very likely target an, to us unknown, immunodominant allergen, that is preferentially recognised by a very specific rearrangement. Similarly, within a given individual, sequences originating from different V genes (belonging to the same IGHV subgroup) but having very similar CDR3-IMGT sequences occurred (Figure 1(c)). We believe that the small size of the H-EPSILON repertoire not only facilitates the identification of these sets of similar but independent sequences, but will also allow us to approach matters of antigen-specific repertoire diversity and diversification in humans in a way not easily achievable with very large H-GAMMA repertoires. As a consequence, we can sequence a sizable fraction not only of an antigen-specifically selected repertoire, as further outlined below, but also of the overall population, and potentially connect the two populations on a sequence level.

Selection of *Phl p 5*-specific antibody fragments

Having established the nature of the H-EPSILON repertoire, we envisaged that the relatively small size of the repertoire would facilitate library construction and use. As relatively few members populated the IGHV repertoire, even a small library would cover many permutations of these members with respect to light (L) chain pairing and allow for selection of allergen-specific antibody specificities. We constructed a single chain antibody fragment (scFv) library using the H-EPSILON VH repertoire. These genes were combined with V light chain domain (V-J-REGION) (VL) genes (with an origin in B cells producing antibodies of a variety of isotypes). As the lymphocyte donors were allergic to grass pollen including that from timothy, and as the lymphocytes from which the library was established were obtained during the grass pollen season, we believed that timothy-specific antibodies were among the specificities encoded for by this repertoire. Selection on *Phleum pratense* extracts and on the major timothy grass pollen allergen *Phl p 5A*, an allergen recognised by serum IgE of the two donors, enriched for clones of which many

Table 5. Mutational frequencies of G nucleotides in intrinsic hot spots (RGYW or DGYW or their complements) in comparison to G found elsewhere in the gene sequence

	In RGYW/ WRCY	In rest of the sequence	In DGYW/ WRCH	In rest of the sequence
Number of G/C	176	1024	264	936
Mutations of G/C	43	52	55	40
Frequency	24%	5%	21%	4%

Results were calculated on eight independent sequences encoding H-EPSILON derived from donors A3 and A4 and originating from germline gene IGHV3-23. The hot spot residue as found on the sense strand that was analysed is indicated in bold type.

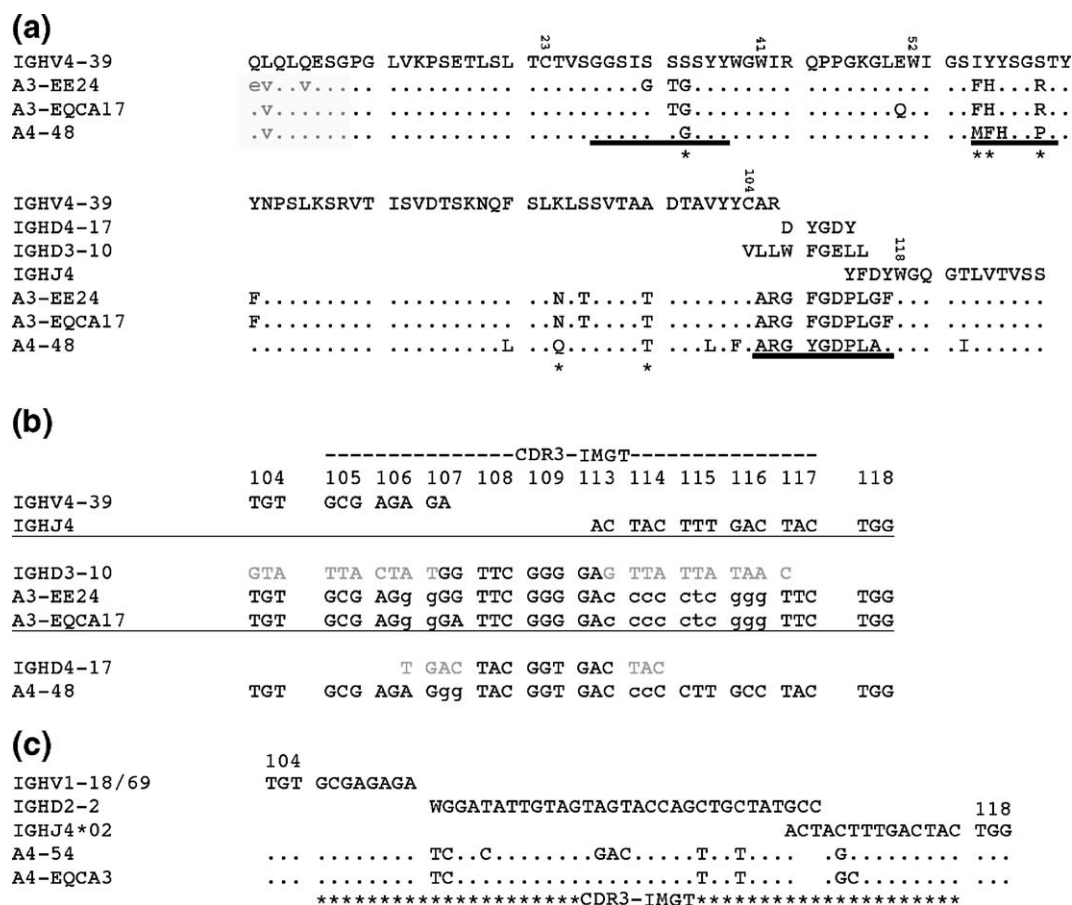


Figure 1. Analysis of randomly picked H-EPSILON VH derived from two individuals identified clones that display a high degree of sequence similarities. (a) Sequences of H-EPSILON VH transcript derived from donor A3 and A4. CDR-IMGT is underlined and five highly conserved residues in VH domains are indicated by their unique sequence number. Residues encoded by primers used for amplification are shown in lower case and grey lettering, as they do not necessarily represent that sequence occurring *in vivo*. The H-EPSILON encoding genes all have an identical genetic origin based on the IGHV and IGHJ gene usage and they use IGHG genes that encode highly similar residues in part of their sequences. The sequences even display a high frequency of identities or similarities including an identical length of the third hypervariable loop. In addition, several substitutions have occurred and been tolerated or even selected for (*) in the affinity maturation process in antibodies derived from the two different donors. The nucleotide sequences of CDR3-IMGT of the three sequences and their germline origins (b) show that sequence similarity does not only have its origin in the germline gene usage but also that N/P nucleotide insertions (lower case letters), as proposed by an IMGT/V-QUEST⁴⁴ and IMGT/Junction Analysis,⁴⁵ or extensive hypermutation in CDR3-IMGT have contributed to the similarity in V domain sequence. The parts of the D-REGION that apparently were not included in the rearrangements are shown in grey. (c) Sequences and proposed genetic origin of two clonally unrelated but similar sequences found in one individual. These sequences originate from IGHV1-18 (A4-54) and IGHV1-69 (A4-EQCA3) but have very similar CDR3 sequences. Note that the CDR3-IMGT is encoded almost entirely by the IGHV, IGHG and IGHJ genes and carries few, if any, N/P nucleotides.

showed specificity for the selecting allergens. Sequencing of these clones identified seven unique sequences while the remaining represented multiple examples of two, dominant, clones (Table 6). Five out of the seven selected clones bound both to *P. pratense* allergen extracts and to recombinant *Phl p 5A*, while the remaining scFv (clones 4.1 and 4.2) bound the pollen extract poorly while they bound strongly to recombinant *Phl p 5A*. The failure of scFv 4.1 and 4.2 to bind pollen extracts seemed to be related to the ELISA blocking agent used, as they bound to pollen extracts (of which only a fraction is the various isoforms of *Phl p 5*) when using BSA as blocking agent in the assay (data not shown). In conclusion, the specificity of the seven

selected scFv were in all cases *Phl p 5*, a major allergen of timothy grass pollen. Three of the seven clones (4.1, 4.3 and 4.10) were clonally related in that they were encoded by identical VH genes in combination with different light chains (Table 6) while other clones were unrelated in sequence. We could show that the scFv specific for *Phl p 5* originated from several different germline IGHV, IGHG and IGHJ genes and subgroups, demonstrating that the IgE response against epitopes on this important allergen involves a diverse set of genes that is not clonally restricted. Thus, the approach allowed us to identify a set of allergen-specific human members that are representative of the IgE population *in vivo*.

Table 6. Genetic origin of genes encoding scFv (GenBank accession numbers DQ065729-DQ056735) specific for *Phl p 5*

ScFv clone	Heavy chain germline gene origin			CDRH3 length	Light chain	Light chain germline gene origin		CDRL3 length	Frequency of clone after selection (n = 16)
	IGHV	IGHD	IGHJ			IGLV	IGLJ		
4.1	IGHV3-11	IGHD1-1	IGHJ3	15	κ	IGKV1D-39 or IGKV1-39	IGKJ3	9	1
4.3	IGHV3-11	IGHD1-1	IGHJ3	15	λ	IGLV2-14	IGLJ3	10	7
4.10	IGHV3-11	IGHD1-1	IGHJ3	15	λ	IGLV2-14	IGLJ3 or IGLJ2	11	1
4.4	IGHV3-23	IGHD3-3	IGHJ6	21	λ	IGLV1-44	IGLJ1	11	4
4.2	IGHV5-a	IGHD2-15	IGHJ4	20	λ	IGLV1-44	IGLJ3	11	1
4.12	IGHV1-69	IGHD6-19	IGHJ4	20	κ	IGKV4-1	IGKJ4	9	1
4.13	IGHV1-18	IGHD3-22	IGHJ6	23	λ	IGLV1-44	IGLJ1	11	1
Clone 5 ^a	IGHV3-30 or IGHV3-30-3	IGHD1-1	IGHJ3	17	κ	IGKV1-12	IGKJ2	9	

^a Data are also given for a previously described Fab (clone 5) specific for *Phl p 5*³⁴ (GenBank accession number X95750).

Epitope specificity and genetic origin of *Phl p 5*-specific antibody fragments

As outlined above, five independent clonotypes (with respect to H-EPSILON sequence) specific for *Phl p 5* were identified among the sequences selected from the herein described scFv library (Table 6). Having defined the extensive sequence diversity of the *Phl p 5*-binding population it became important to establish the diversity of epitope specificity within *Phl p 5*, if any. A combination of various techniques defined, as summarized in Table 7 (experimental evidence is provided as Supplementary Data), that these scFv and the previously defined clone 5 Fab³⁴ recognise at least five different epitopes on this allergen. The first epitope (#1) is recognised by scFv 4.3. Although not strictly confirmed, it is very likely that its close sequence relatives, scFv 4.1 and 4.10 (Table 6), specifically recognise this epitope, too. The previously described clone 5 Fab recognises a different epitope (#2) that, just as epitope #1, is found on the residue 56–165 fragment of *Phl p 5A*.⁴⁶ Both of these epitopes are found on other group 5 antigens like that found in *Lolium perenne*. ScFv 4.2 binds yet another epitope (#3) on the same antigenic fragment of *Phl p 5*, but one that, in contrast to

epitopes #1 and #2, is not found on *Lol p 5*. Epitope #3, however, overlaps with epitope #1 detected by scFv 4.3 and epitope #2 detected by clone 5 Fab. Other scFv bind epitopes not found on the *Phl p 5A* 56–165 fragment. Among these, scFv 4.4 binds an epitope that cannot be detected on *Lol p 5* while scFv 4.12 and 4.13 detect one or two epitopes also found on *Lol p 5*. Thus, we have demonstrated that, despite the fact that overall IgE repertoires are small, they are not confined in their recognition of a single or a few immunodominant epitopes. Rather, the diversity in IgE repertoires is large enough to ensure that antibodies against a given allergen are diverse with respect to epitope specificity.

Clonal relationship of members of the H-EPSILON-encoding variable domains in peripheral blood lymphocytes

Having access to sets of H-EPSILON sequences, we had, as outlined above, been able to identify clonally related sets of sequences in the populations, suggesting that the repertoire is populated by a relatively small set of clones. Indeed, 23 clonally related sets of sequences were identified, 15 of them in the transcripts derived from donor A3. In several cases, these sets were represented by

Table 7. Specificity behaviour of *Phl p 5*-specific antibodies/antibody fragments as determined by the recognition of class 5 antigens

Specificity	Binding specificity						Inhibited in Biacore analysis after preincubation with	
	<i>rPhl p 5A</i>	<i>P. pratense</i> pollen extracts ^a	<i>L. perenne</i> pollen extracts	λ gt11-encoded <i>Phl p 5</i>	λ gt11-encoded <i>Lol p 5</i> fragment 81	<i>rPhl p 5</i> fragment	4.3 IgG	Clone 5 Fab
4.1	+	–	+/-	+	–	+/-	ND ^b	ND
4.2	+	–	–	+	–	+	Yes	Yes
4.3	+	+	+	+	+	+	Yes	No
4.4	+	+	–	+	–	–	ND	ND
4.12	+	+	+	+	–	–	ND	ND
4.13	+	+	+	+	–	–	ND	ND
Clone 5	+ ^{34,46}	+ ³⁴	+ ⁴⁶	ND	+ ⁴⁶	+ ⁴⁶	No	Yes

^a All scFv bound pollen extracts specifically when the assay was blocked by 1% BSA instead of 1% skim milk powder.

^b ND, not determined.

several products allowing us to construct evolutionary trees (Figure 2). Mutational paths involved either those sharing several mutations prior to clonal divergence or those that apparently diverged early during molecular evolution. Importantly, one such set of sequences was also related to the *Phl p 5*-specific scFv 4.2 selected by phage display, identifying a larger set of human VH domain (V-D-J-REGION) sequences with a likely identical and known specificity (Figure 2(c)). We were thus able to show the evolutionary paths taken by a clonotype of known specificity. In order

to show the validity of the relationship between these sequences we utilized several approaches. Firstly, we identified that the sequences shared several mutations (Figure 3(a)), suggesting that they originated from common evolution and selection events. Furthermore, an attempt to confirm the specificity of these randomly picked VH sequences for *Phl p 5* was made. For this purpose two of these sequences were produced as complete H-GAMMA1 chains together with a lambda light chain comprising the VL of scFv 4.2, to obtain intact IgG1. This approach created antibodies that in both cases retained specificity for *Phl p 5A* (Figure 3(b)), supporting the notion that all these VH-genes had evolved through a common affinity maturation process aimed at developing high affinity antibodies against *Phl p 5* or a similar group 5 allergen. Thus, the specific features of H-EPSILON-encoding transcriptomes and our ability to identify allergen specificities of clones from them by library technology, as demonstrated here, strongly suggest that they are highly appropriate for the further analysis of the evolution of antigen-specific responses in humans.

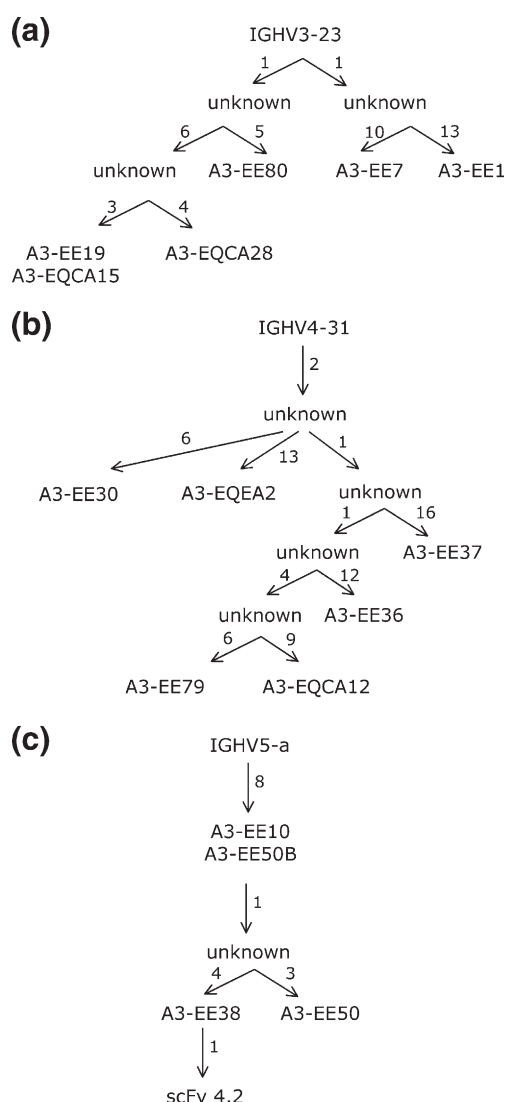


Figure 2. Proposed clonal relationship and mutations of randomly picked sequences encoding H-EPSILON that originated from IGHV3-23 (a) and IGHV4-31 (b) and that harbour related CDRH3 sequences suggesting an origin in a common precursor. (c) Clonal relationship of a set of sequences that in addition to randomly picked H-EPSILON-encoding sequences also include a *Phl p 5A*-specific scFv (4.2) (GenBank accession no. DQ065730), all of which originate from germline gene IGHV5-a. The numbers of base mutations found in-between the 5'-primer used for amplification and the first codon in CDRH3 are shown next to each arrow.

Discussion

The extensive diversity of the receptors and effector molecules produced by the adaptive humoral immune system is supposed to be of paramount importance for its ability to respond to the wide array of potential threats that the organism faces. Consequently, a complex system to establish such diversity has developed. Different species approach this matter in different ways. Some, like man and rodents, harbour in their genome several different germline genes that are recombined in different ways to produce a multitude of different proteins. Others, such as sheep, have fewer functional genes but diversify the receptors by hypermutation early in B cell development (prior to antigen exposure). Yet another approach is to use gene conversion to expand on the initially created diversity established from the few available germline genes. All these different systems and the potential for conformational isomerism of paratopes⁴⁸ create a vast diversity of antigen binding sites that are able to recognise virtually any substance. Further molecular evolution through somatic hypermutation diversifies and perfects these binding surfaces with respect to antigen recognition. Developments in genetic engineering and the determination of protein structure has over the past decades allowed us to begin to understand how these processes occur and which underlying driving forces directs these processes. Much of importance is, however, still to be learned, in particular as it may aid our understanding of the development of humoral immunity, either when it develops to combat disease⁴⁹ or when it develops in a way that by itself causes disease. Importantly, differences in-between species with respect to the

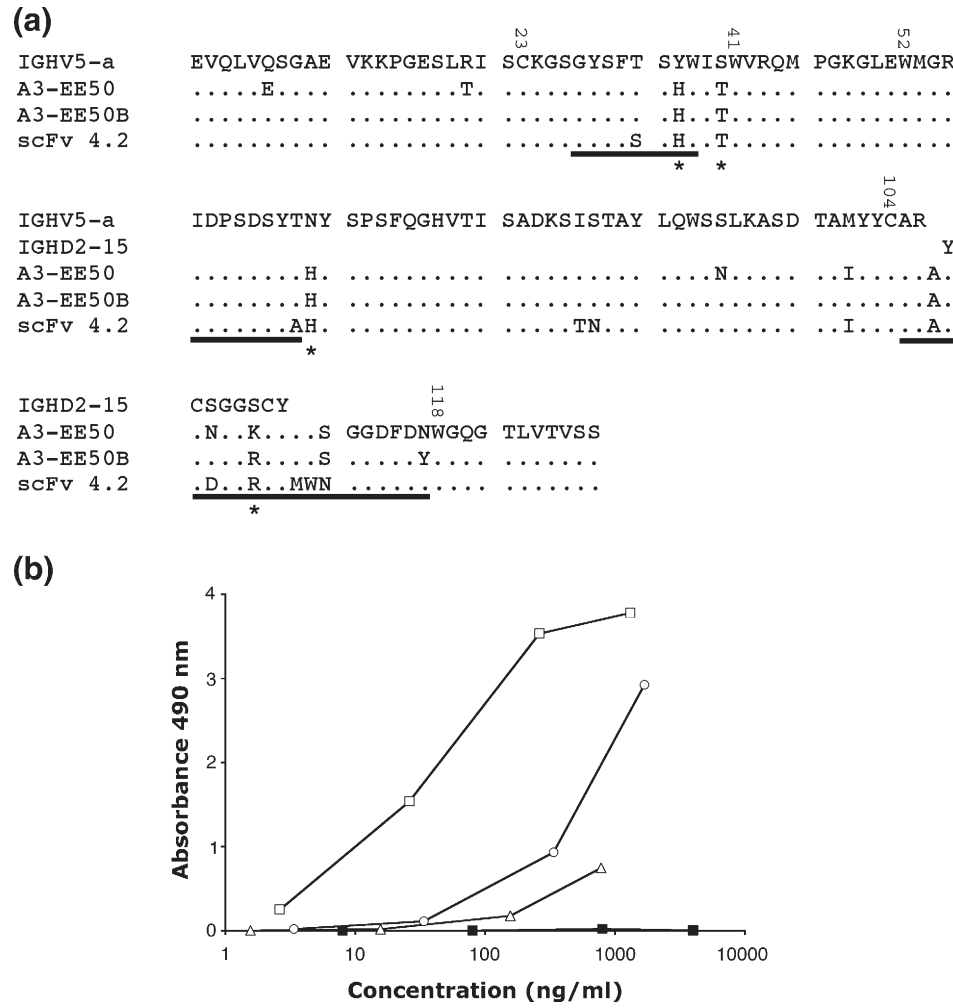


Figure 3. Clonally related VHs and their allergen specificity. (a) Sequences of VH in scFv 4.2 as well as in A3-EE50 and A3-EE50B are shown in comparison to relevant germline IGHV and IGHD genes. The clonal relatedness of these VH sequences (Figure 2), as evidenced by a high level of identical residues in their sequences including many identities in the otherwise very variable CDR3, and the existence of common or very closely related substitutions (stars) is evident. (b) Reactivity with *Phl p 5* of 4.2 in the IgG1 format (open squares), and variants of this antibody carrying the A3-EE50 (triangles) and A3-EE50B (circles) VH together with 4.2 VL. An irrelevant antibody (SG20) specific for the SKBR3 cell line⁴⁷ (filled squares) was included as negative control. The specificity of the reaction was confirmed by the inability of these antibodies to bind to immobilised BSA under the same conditions (data not shown).

germline repertoire and the establishment and evolution of diversity suggest that animal systems are not always appropriate models of immune responses as they occur in humans.

The analysis of immune repertoires and recruitment of specificities has been a long-standing issue in immunology. Many reports have outlined the composition of naïve and antigen-selected repertoires in general but also the composition of antigen-specific repertoires. The latter is in particularly the case with murine antibody responses where repertoires and their development have been elucidated. In the case of human antibody repertoires less is known with respect to the composition of repertoires with known specificities. Developments in the fields of germline immunoglobulin loci and peripheral immunoglobulin gene sequencing, human monoclonal antibody technology and combinatorial library technologies, have, however, recently pro-

vided new insights into the makeup of specific human antibody repertoires. Much of these efforts have gone into attempts to isolate antibodies, mostly from non-immune antibody repertoires, for the treatment of human diseases. Other studies have, however, also begun to unravel the nature of specific repertoires in general^{2,5,11} or demonstrated features associated with given specificities including polysaccharide and complex proteinaceous targets¹²⁻²¹ as they develop *in vivo*. Most of these investigations are made on large peripheral lymphocyte repertoires and thus evaluate only a small fraction of the entire specific repertoire. We have now used existing technologies in an approach to investigate overall repertoires of a single isotype that, we believed, had features facilitating human antibody repertoire analysis.

Firstly, we observed that the peripheral H-EPSILON repertoire, as defined by the transcriptome in

two individuals during the grass pollen season, was complex as it originated from a wide variety of IGHV, IGHD and IGHJ genes. In fact, the composition was in many respects very similar to that found in H chain repertoires of other isotypes. The members of the H-EPSILON repertoires also showed evidence of somatic point mutations similar to other repertoires. In addition, the presence of other mechanisms for antibody diversification like insertion of codons not encoded by the germline repertoire^{50,51} has been seen in other studies.⁵² In conclusion, H-EPSILON repertoires in general represent a normal part of the total immunoglobulin repertoire. Thus, the evidence from our study supports the notion of extensive diversity in H-EPSILON repertoires^{27–29} over the idea suggesting that H-EPSILON repertoires are highly restricted with respect to V gene usage.^{23–26} In fact, there is nothing particular about the composition of H-EPSILON repertoires as identified during active disease and B cells are in this phase not in general preferentially selected into the H-EPSILON-producing population based on e.g. a specific IGHV gene usage *per se*.

Secondly, we observed that the size of the peripheral blood H-EPSILON-encoding transcriptome was very limited as clonally related sequences were frequently encountered. This is in general not observed when sequencing polyclonal repertoires of heavy chain of other isotypes, like H-GAMMA. Such clonally related sets of sequences in H-EPSILON-encoding transcriptomes have, however, previously been identified by others^{29–31,53} as well, suggesting that they are not a result of the particular experimental approach used by us, but rather a factual circumstance. Thus, a very limited number of B cells populate the IgE-producing peripheral B cell population in allergic individuals during active disease. Nevertheless, as outlined above, H-EPSILON-encoding repertoires are complex in their genetic origin. Furthermore, as clonally related sequences show evidence of divergent somatic hypermutation, it is evident that either different members originating from the same clone are recruited into the IgE-producing population, or the few recruited clones that undergo H-EPSILON class switch are further diversified after being committed to IgE production. Findings probably made possible through the limited size of H-EPSILON repertoires are the identification of similar H-EPSILON-encoding sequences in two individuals and the occurrence of similar rearrangements in the H-EPSILON-encoding transcriptome of a single individual. Indeed, the induction of a response against a given epitope may be favoured by the occurrence of similar V-D-J gene rearrangements. This may not be such a rare event as it may sound to be, as it turns out that similar rearrangements are strongly favoured in other immune responses as well and seem to occur independently in different individuals or repeatedly in a given individual.^{21,53,54} Similar rearrangements have also been observed in certain B cell

tumours,^{55–57} suggesting the involvement of a common antigen stimulus as a driving force in the development of these diseases, too. Thus despite the extensive capacity to create diversity in the B cell compartment, very similar rearrangements occur at a significant frequency and a given antigen recruits similar sequences into a humoral response at an observable frequency, also in the case of H-EPSILON repertoires.

The availability of V-D-J genes representing the H-EPSILON transcriptome allows us not only to study the makeup of the transcriptome as such but also to recreate functional binding sites and to identify specific binders by using any one of the many available combinatorial library and display technologies. The experimental setup makes use of the importance of the H chain and in particular the third hypervariable loop in determining many antibody specificities.⁵⁸ Such libraries, based on the H-EPSILON VH genes derived from the two individuals studied herein, identified several different specific sequences originating from three different IGHV subgroups recognising the allergen *Phl p 5*. The antigen selected for these initial studies was chosen as it is an important clinical allergen targeted by H-EPSILON repertoires found in many grass pollen allergic individuals.^{59–61} Delineation of the composition of this type of specific repertoire may therefore have clinical implications in addition to its contribution to basic knowledge. Our approach allowed us to vastly expand on the knowledge of the diversity of human H-EPSILON repertoires targeting this allergen in comparison to previous studies.³⁴ We concluded that despite the very limited overall sizes of these H-EPSILON repertoires they were populated by very different sequences targeting even a single allergen.

Importantly, we were able to define a *Phl p 5*-specific clonotype also among the clones picked at random for sequencing. This, in our view, underscores the importance of the very limited size of H-EPSILON repertoires, allowing us to connect the overall makeup of repertoires encoding IgE with those targeting a specific antigen. This clonotype, although being commonly represented in donor A3, was rarely found among selected clones in phage-displayed libraries. In contrast, another set of clones, harbouring a single VH in combination with different V-KAPPA or V-LAMBDA, was repeatedly selected from the library despite the fact that it was never found among randomly picked clones. It is likely that this effect is caused by different dependencies or promiscuities of different VH with respect to their VL partner. Indeed, many antibody binding sites as they develop *in vivo* are very dependent on a given VH-VL pairing for efficient antigen recognition, for instance as described by us.⁶² As highlighted by this observation of ours, random combinatorial library construction is likely to skew the selected repertoire towards those that are more promiscuous with respect to their VL partner. Consequently, we believe that the herein proposed approach would further benefit from

incorporating technologies like flow cytometry (to select for IgE-producing B cells) in combination with in-cell PCR⁶³ or single cell-PCR and VH-VL re-assembly *in vitro* to recreate VH-VL pairs that exist *in vivo* (e.g. Symplex technology⁶⁴). As the H-EPSILON-encoding repertoires are so small *in vivo*, such an experimental mode of action is likely to allow us to circumvent the display and selection approach altogether and to rely solely on high throughput screening approaches⁶⁵ to identify antigen-specific binders.

Irrespective of the potential for future methodological developments, the current approach allows us to follow somatic evolution of a human antigen-specific repertoire as it occurs *in vivo*. We suggest that a combination of tools allowing for identification of antigen-specific clones by display technology from, and coincident isolation of clonally related sequences defined by random sequencing of, a small, diverse transcriptome like that encoding H-EPSILON, will be a powerful mode to investigate human antibody repertoires and their evolution as it occurs *in vivo* at the molecular level. This approach will further our understanding of human antibody responses in health and disease.

Materials and Methods

Patient samples and allergens

Two individuals, A3 and A4, suffering from grass pollen allergy, donated blood during the grass pollen season as a part of an extensive study of the immunology of allergy.³⁷ The local ethical committee had approved these studies. These patients were both timothy (*P. pratense*) extract skin prick test positive, showed symptoms of rhinitis at the time of blood sampling and had timothy extract-specific IgE titers above 50 kUA/l.³⁷ A3 and A4 were also positive in clinical skin prick tests for cat and birch (*Betula verrucosa*) allergens (A3) and mugwort (*Artemisia vulgaris*) and birch allergens (A4). In addition, both showed evidence of serum IgE against *Phl p 5A* in an in-house enzyme-linked immunosorbent assay (ELISA) (data not shown). Recombinant *Phl p 5A* was obtained from Biomay (Vienna, Austria) while timothy pollen extracts were kind gifts from Dr P. A. Würtzen (ALK-Abelló, Hørsholm, Denmark).

Amplification and cloning of variable domains encoding the H-EPSILON repertoire

Peripheral blood mononuclear cells were isolated by gradient centrifugation on Lymphoprep (Axis-Shield PoC A. S., Oslo, Norway). Monocytes were depleted with microbead-conjugated anti-CD14 antibodies (Miltenyi Biotec, Bergisch Gladbach, Germany) and magnetic cell separation columns (Miltenyi Biotec) and cells were resuspended and frozen in Trizol (Sigma Chemical Co., St Louis, MO, USA). The samples used in these studies contained 200×10^6 (A3) and 190×10^6 (A4) lymphocytes. mRNA was prepared from the Trizol-solubilized samples after extraction with chloroform and precipitation with ice-cold isopropanol. A high quality of the RNA preparations

was confirmed by use of a Nanodrop spectrophotometer and a 2100 BioAnalyzer (Agilent Technologies, Palo Alto, CA, USA) (A_{260}/A_{280} ratio >1.96 ; $1.54 < 18\text{ S}/28\text{ S}$ ratio >2.05). cDNA was prepared by using the Endo Free RTTM Primer Specific Reverse Transcription Kit (Ambion, Austin, TX, USA). Primers specific for the immunoglobulin-encoding gene sequences (Table 8) were synthesised by MWG-Biotech (Ebersberg, Germany). PCR was carried out using AmpliTaq[®] DNA Polymerase (Applied Biosystem, Foster City, CA, USA) in combination with seven primers annealing to the 5'-end of the seven different IGHV subgroups and one of two different primers annealing within the part of the IGHE gene³⁶ encoding the constant domains of IgE. Two consecutive nested PCR were performed employing the same 5'-primers, and a 3'-primer annealing inside the 3'-primer used in the previous PCR step but still within the CH1 of the IGHE gene. The transcriptome encoding the H-GAMMA repertoire was amplified by an approach including a primary PCR and one nested PCR using IGHG gene-specific 3'-primers (Table 8). PCR-products were cloned directly into pCR[®]2.1 vector, and transformed into One Shot[®] *Escherichia coli* Top 10F' chemically competent cells, according to procedures described for the Original TA Cloning[®] Kit (Invitrogen). Independent colonies were picked and plasmid was purified.

Library construction

To construct antibody libraries encoding scFv, VH-encoding genes adapted for cloning in phagemid vectors were amplified from the PCR product obtained above from H-EPSILON transcripts. Primers annealing to all IGHV gene subgroups and all IGHJ genes were used (Table 8). Sequences encoding V-KAPPA and V-LAMBDA were similarly amplified from cDNA prepared using primers annealing to sequences encoding IGKJC and IGLC (Table 8) by using primers annealing to the major gene subgroups (IGKV1-IGKV4, IGLV1-IGLV3, IGKJ1-IGKJ4, IGLJ1-IGLJ3, IGLJ6 and IGLJ7). This procedure ensures amplification of VL-encoding genes but does not ensure their origin in the IgE-encoding population. VH and VL products obtained in this way were assembled in a six-cycles overlap extension PCR procedure into genes encoding VH-VL fusion proteins with a (glycine₄-serine)₃-linker in-between the two domains. The assembled products were amplified using primers specific for each end of the scFv-encoding gene (Table 8). The PCR-products were cleaved with NotI and SfiI (New England BioLabs, Inc, Beverly, MA, USA) and cloned into a modified version of the pFab5c.His phagemid vector.⁶⁶ Ligated DNA was electroporated into *E. coli* Top 10F'. Sequencing of randomly picked clones ($n=6$ (A3), $n=6$ (A4)) confirmed that the library carried members with a different genetic background including VL domains with subgroups IGKV1-IGKV3 and IGLV1-IGLV3. Phage stocks were prepared by using VCSM13 helper phage. This technology, although it generally establishes vectors encoding functional, well-folded proteins, does not ensure faithful recombination of VH-VL pairs as they occurred in immunoglobulins *in vivo*, a problem associated to most conventional library development modes used today. Nevertheless, as much of the specificity of paratopes resides in the heavy chain variable domain and its third hypervariable loop,⁵⁸ and as these were obtained by a stringent PCR amplification process from the H-EPSILON-encoding transcriptome, this procedure in combination with an antigen-specific selection process provides an

Table 8. Primers used for (1) cDNA synthesis, (2) for amplification of VH from H-EPSILON cDNA, (3) for amplification of VH, V-KAPPA and V-LAMBDA and complete scFv for scFv combinatorial library construction, and (4) for amplification of VH and VL for cloning in eukaryotic expression vectors

Purpose	Immunoglobulin nucleotides and amino acids ^a	Primer specificity	Primer sequence ^{b,c}
cDNA synthesis	Nt 18-34, AA 6-12	IGHE CH2	<i>CGACTGTAAACTAGTCACGGTGGGCGGGTG^d</i>
	Nt 262-285, AA 88-95	IGHG1-IGHG4 CH1	<i>GTCCACCTTGGTGTGTCTGGGCTT</i>
	Nt 298-318, AA 100-106	IGKC	<i>AGACTCTCCCTGTGAAGCTCTT</i>
	Nt 292-313, AA 98-105	IGLC1-IGLC2	<i>TGAAGATTCTGTAGGGGCCACTGTCTT</i>
VH amplification for TA-cloning ^e	Nt 1-23, AA 1-8	IGHV1 FR1	<i>CAGGTGCAGCTGGTGCAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV2 FR1	<i>CAGRTACCTTGAAGGAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV3 FR1	<i>GAGGTGCAGCTGGTGGAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV4 FR1	<i>CAGGTGCAGCTGCAGGAGTCCGG</i>
	Nt 1-23, AA 1-8	IGHV5 FR1	<i>GAGGTGCAGCTGGTGCAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV6 FR1	<i>CAGGTACAGCTGCAGCAGTCAGG</i>
	Nt 1-23, AA 1-8	IGHV7 FR1	<i>CAGGTGCAGCTGGTGCATCTGG</i>
	Nt 175-198, AA 59-66	IGHE CH1 ver. 1	<i>ATATGTACACAGAGAGCGTGAGGGT</i>
	Nt 49-72, AA 17-24	IGHE CH1 ver. 2	<i>GGAGGTGGCATTGGAGGGAATGTT</i>
	Nt 9-32, AA 3-11	IGHE CH1 ver. 3 ^f	<i>AAGGGGAAGACGGATGGGCTCTGT</i>
	Nt 10-32, AA 4-11	IGHE CH1 ver. 4 ^f	<i>AAGGGGAAGACGGATGGGCTCTG</i>
	Nt 166-192, AA 56-64	IGHG1-IGHG4 CH1 ver. 1	<i>GGAGTAGAGTCTGAGGACTGTAGGAC</i>
	Nt 91-114, AA 31-38	IGHG1-IGHG4 CH1 ver. 2	<i>CGTCACCGGTTCGGGGAAGTAGTC</i>
VH amplification for library construction	Nt 1-23, AA 1-8	IGHV1 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGGTGCAGCTGGTGCAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV2 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGRTACCTTGAAGGAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV3 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGGTGCAGCTGGTGGAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV4 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGGTGCAGCTGCAGGAGTCGGG</i>
	Nt 1-23, AA 1-8	IGHV5 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGGTGCAGCTGGTGCAGTCTGG</i>
	Nt 1-23, AA 1-8	IGHV6 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGGTGCAGCTGCAGCAGTCAGG</i>
	Nt 1-23, AA 1-8	IGHV7 FR1	<i>GGCCGATTCAATTAATGCAGCTGGCCAGCCGGCCATGGCCAGGTGCAGCTGGTGCATCTGG</i>
	Nt 36-52, AA 12-17	IGHJ	<i>CCGCCGATCCACCTCCGCCTGAACCCGCTCCACCTGAGGAGACGGTGACC</i>
V-KAPPA amplification for library construction	Nt 1-22, AA 1-8	IGKV1 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGGACATCCAGATGACCCAGTCTC</i>
	Nt 1-22, AA 1-8	IGKV2 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGGATATTGTGATGACYCAGWCTC</i>
	Nt 1-22, AA 1-8	IGKV3 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGGAAATTGTGWTGACRCAGTCTC</i>
	Nt 1-22, AA 1-8	IGKV4 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGGACATCGTGATGACCCAGTCTC</i>
	Nt 18-37, AA 6-12	IGKJ1,4	<i>GCCTTTAGCGTCAGACTGTAGTGCGGCCGCTTTATCATCATCATCTTTATAATCTCGTTTGATYTCCACCTTGGTCC</i>
	Nt 18-37, AA 6-12	IGKJ2	<i>GCCTTTAGCGTCAGACTGTAGTGCGGCCGCTTTATCATCATCATCTTTATAATCTCGTTTGATCTCCAGCTTGGTCC</i>
	Nt 18-37, AA 6-12	IGKJ3	<i>GCCTTTAGCGTCAGACTGTAGTGCGGCCGCTTTATCATCATCATCTTTATAATCTCGTTTGATATCCACTTGGTCC</i>
V-LAMBDA amplification for library construction	Nt 1-19, AA 1-7	IGLV1 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGCAGTCTGTGCTGACKCAGC</i>
	Nt 1-19, AA 1-7	IGLV2 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGCAGTCTGCCCTGACTCAGC</i>
	Nt 1-19, AA 1-7	IGLV3 FR1	<i>GGAGGCGGTTCAAGGCGGAGGTGGATCCGGCGGTGGCGGATCGTCTATGAGCTGACWCAGC</i>
	Nt 17-36, AA 6-12	IGLJ	<i>GCCTTTAGCGTCAGACTGTAGTGCGGCCGCTTTATCATCATCATCTTTATAATCACCKAGGACGGTCAGCTTGGTSCC</i>
Amplification of intact scFv-encoding genes		5'-end	<i>GGCCGATTCAATTAATGCAGC</i>
		3'-end	<i>GCCTTTAGCGTCAGACTGTAG</i>
Amplification of VH and VL for eukaryotic expression vectors		VH 5'-end	<i>GCGATGGTGTGCAATCCGAGGTGCAGCTGGTGCAGTCTGG</i>
		VH 3'-end	<i>GCGGATGACGTACGACTCACCTGAGGAGACGGTGACC</i>
		VL 5'-end	<i>GCGATGGTGTGCAATCTCTCTATGAGCTGACACAGCC</i>
		VL 3'-end	<i>GCGAGAGACGTACGTTCTACTCACCTAGGACGGTCAGCT</i>

^a Primers anneal to these nucleotides (Nt) and encodes these amino acids (AA).^b Restriction enzyme sites are underlined, other nucleotides unrelated to immunoglobulin or scFv-encoding sequences are indicated in italics.^c K=G or T; R=A or G; S=G or C; Y=C or T; W=A or T.^d Primer originally designed by Steinberger *et al.*³⁴^e Alternative primers used in different PCR.^f Variant of IGHE CH1 ver. 3 primer to allow for definitive assignment of origin of PCR product in H-EPSILON-encoding transcriptome.

acceptable mimic of antigen-specific antibodies as they occur *in vivo*.

Phage display selection of clones specific for timothy allergens

To identify allergen-specific antibody fragments in the combinatorial libraries, the libraries were selected against recombinant *Phl p 5A* and timothy pollen extract. The selections were performed after coating the antigens at 5 µg/ml in PBS to ELISA microtitre plates (Costar, Cambridge, MA, USA). PBS containing 0.05% Tween 20 (TPBS) containing 3% BSA was used as blocking agent. Unbound phages were washed away and the remaining phages were eluted with trypsin (Sigma Chemical Co.) that cleaves a trypsin-sensitive site between the phage particle and the displayed protein.⁶⁷ After infection of *E. coli*, new phage stocks were prepared for the next round of selection. The selection procedure was repeated three times. After the last round of selection, individual colonies encoding scFv were picked for sequence analysis and determination of binding specificity.

Genetic analysis

TA-cloned genes encoding VH sequences and genes encoding scFv were sequenced directly from plasmids at an external sequencing centre (MWG-Biotech). Their genetic origin and mutational status was determined by using the bioinformatics tools IMGT/V-QUEST⁴⁴ and IMGT/Junction Analysis⁴⁵ associated with IMGT, The International ImLunoGeneTics information system†⁶⁸ Bases that were defined by the primers used for DNA amplification were not included in the analysis. The antibody-encoding gene nomenclature, CDR definitions, and numbering of residues are in agreement with the IMGT definitions.^{36,69} Sequences were considered to be clonally related to each other if they originated from the same IGHV germline gene, encoded a CDR3 (between C104 and W118) of identical length and had at least 80% nucleotide identity in this part of the gene.

Immunological analysis

Antigens, including *Phl p 5A*, a recombinant fragment of *Phl p 5A* (residues 56–165), *Bet v 1*, extracts of timothy grass pollen or *L. perenne* pollen, and BSA, were coated at 5 µg protein/ml on ELISA microtitre plates (Costar or Nunc (Roskilde, Denmark)). Phage stocks displaying scFv obtained from clones picked from the third selection round were diluted in TPBS containing 1% (w/v) skim milk powder (Semper, Stockholm, Sweden) and added to the plates. For detection, HRP-labelled anti-M13 antibody (Amersham) was used in combination with the chromogen *o*-phenylenediamine. Plates were read at 490 nm in a microtitre plate reader. Clones were considered positive when the absorbance was threefold greater than that seen on the blank. Allergen-specific phage-displayed scFv were also assessed for their binding to phage λgt11-encoded recombinant *Phl p 5A* and fragments of *Lol p 5A* and *5B*.⁴⁶ Phage stocks, diluted in TPBS containing 1% skim milk powder, were added to filters carrying these proteins. Bound phages were detected using HRP-labelled anti-M13 and the ECL Plus Western Blotting Detection System

(Amersham). Light emission was recorded with a LAS1000plus cooled CCD camera system (Fujifilm, Tokyo, Japan) and images were evaluated using Image Gauge (Fujifilm).

Analysis of the binding of intact recombinant antibodies was carried out by similar means as those used for the analysis of antigen-specific binding of scFv. HRP-labelled antibodies specific for human IgG (Zymed Laboratories Inc., South San Francisco, CA, USA) were used to detect specific binding to immobilised antigen. Human IgG was quantified as previously described by ELISA using isotype-specific catcher and HRP-labelled antibodies (Zymed Laboratories Inc.). The IgG concentration was determined in comparison to a human immunoglobulin preparation (Gammagard S/D; Baxter, Chicago, USA) of known concentration.

To further assess the fine specificity of antibodies and antibody fragments targeting a major immunogenic fragment (residues 56–165) of *Phl p 5*, a blocking assay was performed employing label-free surface plasmon resonance detection in a Biacore system. Briefly, r*Phl p 5* (approximately 2500 RU) was immobilised on a CM5 sensorchip using the Amine Coupling Kit (Biacore AB, Uppsala Sweden). Samples were diluted in 10 mM Hepes (pH 7.4), 3.4 mM EDTA, 150 mM NaCl, 0.005% P20 surfactant (Biacore AB) and run over the immobilised allergen at a flow rate of 20 µl/min. The binding of a *Phl p 5*-specific antibody or antibody fragment was assessed after injection of buffer alone, non-binding antibodies or antibodies/antibody fragments specific for *Phl p 5* (in the latter case after saturation of their epitopes on the immobilised allergen). The system allows for assessment of both the virtual saturation of the epitopes on the immobilised antigen by the blocking antibody, and the residual binding ability of the second antibody under these conditions. The system was regenerated with 50 mM glycine, 500 mM NaCl (pH 2.55).

Production of intact allergen-specific antibodies

In order to produce intact antibodies (human IgG1) specific for *Phl p 5*, VH and VL of scFv 4.2 were amplified by PCR using primers outlined in Table 8. In addition, two VH, A3-EE50 and A3-EE50B, that are clonally related to the VH found in scFv 4.2 and that were found among randomly picked, TA-cloned H-EPILON VH sequences obtained from donor A3, were amplified in the same way. The products were cloned after BsmI and BsiWI restriction enzyme cleavage into vectors GlnCm5 and Lambda-dhmK7 (BioInvent International AB, Lund, Sweden) suitable for transient production of intact antibody. DNA was transfected into CHO-K1 cells, a generous gift from Professor G. C. Hansson (Department of Medical Biochemistry, Göteborg University, Göteborg, Sweden) with the aid of Lipofectamine 2000 (Invitrogen). Cell culture supernatants containing IgG were harvested after 24 and 48 h and used without further treatment.

Acknowledgements

This study was supported by grants from the Swedish Asthma and Allergy Association's Research Foundation, Alfred Österlund's Foundation, the Swedish Research Council, the Austrian Science

† <http://imgt.cines.fr/>

Fund (grants number F1815 and T165-B09) and by a research grant from Biomay, Vienna Austria.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2006.06.062](https://doi.org/10.1016/j.jmb.2006.06.062)

References

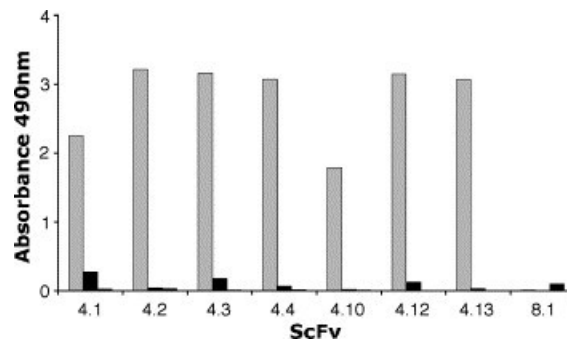
- Guigou, V., Cuisinier, A. M., Tonnelle, C., Moinier, D., Fougereau, M. & Fumoux, F. (1990). Human immunoglobulin VH and VK repertoire revealed by in situ hybridization. *Mol. Immunol.* **27**, 935–940.
- de Wildt, R. M., Hoet, R. M., van Venrooij, W. J., Tomlinson, I. M. & Winter, G. (1999). Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire. *J. Mol. Biol.* **285**, 895–901.
- Kohsaka, H., Carson, D. A., Rassenti, L. Z., Ollier, W. E., Chen, P. P., Kipps, T. J. & Miyasaka, N. (1996). The human immunoglobulin V(H) gene repertoire is genetically controlled and unaltered by chronic autoimmune stimulation. *J. Clin. Invest.* **98**, 2794–2800.
- Yamada, M., Wasserman, R., Reichard, B. A., Shane, D., Caton, A. J. & Rovera, G. (1991). Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. *J. Exp. Med.* **173**, 395–403.
- Dörner, T., Brezinschek, H. P., Foster, S. J., Brezinschek, R. I., Farner, N. L. & Lipsky, P. E. (1998). Comparable impact of mutational and selective influences in shaping the expressed repertoire of peripheral IgM⁺/CD5⁻ and IgM⁺/CD5⁺ B cells. *Eur. J. Immunol.* **28**, 657–668.
- Foster, S. J., Brezinschek, H. P., Brezinschek, R. I. & Lipsky, P. E. (1997). Molecular mechanisms and selective influences that shape the kappa gene repertoire of IgM⁺ B cells. *J. Clin. Invest.* **99**, 1614–1627.
- Küppers, R., Zhao, M., Hansmann, M. L. & Rajewsky, K. (1993). Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *EMBO J.* **12**, 4955–4967.
- Rogozin, I. B. & Diaz, M. (2004). Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCV motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.* **172**, 3382–3384.
- Rogozin, I. B. & Kolchanov, N. A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta*, **1171**, 11–18.
- Spencer, J., Dunn, M. & Dunn-Walters, D. K. (1999). Characteristics of sequences around individual nucleotide substitutions in IgVH genes suggest different GC and AT mutators. *J. Immunol.* **162**, 6596–6601.
- Ohlin, M. & Borrebaeck, C. A. K. (1996). Characteristics of human antibody repertoires following active immune responses in vivo. *Mol. Immunol.* **33**, 583–592.
- Ohlin, M. & Zouali, M. (2003). The human antibody repertoire to infectious agents: implications for disease pathogenesis. *Mol. Immunol.* **40**, 1–11.
- Zhou, J., Lottenbach, K. R., Barenkamp, S. J., Lucas, A. H. & Reason, D. C. (2002). Recurrent variable region gene usage and somatic mutation in the human antibody response to the capsular polysaccharide of *Streptococcus pneumoniae* type 23F. *Infect. Immun.* **70**, 4083–4091.
- Lucas, A. H. & Reason, D. C. (1999). Polysaccharide vaccines as probes of antibody repertoires in man. *Immunol. Rev.* **171**, 89–104.
- Ohlin, M., Owman, H., Rioux, J. D., Newkirk, M. M. & Borrebaeck, C. A. K. (1994). Restricted variable region gene usage and possible rheumatoid factor relationship among human monoclonal antibodies specific for the AD-1 epitope on cytomegalovirus glycoprotein B. *Mol. Immunol.* **31**, 983–991.
- Chardès, T., Chapal, N., Bresson, D., Bés, C., Giudicelli, V., Lefranc, M. P. & Peraldi-Roux, S. (2002). The human anti-thyroid peroxidase autoantibody repertoire in Graves' and Hashimoto's autoimmune thyroid diseases. *Immunogenetics*, **54**, 141–157.
- Edwards, B. M., Barash, S. C., Main, S. H., Choi, G. H., Minter, R., Ullrich, S. *et al.* (2003). The remarkable flexibility of the human antibody repertoire; isolation of over one thousand different antibodies to a single protein, BlyS. *J. Mol. Biol.* **334**, 103–118.
- van Den Brink, E. N., Bril, W. S., Turenhout, E. A., Zuurveld, M., Bovenschen, N., Peters, M. *et al.* (2002). Two classes of germline genes both derived from the V(H)1 family direct the formation of human antibodies that recognize distinct antigenic sites in the C2 domain of factor VIII. *Blood*, **99**, 2828–2834.
- Chan, C. H., Hadlock, K. G., Fong, S. K. & Levy, S. (2001). V(H)1-69 gene is preferentially used by hepatitis C virus-associated B cell lymphomas and by normal B cells responding to the E2 viral antigen. *Blood*, **97**, 1023–1026.
- Ihle, Ø., Beckström, K. J. & Michaelsen, T. E. (2003). Cloning, sequencing and expression of immunoglobulin variable regions of murine monoclonal antibodies specific for the P1.7 and P1.16 PorA protein loops of *Neisseria meningitidis*. *Scand. J. Immunol.* **57**, 453–462.
- McLean, G. R., Olsen, O. A., Watt, I. N., Rathanaswami, P., Leslie, K. B., Babcock, J. S. & Schrader, J. W. (2005). Recognition of human cytomegalovirus by human primary immunoglobulins identifies an innate foundation to an adaptive immune response. *J. Immunol.* **174**, 4768–4778.
- Feeney, A. J., Atkinson, M. J., Cowan, M. J., Escuro, G. & Lugo, G. (1996). A defective V kappa A2 allele in Navajos which may play a role in increased susceptibility to *Haemophilus influenzae* type b disease. *J. Clin. Invest.* **97**, 2277–2282.
- Edwards, M. R., Brouwer, W., Choi, C. H., Ruhno, J., Ward, R. L. & Collins, A. M. (2002). Analysis of IgE antibodies from a patient with atopic dermatitis: biased V gene usage and evidence for polyreactive IgE heavy chain complementarity-determining region 3. *J. Immunol.* **168**, 6305–6313.
- Janežic, A., Chapman, C. J., Snow, R. E., Hourihane, J. O., Warner, J. O. & Stevenson, F. K. (1998). Immunogenetic analysis of the heavy chain variable regions of IgE from patients allergic to peanuts. *J. Allergy Clin. Immunol.* **101**, 391–396.
- Snow, R. E., Chapman, C. J., Frew, A. J., Holgate, S. T. & Stevenson, F. K. (1995). Analysis of Ig VH region genes encoding IgE antibodies in splenic B

- lymphocytes of a patient with asthma. *J. Immunol.* **154**, 5576–5581.
26. Snow, R. E., Djukanovic, R. & Stevenson, F. K. (1999). Analysis of immunoglobulin E VH transcripts in a bronchial biopsy of an asthmatic patient confirms bias towards VH5, and indicates local clonal expansion, somatic mutation and isotype switch events. *Immunology*, **98**, 646–651.
 27. Tilgner, J., Golembowski, S., Kersten, B., Sterry, W. & Jahn, S. (1997). VH genes expressed in peripheral blood IgE-producing B cells from patients with atopic dermatitis. *Clin. Exp. Immunol.* **107**, 528–535.
 28. Eibensteiner, P., Spitzauer, S., Steinberger, P., Kraft, D. & Valenta, R. (2000). Immunoglobulin E antibodies of atopic individuals exhibit a broad usage of VH-gene families. *Immunology*, **101**, 112–119.
 29. Davies, J. M. & O'Hehir, R. E. (2004). VH gene usage in immunoglobulin E responses of seasonal rhinitis patients allergic to grass pollen is oligoclonal & antigen driven. *Clin. Exp. Allergy*, **34**, 429–436.
 30. Efremov, D. G., Batista, F. D. & Burrone, O. R. (1993). Molecular analysis of IgE H-chain transcripts expressed in vivo by peripheral blood lymphocytes from normal and atopic individuals. *J. Immunol.* **151**, 2195–2207.
 31. van der Stoep, N., van der Linden, J. & Logtenberg, T. (1993). Molecular evolution of the human immunoglobulin E response: high incidence of shared mutations and clonal relatedness among epsilon VH5 transcripts from three unrelated patients with atopic dermatitis. *J. Exp. Med.* **177**, 99–107.
 32. de Lalla, C., Tamborini, E., Longhi, R., Tresoldi, E., Manoni, M., Siccardi, A. G., Arosio, P. & Sidoli, A. (1996). Human recombinant antibody fragments specific for a rye-grass pollen allergen: characterization and potential applications. *Mol. Immunol.* **33**, 1049–1058.
 33. Flicker, S., Steinberger, P., Norderhaug, L., Sperr, W. R., Majlesi, Y., Valent, P. *et al.* (2002). Conversion of grass pollen allergen-specific human IgE into a protective IgG(1) antibody. *Eur. J. Immunol.* **32**, 2156–2162.
 34. Steinberger, P., Kraft, D. & Valenta, R. (1996). Construction of a combinatorial IgE library from an allergic patient. Isolation and characterization of human IgE Fabs with specificity for the major timothy grass pollen allergen. *Phl p 5. J. Biol. Chem.* **271**, 10967–10972.
 35. Laukkanen, M. L., Makinen-Kiljunen, S., Isoherranen, K., Haahtela, T., Söderlund, H. & Takkinen, K. (2003). Hevein-specific recombinant IgE antibodies from human single-chain antibody phage display libraries. *J. Immunol. Methods*, **278**, 271–281.
 36. Lefranc, M. P. & Lefranc, G. (2001). *The Immunoglobulin FactsBook*. Academic Press, London.
 37. Lindstedt, M., Schiött, A., Bengtsson, A., Larsson, K., Korsgren, M., Greiff, L. & Borrebaeck, C. A. K. (2005). Genomic and functional delineation of dendritic cells and memory T cells derived from grass pollen-allergic patients and healthy individuals. *Int. Immunol.* **17**, 401–409.
 38. Corbett, J. S., Tomlinson, I. M., Sonnhammer, E. L. L., Buck, D. & Winter, G. (1997). Sequence of the human immunoglobulin diversity (D) segment locus: A systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segment or D-D recombination. *J. Mol. Biol.* **270**, 587–597.
 39. Souto-Carneiro, M. M., Longo, N. S., Russ, D. E., Sun, H. W. & Lipsky, P. E. (2004). Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *J. Immunol.* **172**, 6790–6802.
 40. Wu, T. T., Johnson, G. & Kabat, E. E. (1993). Length distribution of CDRH3 in antibodies. *Proteins: Struct. Funct. Genet.* **16**, 1–7.
 41. Yu, K., Huang, F. T. & Lieber, M. R. (2004). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J. Biol. Chem.* **279**, 6496–6500.
 42. Keohavong, P. & Thilly, W. G. (1989). Fidelity of DNA polymerase in DNA amplification. *Proc. Natl Acad. Sci. USA*, **86**, 9253–9257.
 43. Betz, A. G., Neuberger, M. S. & Milstein, C. (1993). Discriminating intrinsic and antigen-selected mutational hotspots in immunoglobulin V genes. *Immunol. Today*, **14**, 405–411.
 44. Giudicelli, V., Chaume, D. & Lefranc, M. P. (2004). IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucl. Acids Res.* **32**, W435–W440.
 45. Monod, M. Y., Giudicelli, V., Chaume, D. & Lefranc, M. P. (2004). IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics*, **20**, I379–I385.
 46. Flicker, S., Vrtala, S., Steinberger, P., Vangelista, L., Bufe, A., Petersen, A. *et al.* (2000). A human monoclonal IgE antibody defines a highly allergenic fragment of the major timothy grass pollen allergen, Phl p 5: molecular, immunological, and structural characterization of the epitope-containing domain. *J. Immunol.* **165**, 3849–3859.
 47. Fransson, J., Ek, S., Ellmark, P., Söderlind, E., Borrebaeck, C. A. K. & Furebring, C. (2004). Profiling of internalizing tumor-associated antigens on breast and pancreatic cancer cells by reversed genomics. *Cancer Letters*, **208**, 235–242.
 48. Foote, J. & Milstein, C. (1994). Conformational isomerism and the diversity of antibodies. *Proc. Natl Acad. Sci. USA*, **91**, 10370–10374.
 49. Feeney, A. J., Tang, A. & Ogward, K. M. (2000). B-cell repertoire formation. Role of the recombination signal sequence in non-random V segment utilization. *Immunol. Rev.* **175**, 59–69.
 50. Ohlin, M. & Borrebaeck, C. A. K. (1998). Insertions and deletions in hypervariable loops of antibody heavy chains contribute to molecular diversity. *Mol. Immunol.* **35**, 233–238.
 51. Wilson, P. C., de Bouteiller, O., Liu, Y. J., Potter, K., Banchereau, J., Capra, J. D. & Pascual, V. (1998). Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J. Exp. Med.* **187**, 59–70.
 52. Snow, R. E., Chapman, C. J., Frew, A. J., Holgate, S. T. & Stevenson, F. K. (1997). Pattern of usage and somatic hypermutation in the V(H)5 gene segments of a patient with asthma: implications for IgE. *Eur. J. Immunol.* **27**, 162–170.
 53. Bando, Y., Shimizu, A. & Ra, C. (2004). Characterization of V_Hε gene expressed in PBL from children with atopic diseases: detection of homologous V_H1-69 derived from transcripts from three unrelated patients. *Immunol. Letters*, **94**, 99–106.
 54. Seidl, K. J., MacKenzie, J. D., Wang, D., Kantor, A. B., Kabat, E. A. & Herzenberg, L. A. (1997). Frequent occurrence of identical heavy and light chain Ig rearrangements. *Int. Immunol.* **9**, 689–702.

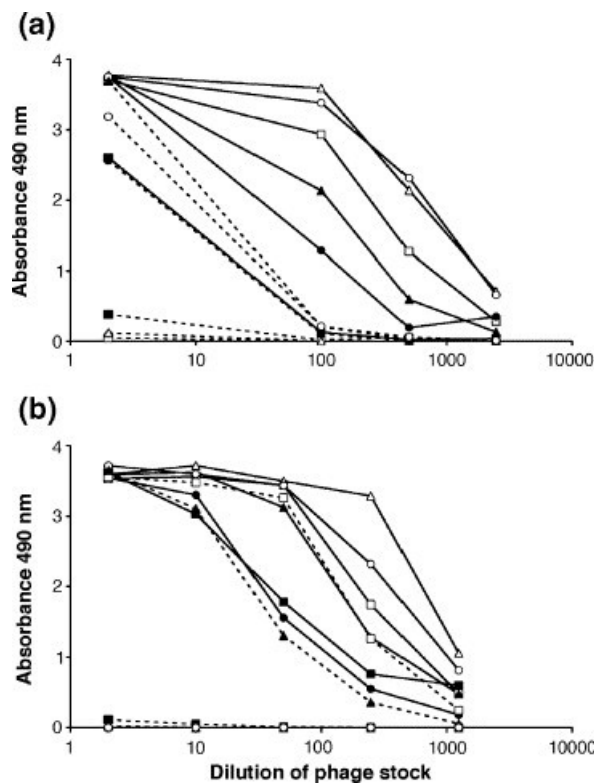
55. Tobin, G., Thunberg, U., Karlsson, K., Murray, F., Laurell, A., Willander, K. *et al.* (2004). Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. *Blood*, **104**, 2879–2885.
56. Stamatopoulos, K., Belessi, C., Hadzidimitriou, A., Smilevska, T., Kalagiakou, E., Hatzi, K. *et al.* (2005). Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. *Blood*, **106**, 3575–3583.
57. Walsh, S. H. & Rosenquist, R. (2005). Immunoglobulin gene analysis of mature B-cell malignancies: reconsideration of cellular origin and potential antigen involvement in pathogenesis. *Med. Oncol.* **22**, 327–341.
58. Xu, J. L. & Davis, M. M. (2000). Diversity in the CDR3 region of V_H is sufficient for most antibody specificities. *Immunity*, **13**, 37–45.
59. Vrtala, S., Sperr, W. R., Reimitzer, I., van Ree, R., Laffer, S., Müller, W. D. *et al.* (1993). cDNA cloning of a major allergen from timothy grass (*Phleum pratense*) pollen; characterization of the recombinant Phl pV allergen. *J. Immunol.* **151**, 4773–4781.
60. Andersson, K. & Lidholm, J. (2003). Characteristics and immunobiology of grass pollen allergens. *Int. Arch. Allergy Immunol.* **130**, 87–107.
61. Niederberger, V., Laffer, S., Froschl, R., Kraft, D., Rumpold, H., Kapiotis, S. *et al.* (1998). IgE antibodies to recombinant pollen allergens (Phl p 1, Phl p 2, Phl p 5, and Bet v 2) account for a high percentage of grass pollen-specific IgE. *J. Allergy Clin. Immunol.* **101**, 258–264.
62. Ohlin, M., Owman, H., Mach, M. & Borrebaeck, C. A. K. (1996). Light chain shuffling of a high affinity antibody results in a drift in epitope recognition. *Mol. Immunol.* **33**, 47–56.
63. Embleton, M. J., Gorochoy, G., Jones, P. T. & Winter, G. (1992). In-cell PCR from mRNA: amplifying and linking the rearranged immunoglobulin heavy and light chain V-genes within single cells. *Nucl. Acids Res.* **20**, 3831–3837.
64. Haurum, J. & Bregenholt, S. (2005). Recombinant polyclonal antibodies: therapeutic antibody technologies come full circle. *Idrugs*, **8**, 404–409.
65. Hallborn, J. & Carlsson, R. (2002). Automated screening procedure for high-throughput generation of antibody fragments. *Biotechniques*, **33**, Suppl: 30–37.
66. Engberg, J., Andersen, P. S., Nielsen, L. K., Dziegiel, M., Johansen, L. K. & Albrechtsen, B. (1996). Phage-display libraries of murine and human antibody Fab fragments. *Mol. Biotechnol.* **6**, 287–310.
67. Johansen, L. K., Albrechtsen, B., Andersen, H. W. & Engberg, J. (1995). pFab60: a new, efficient vector for expression of antibody Fab fragments displayed on phage. *Protein Eng.* **8**, 1063–1067.
68. Lefranc, M. P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D. *et al.* (2005). IMGT, the international ImMunoGeneTics information system. *Nucl. Acids Res.* **33**, D593–D597.
69. Lefranc, M. P., Pommie, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L. *et al.* (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* **27**, 55–77.

Edited by A. R. Fersht

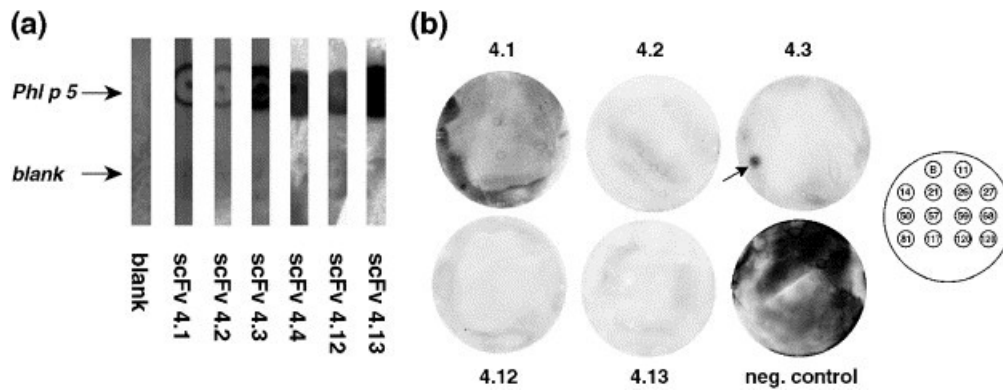
(Received 28 February 2006; received in revised form 1 June 2006; accepted 5 June 2006)
Available online 14 August 2006



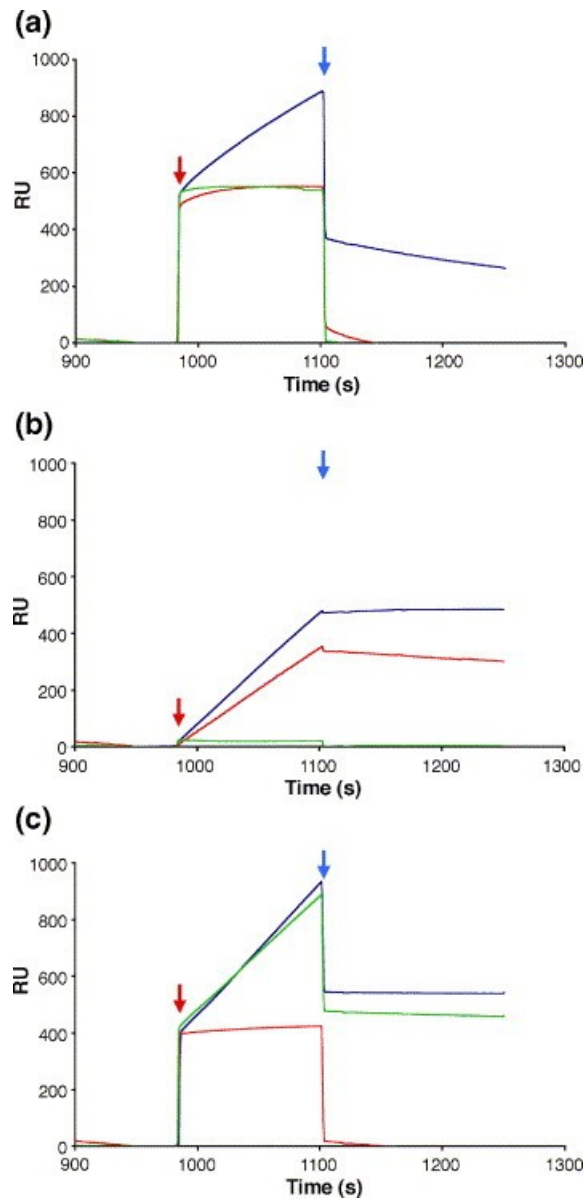
Supplementary Figure S1. Specificity of scFv selected on timothy allergens as demonstrated by binding of scFv displayed on phage to *rPhl p 5A* (light grey), BSA (black) and the unrelated recombinant allergen *Bet v 1* (dark grey). ScFv 8.1 with an unknown specificity was included as a negative control.



Supplementary Figure S2. Binding of scFv displayed on phage to recombinant *Phl p 5A* or fragments thereof and extracts of *Lolium perenne*. A. ELISA demonstrated the ability of scFv 4.3 (filled triangle), 4.12 (filled circle), 4.13 (open circle) and to some extent 4.1 (filled squares) to recognise epitopes not only in *Phl p 5* (solid line) but also in *Lolium perenne* extracts (broken line), presumably on *Lol p 5*. In contrast, scFv 4.2 (open squares) and 4.4 (open triangles) were unable to detect any epitope in *Lolium perenne* antigen preparations. B. Among the scFv only 4.2 and 4.3 strongly recognised the *Phl p 5A* fragment 56-165 (broken line) immobilised onto microtitre plates, while scFv 4.1 showed minimal reactivity, although all sFv recognised intact *Phl p 5A* (solid line). Other tested scFv (4.4, 4.12 and 4.13) were non-reactive in this assay for binding to the *Phl p 5A* fragment 56-165.



Supplementary Figure S3. Specificity of scFv selected against *Phl p 5*. A. Immunoblot analysis of scFv displayed on phage and their binding to bacteria infected with λ gt11 encoding *Phl p 5A* or a control vector (blank) without a recombinant insert. All selected scFv bound specifically to bacteria producing recombinant *Phl p 5A* although with different intensities. ScFv 4.10 was not included in this assay. B. Immunoblot analysis of scFv displayed on phage and their binding to recombinant fragments of *Lol p 5*. The proteins, the positioning of which are indicated to the right, are those described by Flicker *et al.* (2000). B indicates a vector without an insert. The control antibody displayed on phage is specific for human mucin-1 (Persson *et al.*, 2005). ScFv 4.3 was the only antibody fragment that gave a positive signal and it did so only with antigen fragment 81 of *Lol p 5B* (arrow). ScFv 4.12 and 4.13 did not recognise any fragment despite their strong binding to *Phl p 5* in this assay format. References: (Flicker, S. *et al.* 2000. *J. Immunol.* 165: 3849–3859; Persson, J. *et al.* 2005. *Mol. Immunol.* 42: 1321–1330).



Supplementary Figure S4. Blocking of binding of 4.2 IgG (A), 4.3 IgG (B) and clone 5 Fab (C) to immobilised *Phl p 5* after prior binding of clone 5 Fab (red) or 4.3 IgG (green) to allergen as assessed by Biacore technology. Binding of the respective antibodies/fragments in the absence of the addition of a blocking reagent is also shown (blue). The start (red arrow) and end (blue arrow) of the injection of the antibody/Fab fragment under study is highlighted. A bulk effect originating from the solution injected in panel A and C is observed during this period but lost immediately after the injection phase. The binding of the blocking reagent occurred prior to the events shown in this diagram but was evident by an increase in resonance units (RU). As the blocking reagent was injected three times an almost complete saturation of its binding sites could be observed during this period of time (not shown).