Epitope-specificity of recombinant antibodies reveals promiscuous peptide-binding properties

Niclas Olsson,^{1,2} Stefan Wallin,³ Peter James,^{1,2} Carl A. K. Borrebaeck,^{1,2} and Christer Wingren^{1,2*}

¹Department of Immunotechnology, Lund University, Lund, Sweden

²CREATE Health, BMC D13, Lund, Sweden

³Department of Astronomy and Theoretical Physics, Computational Biology and Biological Physics Unit, Lund University, Lund, Sweden

Received 6 July 2012; Accepted 26 September 2012

DOI: 10.1002/pro.2173

Published online 3 October 2012 proteinscience.org

Abstract: Protein–peptide interactions are a common occurrence and essential for numerous cellular processes, and frequently explored in broad applications within biology, medicine, and proteomics. Therefore, understanding the molecular mechanism(s) of protein–peptide recognition, specificity, and binding interactions will be essential. In this study, we report the first detailed analysis of antibody–peptide interaction characteristics, by combining large-scale experimental peptide binding data with the structural analysis of eight human recombinant antibodies and numerous peptides, targeting tryptic mammalian and eukaryote proteomes. The results consistently revealed that promiscuous peptide-binding interactions, that is, both specific and degenerate binding, were exhibited by all antibodies, and the discovery was corroborated by orthogonal data, indicating that this might be a general phenomenon for low-affinity antibody–peptide interactions. The molecular mechanism for the degenerate peptide-binding specificity appeared to be executed through the use of 2–3 semi-conserved anchor residues in the C-terminal part of the peptides, in analogue to the mechanism utilized by the major histocompatibility complex–peptide complexes. In the long-term, this knowledge will be instrumental for advancing our fundamental understanding of protein–peptide interactions, as well as for designing, generating, and applying peptide specific antibodies, or peptide-binding proteins in general, in various biotechnical and medical applications.

Keywords: antibody specificity; anti-peptide antibody; peptide binding; immunoaffinity peptide capture; mass spectrometry

Abbreviations: CIMS, context independent motif specific; CDRs, complementarity-determining regions; FDR, false discovery rate; MHC, major histocompatibility complex; RMSD, root-mean-square-deviation; scFv, single-chain fragment variable; TCR, T-cell receptors.

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Swedish National Research Council (VR-NT), the Foundation for Strategic Research (SSF) (Strategic Centre for Translational Cancer Research—CREATE Health); VINNOVA.

*Correspondence to: Christer Wingren, Department of Immunotechnology, Lund University, BMC D13, 221 84 Lund, Sweden. E-mail: christer.wingren@immun.lth.se

Introduction

Protein–peptide interactions frequently play key roles in many cellular processes in cells, such as signal transduction, regulation, formation of protein networks, and immune recognition.^{1,2} In the latter case, specific recognition and binding of foreign peptides by molecules of the immune system, such as T-cell receptors (TCR) and antibodies, mediate immune recognition and the induction of immune responses. Protein–peptide interactions, and in particular antibody–peptide interactions, have been explored in various applications, such as biosensors,

biomarkers, therapeutics, and functional modulation of proteins.^{2–7} Hence, increasing our fundamental understanding of protein–peptide recognition, and having the possibility to design, control, and even predict novel protein–peptide interactions will have a significant impact on broad applications within the fields of proteomics, biology, and medicine.^{2,6}

Antibodies and other proteins can interact specifically with short peptide sequences in a variety of ways, and the bound peptides are frequently in an extended conformation, although they might also adopt beta-turns and alpha-helices as motifs for recognition.^{8,9} Peptide-specific antibodies often display moderate affinities (μM to nM range¹⁰), and the peptides can be completely buried in cavities, bound in pockets or grooves, or form beta-strand type interactions at the protein surface.^{8,9} While antibodies, and in particular peptide-binding receptors, for example, hormone receptors, display high specificity, often binding a single peptide, the major histocompatibility complex (MHC) Class I and II molecules have solved the issue of diversity by generating a smaller number of different MHC molecules, each capable of binding a wide range of peptides.¹¹ The degenerate peptide-binding specificity of a MHC molecule is accomplished by invariant (main stabilizing) and variable (broad specificity) contacts, and the latter is generated through interactions with conserved amino acids at 2–3 positions, so called anchor residues, in the 8–10 (MHC Class I) or 13–17 (MHC Class II) amino acids long peptides, accommodated by the MHCs.¹² In fact, even the TCRs have been shown to exhibit promiscuous binding properties to non-overlapping peptides.^{13–16} It has been argued that the individual TCR, as well as antibody combining sites, is composed of a packet of specificities randomly plucked from the repertoire, hence being poly-specific, which in turn should be dissected into two features, specificity and degeneracy.¹⁷ Notably, recent studies based on X-ray crystallography have supported the possibility of peptide binding promiscuity, or poly-specificity (cross-reactivity) also for antibodies.^{18–20} Still, our understanding of the fundamental fine specificity of antibodies targeting peptides on a larger scale is yet surprisingly limited. This information will be essential when designing and generating peptide-specific antibodies for any application.

In this context, we have recently presented a novel discovery platform for quantitative protein expression profiling of complex proteomes, denoted global proteome survey (GPS), by combining the best features of affinity proteomics and mass spectrometry.^{21–23} In this set-up, we analyze proteomes generated by tryptic digestion using human recombinant single-chain Fv (scFv) antibodies directed against short binding motifs. Each motif, 4–6 amino acids long, is shared among several proteins (up to several hundred), enabling each antibody to target numer-

ous proteins in a specific independent manner. To this end, this novel breed of peptide specific antibodies has been denoted context independent motif specific (CIMS) antibodies. Initial data analysis showed that the CIMS antibodies displayed a complex, degenerate peptide reactivity pattern.^{21,23} Clearly, additional experiments will be warranted in order to unravel the antibody–peptide interaction (specificity) on a molecular level.

Here, we report the first detailed analysis of the CIMS antibody–peptide interaction (specificity) features, by combining large-scale experimentally determined peptide-binding data with structural data analysis for eight CIMS antibodies and numerous peptides. The results showed that the antibodies exhibited promiscuous peptide-binding properties and that mainly two or three semi-conserved motif residues, or anchor residues, were sufficient for establishing a specific antibody–peptide interaction, analogues to the MHC–peptide interaction. This knowledge could pave the way for exploring and exploiting protein–peptide interactions in future biomedical applications.

Results

Specific interaction antibody—Selection peptide

The peptide-binding specificity of eight CIMS-antibodies (Table I) was first investigated by evaluating their reactivity toward a synthetic peptide mixture consisting of 10 different peptide selection binding motifs, 4 or 6 amino acids long (plus a conserved 4 amino acids long linker). The results showed that the CIMS-antibodies, individually immobilized to beads, specifically bound their selection peptide only (Supporting Information Fig. 1).²¹ Further, the dissociation constant (K_D) was found to be in the range of 0.2–6.8 μM (Table I). Hence, the CIMS-antibodies displayed specificities and sensitivities as could be expected for anti-peptide antibodies directly selected from a scFv phage-display library and not subjected to affinity maturation.

Degenerate peptide-binding specificity

Next, we determined the detailed peptide-binding specificity of the CIMS-antibodies in a large-scale approach by targeting tryptic digests of human colon tissue extracts (Supporting Information Table 1)²¹ and yeast cell lysates (Supporting Information Table 2)²¹ based on the GPS-assay (Table I and Fig. 1). The set of tryptic peptides bound by each individual CIMS antibody was thus detected and identified using tandem-mass spectrometry. The results showed that peptides of a median length of 10 amino acids (range 6–22) and with a median pI of 6.5 (range 3.8–12) were captured in a clone-dependent manner (Table I). Noteworthy, as illustrated for the CIMS-17 sister pair, the affinity (K_D) for an

Table I. CIMS-Binder Characteristics

Selection peptide (binding motif in bold)	CIMS-antibody	Affinity ^a (K_D) (μ M)	Captured peptides	
			Length ^b	pI ^b
M-1	Biotin-SGSGEDFR	CIMS-1-A05 ^c	12 (7–21)	4.5 (3.8–12)
		CIMS-1-B03 ^c	10 (6–17)	6.3 (3.9–12)
M-17	Biotin-SGSGSSAYSR	CIMS-17-C08 ^c	9 (6–19)	6.8 (3.8–12)
		CIMS-17-E02 ^c	10 (6–19)	6.8 (4.0–12)
M-33	Biotin-SGSGLSADHR	CIMS-33-3C-A09 ^c	10 (7–15)	8.0 (4.1–10)
		CIMS-33-3D-F06 ^c	10 (6–21)	5.8 (3.8–12)
M-15	Biotin-SGSGLTFAK	CIMS-15-A06	9 (6–22)	6.0 (3.9–10)
M-34	Biotin-SGSGSEHLR	CIMS-34-3A-D10	11 (6–17)	7.9 (3.9–12)

^a Unless otherwise stated, determined against selection peptide.

^b Median (range).

^c Antibodies directed against the same selection peptide, but displaying different amino acid sequences in the antigen-binding sites were defined as sister clones.

^d Data adopted from Ref. 21.

^e Determined against a captured yeast peptide (FAGGNAYQR).

experimentally verified captured yeast peptide was found to be in the same range as that observed for the original selection peptide (Table I).

When the sequences of the experimentally verified captured peptides were compiled and expressed as WebLogos, a degenerate and clone-dependent peptide-binding pattern was revealed for all CIMS antibodies [Fig. 1(A–E)]. The results showed that the last four C-terminal residues (denoted P1–P4) of the peptides were essential for establishing specificity, whether the original selection motif had been four or six amino acids long. More precisely, mainly two or three semi-conserved positions were found to be critical, acting as potential anchor-residues, while a significantly higher degree of wobbling was allowed in the other positions [Fig. 1(F)]. This feature, combined with the fact that most captured peptides were fairly short, further indicated the critical nature of these interactions for binding. Using trypsin as digestive enzymes, the peptides will be limited toward C-terminal lysine or arginine in P1. However, the data showed that the frequency of these two residues in P1 was about 50–50% for three CIMS antibodies (33-3C-A09, 1-B03, and 34-3A-D10), but almost 100% lysine or 100% arginine in the other five CIMS antibodies (Fig. 1). Thus, the data indicated the importance of also a conserved P1 residue for binding for five of eight CIMS antibodies. Like MHC bound peptides, a tendency toward utilizing different combinations of charged and/or hydrophobic residues at the candidate anchor positions was observed. Hence, the antibody–peptide interaction was in general found to display key features very similar to that of the MHC–peptide complex formation.

Antibody structure reflects the overall peptide-binding specificity

The CIMS antibodies were found to bind different sets of peptides (Fig. 1 and Supporting Information

Tables 1 and 2),²¹ and the cross-talk was low. Further, the CIMS-33 and CIMS-1 sister pairs were also found to capture different set of peptides, although they had been selected against the same selection peptide binding motif. Interestingly, the CIMS-17 sister clones were each found to enrich a unique set of peptides, as well as an overlapping set of peptides bound by both clones. In order to gain further insights into the antibody–peptide interaction, primary and tertiary (homology modeling) structure analysis of the CIMS-scFv antibodies, all based on the same framework, was performed (Fig. 2 and Supporting Information Fig. 2). The results showed that the main structural differences were located to complementarity-determining region (CDR)-H2, –H3, and –L3, respectively (Fig. 2). Noteworthy, the same structural differences, albeit more focused toward CDR-H3, were observed between the individual sister clones [Fig. 2(B–D)]. Hence, the antibody structures could indirectly explain the different overall peptide-binding specificity displayed by the CIMS-antibodies, but not the degenerated binding pattern displayed by each individual clone.

Peptide structure reflects the degenerate peptide-binding specificity

In attempt to address the observed antibody–peptide interactions from the ligand point of view, we modeled the three-dimensional structure of in total 120 peptides, including 5 selection motifs, 100 experimentally verified captured peptides, and 15 peptides defined as general, non-specific background binders (Supporting Information Table 3). The peptides were modeled using PEP-FOLD, and a majority of the peptides were modeled into one or two main clusters with adequate statistics, from which the top ranked structure was selected (Supporting Information Fig. 3).

The results showed that all peptides interacting with an individual antibody, but not the

Figure 1. Experimentally determined peptide binding motifs for eight CIMS-antibodies. The last six residues, denoted P1–P6 starting from the C-terminal, are displayed, and the frequency of each individual residue is shown. The original peptide selection motif and the hydrophilic index for each position based on the Hoop–Woods scale (most hydrophobic amino acids get negative values) are displayed. A: Sister clones 17-C08 and 17-E02. B: Sister clones 33-3C-A09 and 33-3D-F06. C: Sister clones 1-A05 and 1-B03. D: Clone 15-A06. E: Clone 34-3A-D10. F: The candidate key semi-conserved anchor residue positions are indicated.

background-binding peptides, overlapped well in the last three (P1–P3) or four (P1–P4) C-terminal positions, while no structural consensus was readily observed for the middle or N-terminal parts of the peptides (Figs. 3 and 4, and Supporting Information Figs. 4–6). Using the CIMS-15 captured peptides as a general showcase, the data demonstrated that the peptides adopted an overlapping alpha-helices like structure in P1–P4 [Fig. 3(A)]. Most importantly, the side chains of P1 and P4 were found to point in the same direction, while those of P2 and P3 pointed elsewhere [Fig. 3(B,C)]. Hence, the folding of the

peptides revealed how P1 and P4 potentially could act as semi-conserved anchor-residues in the antibody–peptide interaction, while a higher degree of wobble would be tolerated in positions P2 and P3 [cf., Figs. 1(D) and 3]. This was further highlighted by the fact that a majority of the background binding peptides overlapped poorly with the selection peptide and adopted different structures [Fig. 3(D)]. Of note, two of the peptides defined as a background binding peptide was found to adopt a primary and tertiary structure similar to that of the refined binding motif [Fig. 3(E)], why it might be argued that

Figure 2. Structural analysis of CIMS-antibodies. Three-dimensional structural homology models were generated for three pairs of sister clones. A: Overlapped ribbon structures (top view) with color coded CDRs for all six CIMS-antibodies. B: Overlapped ribbon structures (top view) with color coded CDRs for sister clones 17-C08 and 17-E02, and separate space filling models onto which the CDRs and electrostatics have been mapped. C: Overlapped ribbon structures (top view) with color coded CDRs for sister clones 33-3C-A09 and 33-3D-F06, and separate space filling models onto which the CDRs and electrostatics have been mapped. D: Overlapped ribbon structures (top view) with color coded CDRs for sister clones 1-A05 and 1-B03, and separate space filling models onto which the CDRs and electrostatics have been mapped. E: Amino acid sequences (color coded) of the CDRs for all six CIMS antibodies.

this peptide might also be specifically captured by this particular CIMS-antibody. Taken together, the folding of the peptides appeared to reflect the degenerate peptide-binding specificity displayed by the CIMS antibodies and that a few semi-conserved anchor residues were sufficient for establishing specificity.

The structural overlap on the entire motif (P1–P4) between peptides captured by different sister clones was best for the CIMS-17 pair (Fig. 4 and Supporting Information Figs. 5 and 6), which also was the only sister pair with significant cross-talk, that is, each of the clones were found to bind a unique as well as a shared cohort of peptides. While P1 and P3 might explain the cross-talk, as well as the degenerate peptide-binding specificity, P3 might also confer the clone-dependent binding feature (cf., Figs. 1 and 4).

Antibody–peptide complexes reflect specific and degenerated binding-specificity

Next, we modeled the antibody–peptide complexes formed by three CIMS antibodies with two or four different peptides each, and only those complexes that showed coherent results, that is, when one antibody was found to bind the different peptides in the same position and manner was further evaluated (Fig. 5 and Supporting Information Fig. 7). In the case of CIMS-17-E02, the antibody was found to bind the four different peptides in a similar fashion at the same position (Fig. 5). Noteworthy, the peptides interacted with mainly CDR-H3 and –L3, previously pin-pointed as essential in defining the specificity of this particular CIMS-antibody (cf., Figs. 2 and 5). Furthermore, the modeled antibody–peptide complex indicated P1 and P3 as main anchor residues, in line with the sequence analysis of the

Figure 3. Structural analysis of peptides captured by CIMS-antibody 15-A06. Using PEP-FOLD, structural models were generated for the selection peptide, 10 experimentally captured peptides, and 15 peptides experimentally defined as general non-specific background binding peptides. A: Selection peptide and captured peptides overlapped on P1–P4 (cyan). B: Truncated (P1–P4 shown) selection peptide and captured peptides overlapped on P1–P4. C: As (B), but showing all sides-chains of P1–P4. D: Selection peptide (colored in pink and cyan (P1–P4) and all background-binding peptides (red) overlapped on P1–P4. E: Selection peptide (colored in pink and cyan (position P1–P4), and two background binding peptides (red) overlapped on P1–P4.

experimentally verified captured peptides and the modeling data of the peptides (cf., Figs. 1, 4, and 5). Interestingly, the bound peptides were also found to

overlap well (P1–P3) with the conformation of the modeled selection peptide [Fig. 5(E)]. In this context, it should be noted that similar observations were

Figure 4. Structural analysis of peptides captured by the CIMS-antibody sister clones 17-C08 and 17-E02. Using PEP-FOLD, structural models were generated for the selection peptide, 10 experimentally verified peptides captured by 17-C08, 10 experimentally verified peptides captured by 17-E02, and 10 experimentally verified peptides captured by both 17-C08 and 17-E02. A: Selection peptide and peptides captured by both 17-C08 and 17-E02 overlapped on P1–P4 (cyan), and a close-up of P1–P5 with P2 colored in red. B: Selection peptide and peptides captured by 17-C08 overlapped on P1–P4 (cyan), and a close-up of P1–P5 with P2 colored in red. C: Selection peptide and peptides captured by 17-E02 overlapped on P1–P4 (cyan), and a close-up of P1–P5 with P2 colored in red.

Figure 5. Structural analysis of peptide—CIMS—antibody complexes. Three dimensional structural models were generated for CIMS antibody 17-E02 and four experimentally verified captured peptides. A: Overlapped ribbon structures (top view) with color coded CDRs and peptides (orange). B: A close-up of (A) and the side-chains of position P1–P3 are displayed. C: A space-filling model of the antibody with color coded CDRs with one of the bound peptides. D: As (C), but now showing all four bound peptides. E: Overlap (P1–P3) of the selection peptide (red) modeled free in solution, with the four bound peptides (blue, cyan, green, and yellow).

made for the complex formed between CIMS-17-C08 and two different peptides (cf., Figs. 1, 4, and Supporting Information Fig. 7), thus further corroborating the findings. In comparison to its sister clone 17-E02, the peptide-binding site on 17-C08 appeared to be somewhat displaced, involving also CDR-H2, further illustrating the clone-dependent differences.

Design of next generation binding motifs

Based on the experimental observations above, we then implemented the concept of semi-conserved anchor residues in the design of shorter peptide binding selection motifs, in order to evaluate the next generation of candidate selection motifs for broad protein coverage. To this end, we evaluated the impact of defining P1–P4 (Fig. 6), P1 and P3 [Supporting Information Fig. 8(A,B)], as well as P1 and P2 [Supporting Information Fig. 8(C)] as anchor residues, respectively. Using the P4-x-x-P1 motif as a showcase showed that 718–20,005 tryptic peptides per potential motif was found (Fig. 6). Selecting a pair of the most frequent motifs, ExxK/ExxR, and filtering for single (unique) peptide sequences reduced the number of potential peptides to 17,724/16,385, corresponding to 8843/8787 protein accessions, clearly demonstrating a broad potential coverage. Adding an additional filter, to exclude

high-abundant protein species that could impair the subsequent mass spectrometry (MS) step (Supporting Information Table 4), 11,730/13,161 single peptides were still defined, representing 6,925/7,701 protein accessions. In a similar fashion, P3-x-P1 and P2-P1 motifs could be defined, potentially generating a coverage of 4401 to 9020 single peptides, corresponding to 3416 to 5902 protein accessions (Supporting Information Fig. 8). As illustrated for a few selected motifs, the proteins carrying the motifs were well distributed between cytoplasm, nucleus, plasma membrane, and the extracellular space, further highlighting the broad coverage (Supporting Information Fig. 9). In the end, the precise coverage will also be dependent on a combination of additional key factors, such as the properties of the individual CIMS antibody, the allowed wobbling of the non-anchor residues, and how well the peptide can be targeted by MS and so forth, which most likely will result in a lower coverage that in the end will have to be experimentally determined and verified.

Discussion

Protein–peptide interactions are fundamental and essential for numerous cellular processes and far more common than was initially estimated. Recently, they have been proposed to account for at least

Figure 6. Evaluation of the frequency of candidate C-terminal motifs in a human tryptic proteome. A motif with two anchor residues (P4-x-x-P1) was targeted, with wobbling allowed in position P2 and P3. A: P4 was set to K, and detailed statistics are shown for the motif ExxK, including the impact of filtering the motifs for presence in high-abundant proteins. B: P4 was set to R, and detailed statistics are shown for ExxR, including the impact of filtering the motif for presence in high-abundant proteins. Total hits represent the number of tryptic peptides in the database carrying the motif(s). Single peptide sequences represent the number of unique peptides carrying the motif(s). The motifs that were filtered for presence in high abundant proteins are plotted (in gray) at the right hand side of the graph.

15–40% of all interactions within a cell.² The importance of such interactions are evident from the high number of domains, for example kinase catalytic and WD40 domains,²⁴ now known to interact with peptides sharing a common sequence pattern or linear motifs. However, nature has also evolved protein species capable of displaying multiplicity, or promiscuity in ligand selection, that is a protein (or domain) able of binding multiple peptide sequences. Promiscuous binding properties are typified by the Src homology 3 (SH3) domain and many other interaction modules that commonly recognize and bind

proline-rich sequences,²⁵ giving us a glimpse of the complex nature of protein–peptide interactions. In this particular case, the shallow binding surface of SH3, although ideal for accommodating this type of ligands, lacks the intrinsic features that would allow it to distinguish minor differences between two proline-rich sequence.²⁶ In comparison, MHC Class I and II molecules possess an intrinsic ability to bind a diverse set of ligands in a peptide binding groove, utilizing a few conserved anchor residues within the accommodated peptide to display a degenerate peptide-binding specificity.¹¹

In contrast, the hallmark of antibodies is to specifically recognize and bind its antigen, for example a peptide, and they are routinely referred to as one-ligand binders. This feature can be corroborated on a structural level, by the fact that the structure of the antigen-binding site represents a dynamic surface being rich in pockets and clefts, and that is different from one antibody to another.^{8,9} However, in this study, we presented evidence for degenerated peptide-binding specificity among recombinant antibodies, instead displaying features mimicking the MHC molecules. These findings were supported by recent structural work that revealed that monoclonal antibodies could display peptide-binding promiscuity or poly-specificity,^{18–20} a feature which should be further dissected into specificity and degeneracy.¹⁷ In these studies, one monoclonal antibody was found to bind a set of non-homologous peptides, but the peptides adopted different binding conformations and formed their critical interaction with different antibody residues.^{19,20} This might indicate that structurally different (peptide) binding motifs could be obtained for any monoclonal antibody,¹⁹ which if established, would impact, for example the way we search for auto-antigens and design peptide motifs for immunization. In the other example, a set of peptides were observed to adopt spatially different footprints within the antigen-combining site of the antibody, but they shared a conformation-specific lock involving two residues on the antibody.¹⁸ The authors interpreted the multiple binding modes of diverse epitopes within a common paratope conformation of a germline antibody as a mechanism for expanding the primary antibody repertoire. Hence, the observed poly-specificities of these antibodies coincided with the argument that an individual antibody combining site is composed of a package of specificities randomly plucked from the repertoire.¹⁷ Still, our fundamental knowledge about the antibody–peptide interactions in terms of fine specificity and potential binding-promiscuity is still very limited.

In this study, we showed novel examples of degenerate antibody peptide-binding specificity, involving a molecular mechanism more similar to the one adopted by MHCs. To the best of our knowledge, this phenomenon has not previously been described for antibodies. In analogue to MHC-peptide, our data implied that mainly 2–3 semi-conserved anchor residues were sufficient for establishing specificity, while a significantly higher degree of wobbling was allowed in the other positions, enabling the degenerate binding of multiple motif-carrying peptides. However, the MHC peptides are bound in mainly an elongated conformation,¹¹ while the modeling experiments indicated that the motif-containing end of our peptides adopted a more α -helices-like conformation. Although it cannot be

excluded that this was a bias, at least partly, introduced by the peptide modeling tool, it should be noted that motif-end of the peptide containing the anchor-residues adopted a similar conformation when the antibody–peptide complexes were modeled using a different modeling tool.

In comparison, the combining site of the TCR, which is built up in a similar fashion to that of an antibody, has been suggested to exhibit a different molecular mechanism for promiscuous binding, involving flexibility (plasticity).^{15,16} The available crystal structures of liganded and unliganded TCRs have implied that the flexibility of the TCR combining site facilitates TCR adaptation to rather conformationally constrained MHC–peptide surfaces.^{27–29}

Based on the SH3 work,^{25,26} the importance of the affinity for peptide binding promiscuity has been highlighted, and such interactions tend, as might be expected, to be of low affinity. In accordance, our peptide–antibodies displayed affinities in the μ M range, which for antibodies means low-affinity interactions. Additional experiments will be required in order to elucidate whether the feature of degenerate peptide-binding specificity could be minimized or even eliminated by affinity maturing the scFv antibodies. In the case of a promiscuous TCR, recognizing non-overlapping peptides, the promiscuity, that is the degree of degeneracy, was indeed found to correlate with increased affinity/avidity.¹³ In this context, it should be noted that the conditions for selecting antibodies from recombinant phage-display libraries could be manipulated in order to promote the selection of more high- or low-affinity binders. In addition, the sequence of the peptide can also be optimized in order to generate more high-affinity binders, as recently illustrated for MHC peptides.³⁰ Hence, it might be important to synchronize the selection procedures with the type of binding-specificity, that is degenerate vs. non-degenerate, that the retrieved clones are desired to display.

Examining the nature of the antibody bound peptides motifs, and in particular the semi-conserved anchor residues, revealed a tendency of utilizing different combinations of hydrophobic and/or charged residues in these position, in analogue to what has been observed for peptides bound by MHCs.¹² The location of the anchor residues differed, however. While the MHC bound peptide anchor residues are not located at the N- and C-terminal, the antibody bound peptides anchor residues were located among the last four C-terminal residues, including the C-terminal. However, in our case, this was a deliberate motif strategy based on the fact that the GPS set-up was designed to target tryptic peptides.^{21–23} It could be anticipated that the same phenomena would be observed using proteases favoring different cutting sites, although this remains to be experimentally verified. The impact of

adding extra residues to the C-terminal end of the peptides also remains to be directly evaluated. In this context it should be noted that none of the bound peptides contained an internal missed, that is non-cleaved trypsin site, but some of the CIMS antibodies have been shown to bind intact, semi-denatured full length proteins.²¹ Hence, the data again indicated on a clone dependent importance of the C-terminal, that is position P1, for binding. No N-terminal motifs, or reactivity against N-terminal was observed for the selected C-terminal specific CIMS antibodies. However, adopting the reverse approach, there are no obvious reasons to why not a N-terminal motif-based system could be established. Depending on the identity of the semi-conserved residues at P1–P4, and the properties of the surrounding surface provided by the antibody, the range of normal type of ligand–antibody interactions (e.g., hydrophobic interactions, van der Waals interactions) would be anticipated, as supported by the peptide–antibody modeling work. Hot spot residues, important for protein–peptide as well as protein–protein interactions, has often been found to involve residues, such as arginine, tryptophan, glutamic acid, and tyrosine.³¹ In this context, it was of interest to note that a tryptophan in CDR-L3 of CIMS antibody 17-E02 was in close proximity to one of the anchor residues of a bound (modeled) peptide and available for interaction (data not shown). Additional work, involving 3D-structural determinations, will be required in order to validate and further refine these findings.

In this study, we used human recombinant scFv antibodies selected from a large phage display library³² as binding probes. In order to make sure that the phenomenon of degenerate peptide-binding specificity was biologically relevant and not a bias introduced by the library design, we examined the reactivity pattern of two conventional antibodies raised in rabbits against similar tryptic peptide motifs.³³ Of note, these antibodies are used within a set-up similar to our GPS technology platform,^{21–23} denoted Triple X proteomics (TXP).³³ Most interestingly, we found three semi-conserved residues (P4-x-P2-P1 or P4-P3-x-P1) to be essential for establishing the peptide-binding specificity also for these antibodies, while a higher degree of wobbling was allowed in the other positions (Supporting Information Fig. 10). Hence, these findings further corroborated our results of promiscuous peptide-binding specificity among antibodies, and implied that it could be a more general biological feature for at least low-affinity peptide-specific antibodies.

In the process of generating antibodies, the aspect of promiscuous peptide-binding will also impact the way the target peptide (antigen) should be designed. In this study, we used continuous linear motifs, four or six amino acids long, as antigens, and

they were selected as target motifs based on the frequency that the entire motif occurred in the human proteome.^{21–23} In comparison, a similar motif strategy was adopted by Planatscher *et al.*, and using bioinformatics, they showed that only ~2000 solutions (motifs) would be required in order to cover the entire human proteome, and to provide multi-coverage for around 13,800 proteins.³⁴ However, based on the current study, it has now become evident that the semi-conserved anchor residues are a key characteristic to consider in this design process. Adopting such a strategy, we here outlined a novel path forward in designing candidate motifs for broad protein coverage, although the motifs in the end will have to be experimentally verified. In fact, this molecular mechanism might in the end facilitate the design of motifs for broad protein coverage. Considering the fact that some captured peptide groups aligned well and adopted, for example, a longer alpha-helices like structure, one might speculate that additional up-stream anchor-residue(s) could also be defined (data not shown). If instead, a certain group of proteins or protein families are to be targeted, the motifs should be designed adopting a different bioinformatic strategy. In this case, the protein targets should first be defined and any shared sequences pin-pointed, and it is then within these common sequences any potential peptide selection binding motif(s) should be defined.

In conclusion, we have shown that promiscuous peptide-binding specificity was displayed by anti-peptides antibodies and that this could be a general phenomenon. The molecular mechanism for the degenerate peptide-binding specificity appeared to be accomplished by the use of 2–3 semi-conserved anchor residues in the peptides, in analogue to the mechanism evolved for MHC molecules. In a longer perspective, this information will foster our understanding of antibody specificity and will be essential for designing, generating, and applying peptide specific antibodies, or peptide-binding proteins in general, in applications within the fields of proteomics, biology, and medicine.

Material and Methods

CIMS antibodies

Eight human recombinant CIMS scFv antibody fragments directed against five short C-terminal amino acid peptide motifs were selected from the n-CoDeR library,³² and kindly provided by Bioinvent International AB, Sweden (Table I). CIMS antibodies selected against the same selection peptide, but displaying different amino acid sequences in their antigen-binding regions, were denoted sister clones.

The CIMS antibodies were produced in *Escherichia coli* cultures, purified using affinity chromatography on Ni²⁺-NTA agarose (Qiagen, Hilden,

Germany), and quality controlled (concentration, integrity, and purity) as recently described.²¹

Purified CIMS antibodies (180–250 µg) were individually coupled to ~9 mg (300 µL) of magnetic beads (M-270 carboxylic acid-activated, Invitrogen Dynal, Oslo, Norway), using manufacture provided protocols with some minor modifications as described before.²¹

Analysis of synthetic peptide mixture

The peptide binding specificity of the CIMS antibodies was evaluated on a synthetic peptide mixture in a similar manner as earlier described.²¹ Briefly, for each capture experiment, a synthetic peptide mixture consisting of ten different selection peptides was applied to a solution of CIMS-conjugated beads and a two-step washing procedure was then performed by adding phosphate buffered saline (PBS). Subsequently, the captured peptides were eluted by adding 5% (vol/vol) acetic acid. Prior to mass spectrometry analysis, a sample clean-up was performed using C18 tips (Thermo Scientific (formerly Proxeon A/S), Odense, Denmark). Mass spectrometry was performed using a matrix-assisted laser desorption/ionization linear trap quadrupole (MALDI-LTQ) (Thermo) instrument. Standard 96-well stainless steel MALDI target plate (Thermo) was used, where 1 µL sample was spotted onto the target plate, dried followed by another 1 µL sample together with 1 µL matrix (2.5 mg/mL CHCA (α -cyano-4-hydroxycinnamic acid) in 50% (vol/vol) acetonitrile, 0.05% (vol/vol) trifluoroacetic acid).

GPS capture of tryptic proteome digests

Captures of tryptic proteomes, including human colon tissue and yeast cell lysate (*Saccharomyces cerevisiae*), were performed as previously described.²¹ Briefly, two replicate captures were performed for each sample (yeast and colon on the same digested sample batches) and the elute was then directly used for mass spectrometry analysis without any additional cleanup. An ESI-LTQ-Orbitrap XL (Thermo Electron, Bremen, Germany) coupled to an Eksigent 2D nano HPLC (Eksigent technologies, Dublin, CA) was used. All raw-data files generated were processed and converted into mzData format, using Proteome Discovery software (v 1.0) and then uploaded in Proteios SE.³⁵ The Proteios platform was then used for performing automated database searches in both Mascot and X! Tandem as previously described.²¹ The human IPI protein database version 3.71 (consisting of forward and random sequences with the same amino acid distribution and size resulting in a total of 173490 sequences for the combined database) was used for the colon sample. The *S. cerevisiae* Swiss-Prot (13-Oct 2009 consisting of forward and random sequences with the same amino acid distribution and size resulting in a

total of 13,780 sequences) was used for the yeast samples. In both cases, the following parameters were applied: enzyme—trypsin; missed cleavages—1; fixed modification—carbamidomethyl (C); variable modification—methionine oxidation. A peptide mass tolerance of 3 ppm and fragment mass tolerance of 0.5 Da was used. This was followed by automated combination of the search results from the two different search engines with predetermined false discovery rates (FDR). A FDR of 0.01 was used for both the human colon set and the yeast set. In order to extend the peptide binding data, the data set was merged with old mass spectrometry data,²¹ recompiled, and re-analyzed. The tool WebLogo was used to generate the motif figures for selected binders enriched sequence motifs.³⁶

Affinity measurements of CIMS-binders

As recently described,²¹ a competitive inhibition affinity assay was used to determine the affinity of the CIMS antibodies for the corresponding fluorescently labeled (i) synthetic selection peptides, or (ii) a synthetic yeast peptide (FAGGNAYR) (clones 17-C08 and 17-E02). The latter peptide was selected to represent a “real” peptide, experimentally verified to be captured in GPS experiments of a tryptic yeast proteome. The amount of bound versus free peptide were calculated and Scatchard plots made in order to determine the dissociation constant (K_D). Synthetic peptides (80% degree of purity), with or without a fluorescent label (TAMRA) were purchased from Thermo Biopolymers, Germany.

Structural analysis of CIMS antibodies

Structural homology models were generated for all CIMS antibodies using Web Antibody Modelling server following the standard guidelines (<http://antibody.bath.ac.uk>). In more detail, the antibodies were aligned using the manual method, but identical results were obtained if the auto-align approach was used instead. The screening and side-chain building methods were selected based on the properties of the CDRs according to the manual. The structures were examined and visualized, using Pymol (Delano Scientific LCC).

Structural analysis of peptides

Structural models were generated for in total 120 peptides, including all selections peptides as well as a set of peptides, experimentally verified to be captured by the CIMS antibodies in this and a previous study,²¹ and a cohort peptides defined as general, non-specific background binders. The specific peptides were selected based on (i) having a high MS identification score, and (ii) the WebLogos in the sense that they should contain the refined binding motif(s), but still displaying large sequence differences upstream of the refined motif including the

N-terminal. The peptides were modeled using mainly PEP-FOLD,³⁷ but also PROFASI.³⁸ The peptide–antibody complexes were modeled using PROFASI, and to ensure that we did not introduce a bias in the peptide modeling, we performed dual modeling of the peptides. Notably, when comparing the peptide modeling data from PEP-FOLD and PROFASI similar results were obtained (data not shown). In the case of the eight amino acid long peptide, M1, an extra small, neutral residue (glycine) was added at the N-terminal end, that is as far as possible from the C-terminal, in order to meet the length requirement for modeling in PEP-FOLD (≥ 9 amino acids). The structures were visualized using Pymol (Delano Scientific LLC) and/or Swiss-Pdb Viewer.³⁹

PEP-FOLD is a *de novo* approach, aimed at predicting the free peptide structures from amino acid sequences. This method, based on structural alphabet SA letters to describe the conformations of four consecutive residues, couples the predicted series of SA letters to a greedy algorithm and a coarse-grained force field. In each PEP-FOLD run, 100 simulations were made, returning an archive of all the models generated, and the best conformations of the best five clusters. The clusters are based on root-mean-square-deviation (RMSD) between the models, and from this cluster, the model with the lowest energy was selected. In the case of PROFASI, each of the peptide sequences were studied as free chains by a single Monte Carlo simulation with 10 billion elementary steps at the temperature $T = 298$ K. These single runs were long enough to provide full representative sampling of the equilibrium behavior. In order to characterize the most populated structural states for each peptide, we used so-called Complete Linkage Hierarchical clustering.⁴⁰ To determine the similarity between peptide conformations, we used the standard RMSD measure. The RMSD cutoff value used was 3.3 Å. Clusters were ranked according to their size, representing their populations, and not based on energy, since we reasoned that structural clustering would be a more robust approach. Peptides/ligands do not often adopt one fixed structure, not even in the bound state, but rather exist in ensembles of different conformations.⁴¹ The clustering aims to characterize the most typical conformations and we then selected clusters based on size (largest population).

Structural analysis of peptides–CIMS–antibody complexes

The PROFASI program package was used for all the peptide–antibody modeling. First, the full structures of the CIMS antibodies were built, using WAM. Thereafter, the structures were reduced in size to fragments of around 93–102 amino acids, including all of the six CDRs in full. To maintain the native state of the antibody fragment while allowing some

chain flexibility during simulations, an auxiliary energy term proportional to RMSD was added to the total potential energy function of the model. Initial simulations of the antibody fragments were performed to relax the CIMS antibody structures under these constraints. Thereafter, a peptide chain was added and allowed to explore a cylinder-shaped volume surrounding the top of the antigen-binding site, that is, the CDR regions, using an additional script. The precise size and location of the cylinder was determined manually, with the goal of allowing the peptides to freely explore the top of the binding-site, while disallowing (non-physical) binding on the side or the bottom of the CIMS antibody fragments. For each antibody–peptide pair, five independent simulations were performed at $T = 298$ K with 2 billion MC steps. To characterize typical antibody–peptide bound structures, the same basic clustering procedure was performed as for the free PROFASI modeled peptides with the difference that only the five most C terminal amino acids of the peptides were taken into account in the procedure, that is covering the experimentally determined binding site on the peptides. For the clustering of antibody–peptide conformations, we used a cut-off corresponding to 2.0 Å.

Due to computational limitations, only three CIMS antibodies (based on the fact that their selection peptide modeled the best, see Supporting Information Fig. 3) and two experimentally verified captured peptides each were selected for modeling. The peptides were selected based on (i) giving an adequate MS identification score, and (ii) the WebLogos in the sense that they contained the refined binding motif(s). To this end, the following complexes were modeled, including CIMS-15-A06 and IGSAIEYQK/TVITNEYMK, CIMS-17-C08 and EDLYQTFAR/HDIFLQFAR, and CIMS-17-E02 and AIPQLQGYLE/FHTITTSYYR. In addition, two peptides experimentally observed to be bound by both 17-C08 and 17-E02 (sister clones) were also modeled (AAGLTAAYAR/FAGGNAYQR) for complex formation.

Data analysis of candidate tryptic peptide motifs

The *in silico* trypsinization of the human proteome (release 56.1 of UniProtKB/Swiss-Prot (composed of 20,325 non redundant human proteins)) was generated, using SignPept, an in-house developed software (Dr. F. Levander, Department. of Immunotechnology). The theoretical cleavage specificities of trypsin have been built into the SignPept software, and mimics the rules employed by PeptideCutter (<http://expasy.org/tools/peptidecutter>). A complete digest without missed cleavages or miss-cleavages was assumed. The frequency of all motifs as well as refined (targeted) binding motifs was then evaluated using in house developed scripts (Dr. A. Carlsson, Department. of Immunotechnology). A peptide

molecular weight cut-off of 800–2000 Da and/or abundance level cut-offs were used.

Acknowledgments

Authors gratefully acknowledge Dr. Karin Hansson for assistance with the LTQ-Orbitrap instrument, Dr. Fredrik Levander, supported by BioInformatics Infrastructure for Life Science (BILS), for assistance with Proteios SE and the SignPept software, and Dr. Anders Carlsson for assistance with in-house designed software for identifying candidate CIMS binding motifs. CW and CB are co-founder of a company that holds a granted patent for the basic GPS technology.

References

- Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol* 3:e405.
- Petsalaki E, Russell RB (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr Opin Biotechnol* 19:344–350.
- Hancock RE, Nijnik A, Philpott DJ (2012) Modulating immunity as a therapy for bacterial infections. *Nat Rev Microbiol* 10:243–254.
- Hanash SM, Pitteri SJ, Faca VM (2008) Mining the plasma proteome for cancer biomarkers. *Nature* 452:571–579.
- Tovar C, Rosinski J, Filipovic Z, Higgins B, Kolinsky K, Hilton H, Zhao X, Vu BT, Qing W, Packman K, Myklebost O, Heimbrook DC, Vassilev LT (2006) Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. *Proc Natl Acad Sci USA* 103:1888–1893.
- Karanicolas J, Kuhlman B (2009) Computational design of affinity and specificity at protein-protein interfaces. *Curr Opin Struct Biol* 19:458–463.
- Ussar S, Vienberg SG, Kahn CR (2011) Receptor antibodies as novel therapeutics for diabetes. *Sci Transl Med* 3:113ps47.
- Stanfield RL, Wilson IA (1995) Protein-peptide interactions. *Curr Opin Struct Biol* 5:103–113.
- Wilson IA, Stanfield RL (1994) Antibody-antigen interactions: new structures and new conformational changes. *Curr Opin Struct Biol* 4:857–867.
- Pope ME, Soste MV, Eyford BA, Anderson NL, Pearson TW (2009) Anti-peptide antibody screening: selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique. *J Immunol Methods* 341:86–96.
- Neefjes J, Jongasma ML, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11:823–836.
- Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50:213–219.
- Li L, Wang B, Frelinger JA, Tisch R (2008) T-cell promiscuity in autoimmune diabetes. *Diabetes* 57:2099–2106.
- Eisen HN (2001) Specificity and degeneracy in antigen recognition: yin and yang in the immune system. *Annu Rev Immunol* 19:1–21.
- Markovic-Plese S (2009) Degenerate T-cell receptor recognition, autoreactive cells, and the autoimmune response in multiple sclerosis. *Neuroscientist* 15:225–231.
- Mazza C, Auphan-Anezin N, Gregoire C, Guimezanes A, Kellenberger C, Roussel A, Kearney A, van der Merwe PA, Schmitt-Verhulst AM, Malissen B (2007) How much can a T-cell antigen receptor adapt to structurally distinct antigenic peptides? *EMBO J* 26:1972–83.
- Cohn M (2008) An in depth analysis of the concept of “polyspecificity” assumed to characterize TCR/BCR recognition. *Immunol Res* 40:128–147.
- Sethi DK, Agarwal A, Manivel V, Rao KV, Salunke DM (2006) Differential epitope positioning within the germ-line antibody paratope enhances promiscuity in the primary immune response. *Immunity* 24:429–438.
- Kramer A, Keitel T, Winkler K, Stocklein W, Hohne W, Schneider-Mergener J (1997) Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody. *Cell* 91:799–809.
- Keitel T, Kramer A, Wessner H, Scholz C, Schneider-Mergener J, Hohne W (1997) Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity. *Cell* 91:811–820.
- Olsson N, Wingren C, Mattsson M, James P, O’Connell D, Nilsson F, Cahill DJ, Borrebaeck CAK (2011) Proteomic analysis and discovery using affinity proteomics and mass spectrometry. *Mol Cell Proteomics* 10:M110003962.
- Wingren C, James P, Borrebaeck CAK (2009) Strategy for surveying the proteome using affinity proteomics and mass spectrometry. *Proteomics* 9:1511–1517.
- Olsson N, James P, Borrebaeck CAK, Wingren C (2012) Quantitative proteomics targeting classes of motif-containing peptides using immunoaffinity-based mass spectrometry. *Mol Cell Proteomics* 11:342–354.
- Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinski F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E (2005) Ensembl 2005. *Nucleic Acids Res* 33:D447–D453.
- Li SS (2005) Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem J* 390:641–653.
- Ladbury JE, Arold S (2000) Searching for specificity in SH domains. *Chem Biol* 7:R3–R8.
- Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, Teyton L, Wilson IA (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* 279:1166–1172.
- Reiser JB, Gregoire C, Darnault C, Mosser T, Guimezanes A, Schmitt-Verhulst AM, Fontecilla-Camps JC, Mazza G, Malissen B, Housset D (2002) A T cell receptor CDR3beta loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex. *Immunity* 16:345–354.
- Kjer-Nielsen L, Clements CS, Purcell AW, Brooks AG, Whisstock JC, Burrows SR, McCluskey J, Rossjohn J (2003) A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* 18:53–64.

30. Knapp B, Giczi V, Ribarics R, Schreiner W (2011) PeptX: using genetic algorithms to optimize peptides for MHC binding. *BMC Bioinformatics* 12:241.
31. Ben-Shimon A, Eisenstein M (2010) Computational mapping of anchoring spots on protein surfaces. *J Mol Biol* 402:259–277.
32. Soderlind E, Strandberg L, Jirholt P, Kobayashi N, Alexeiva V, Aberg AM, Nilsson A, Jansson B, Ohlin M, Wingren C, Danielsson L, Carlsson R, Borrebaeck CAK (2000) Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat Biotechnol* 18:852–856.
33. Hoeppe S, Schreiber TD, Planatscher H, Zell A, Templin MF, Stoll D, Joos TO, Poetz O (2011) Targeting peptide termini, a novel immunoaffinity approach to reduce complexity in mass spectrometric protein identification. *Mol Cell Proteomics* 10:M110 002857.
34. Planatscher H, Supper J, Poetz O, Stoll D, Joos T, Templin MF, Zell A (2010) Optimal selection of epitopes for TXP-immunoaffinity mass spectrometry. *Algorithms Mol Biol* 5:28.
35. Hakkinen J, Vincic G, Mansson O, Warell K, Levander F (2009) The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. *J Proteome Res* 8:3037–3043.
36. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190.
37. Maupetit J, Derreumaux P, Tuffery P (2009) PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res* 37:W498–W503.
38. Irback A, Mohanty S (2006) PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *J Comput Chem* 27:1548–1555.
39. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723.
40. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20:1453–1454.
41. Mobley DL, Dill KA (2009) Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure* 17:489–498.