A STRUCTURAL COMPARISON WORKFLOW FOR INVESTIGATING IGE CROSS-REACTIVITY BETWEEN COW'S MILK AND RESPIRATORY LIPOCALIN ALLERGENS

Ikebek S. Peter

A graduate thesis submitted to the Department of Chemistry of Lakehead University in partial fulfillment

of the requirements for the degree of

Master of Science in Chemistry

Monday, September 21, 2026

ABSTRACT

Background: Cow's milk allergy and allergic asthma can occur in the same individuals, but it is not well understood at the molecular level. Bovine β-lactoglobulin (Bos d 5) is a major cow's milk allergen and a member of the lipocalin protein family. Several clinically important respiratory allergens are lipocalins. Their shared fold provides a basis for comparing surfaces that may be relevant to IgE recognition.

Objective: To develop a structural workflow for comparing Bos d 5 IgE-binding regions with corresponding regions in milk, respiratory and human lipocalins.

Methods: Literature-reported linear IgE-binding regions of Bos d 5 and the conformational epitope defined from the Bos d 5 IgE–Fab complex were mapped onto homologous lipocalin structures using structural alignment. In parallel, comparative Fab–query allergen models were generated by positioning each query structure in the binding orientation defined by the Bos d 5 IgE–Fab complex. Structural and physicochemical descriptors were calculated after standardized preparation without additional energy minimization (RAW) and with minimization (MIN). Similarity to Bos d 5 was assessed using principal component analysis (PCA), Euclidean distance-based ranking, and group comparisons using permutation tests.

Results: The dataset comprised the Bos d 5 reference structure (n = 1), respiratory allergens (n = 23), human lipocalins (n = 18), and β-lactoglobulin (βLG) homologs (n = 17 Across mapped linear IgE-binding regions, non-bovine ruminant βLGs most frequently ranked closest to Bos d 5, occupying the top rank in 8 of 12 analyses under the RAW protocol and 7 of 12 analyses under the MIN protocol. For the mapped conformational epitope region, non-bovine ruminant βLGs showed the smallest median distance to Bos d 5 under both protocols (RAW: 0.82; MIN: 1.10). The comparison with respiratory allergens was significant under RAW ($p = 0.0311$), but not under MIN ($p = 0.0774$). In the MIN composite analysis, median distances were 1.60 for non-bovine ruminant β-lactoglobulins, 4.46 for respiratory allergens and 4.51 for human lipocalins. Several respiratory lipocalins, including Fel d 4, Can f 4, Mus m 1, Fel d 7, and Cav p 1, repeatedly appeared among nearer neighbors in selected descriptor sets; however, this pattern was not consistent across all descriptor sets.

Conclusion: This study developed a reproducible structural comparison workflow for evaluating IgE-binding surface similarity across homologous lipocalins. Application to Bos d 5 consistently recovered close similarity among non-bovine ruminant βLGs, as expected from their close evolutionary and structural relationship. A subset of respiratory lipocalins showed partial structural similarity in selected analyses, but this pattern was not consistent across all comparisons. The workflow therefore provides a way to prioritize candidate proteins and surface regions for experimental testing of IgE binding and possible cross-reactivity.

ACKNOWLEDGEMENTS

I would like to thank my supervisory committee for their guidance and support throughout this research project. I am especially grateful to my principal supervisor, Dr. Floriano, for her continued supervision, support, and constructive feedback. Her guidance helped keep the work focused when attention to detail became overwhelming, and her advice consistently helped direct the project toward its main objectives.

I would also like to thank Dr. Dayer for his valuable input in identifying key conceptual and interpretative gaps throughout the project. I also appreciate his willingness to assist with training in relevant experimental techniques and guide future validation work.

I am grateful to Dr. Oba for careful feedback and for pointing out details that required clarification and improvement. I also appreciate his encouragement to participate in academic discussions and events that supported my professional development.

I also acknowledge access to computational resources provided through the Fir cluster, a computing system of the Digital Research Alliance of Canada. Access to the cluster was provided by Dr. Mawhinney.

This work was supported by the Graduate Student Award in Asthma Research (MSc), jointly funded by Asthma Canada, the Canadian Allergy, Asthma and Immunology Foundation (CAAIF), and the Canadian Institutes of Health Research – Institute of Circulatory and Respiratory Health (CIHR-ICRH), for the 2024–2025 funding period.

TABLE OF CONTENTS

ABSTRACT II
ACKNOWLEDGEMENTS III
TABLE OF CONTENTS IV
LIST OF TABLES V
LIST OF FIGURES VI
LIST OF ABBREVIATIONS VII
1 INTRODUCTION 1
1.1 Literature review 1
1.2 Study rationale, goal, and objectives 27
2 METHODS 29
2.1 Structural dataset assembly and preprocessing 29
2.2 Workflow development 30
2.3 Workflow validation 46
2.4 Workflow application 51
3 RESULTS 62
3.1 Structural dataset 62
3.2 Workflow development and validation 67
3.3 Workflow application 87
4 DISCUSSION 129
5 CONCLUSION 136
6 FUTURE WORK 137
7 REFERENCES 139
APPENDICES 153
APPENDIX I 154
APPENDIX II 155

LIST OF TABLES

Table 1-1. Representative IgE cross-reactivity between lipocalin proteins. 20
Table 3-1. Experimentally determined IgE Fab–allergen complexes. 63
Table 3-2. Lipocalin allergens assembled from the WHO/IUIS database. 64
Table 3-3. Reviewed human lipocalin proteins retrieved from UniProtKB..... 65
Table 3-4. Other beta-lactoglobulin proteins sourced from UniProtKB. 66
Table 3-5. Identity-control descriptor values under RAW and MIN protocols. 69
Table 3-6. Structural alignment validity and quality by template complex. 84
Table 3-7. Lipocalin query groups used for comparison with Bos d 5..... 87
Table 3-8. Minimally redundant Bos d 5 linear IgE-binding regions. 89
Table 3-9. Residue composition of the Bos d 5 conformational epitope. 91
Table 3-10. Retained structural and physicochemical descriptors. 97
Table 3-11. Variance explained by PC1–PC2 and retained PCs for the mapped linear epitope regions. 103
Table 3-12. PCA variance summary for retained descriptor sets. 107
Table 3-13. Aeroallergens among the ten closest query structures for each mapped linear epitope region. 117
Table 3-14. Ten closest query structures to the Bos d 5 reference for the mapped conformational epitope region. 118
Table 3-15. Ten closest query structures to the Bos d 5 reference for each descriptor set. 118
Table 3-16. Pairwise permutation-test results for the mapped conformational epitope median Euclidean distances. 127
Table 3-17. Pairwise permutation-test results for comparative Fab-allergen model descriptor distances. 128

LIST OF FIGURES

Figure 1-1. Overview of allergic sensitization and the IgE-mediated effector response. . 2
Figure 1-2. Structural organization of IgE and IgG antibodies. 4
Figure 1-3. Structural basis of allergen–antibody recognition. 10
Figure 1-4. Lipocalin fold conservation. 12
Figure 1-5. Contact surfaces in IgE–allergen complexes. 14
Figure 2-1. Overview of the epitope mapping workflow. 32
Figure 2-2. Overview of the template-based comparative structural workflow. 33
Figure 3-1. Concordance between experimental and reconstructed IgE Fab–allergen complexes. 69
Figure 3-2. Protocol sensitivity of interface descriptors. 75
Figure 3-3. Distribution of protocol-level descriptor changes after minimization. 76
Figure 3-4. Structural effects of energy minimization on the IgE Fab–Bos d 5 interface. 77
Figure 3-5. Workflow output by template complex and protocol. 83
Figure 3-6. Structural alignment quality by template complex. 85
Figure 3-7. Runtime of workflow execution by template complex. 86
Figure 3-8. Bos d 5 linear IgE-binding regions mapped onto the Bos d 5 structure. 90
Figure 3-9. Bos d 5 conformational epitope mapped on Bos d 5 structure. 92
Figure 3-10. Mapping quality for Bos d 5 MLERs across query lipocalins. 94
Figure 3-11. Mapping quality for the contact-defined Bos d 5 MCER. 95
Figure 3-12. CEV of PCA models fitted to retained MLER descriptors. 104
Figure 3-13. PCA biplots of retained descriptor differences for MCER. 105
Figure 3-14. PCA biplots of retained composite descriptors. 106
Figure 3-15. Nearest-neighbor rankings for the MCER. 115
Figure 3-16. Nearest-neighbor rankings for composite descriptor sets. 116
Figure 3-17. Group differences in RAW MLER Euclidean distances to Bos d 5. 123
Figure 3-18. Group differences in MIN MLER Euclidean distances to Bos d 5. 124
Figure 3-19. Group differences in MCER Euclidean distances to Bos d 5. 125
Figure 3-20. Group differences in composite descriptor Euclidean distances to Bos d 5. 126

LIST OF ABBREVIATIONS

AFDB	AlphaFold Protein Structure Database
B cell	B lymphocyte
BSA	Buried surface area
CD4+	Cluster of differentiation 4 positive
CD40	Cluster of differentiation 40
CD40L	CD40 ligand
CDR	Complementarity-determining region
CEV	Cumulative explained variance
CMA	Cow's milk allergy
CMP	Cow's milk protein
DBPCFC	Double-blind placebo-controlled food challenge
ELISA	Enzyme-linked immunosorbent assay
ESP	Electrostatic surface potential
FA	Food allergy
Fab	Fragment antigen-binding
Fc	Fragment crystallizable
FcεRI	High-affinity IgE receptor
FcγR	Fc gamma receptor
FcRn	Neonatal Fc receptor
FDR	False discovery rate
GRAVY	Grand average of hydropathy
H-bond	Hydrogen bond
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IL	Interleukin
ILC2	Group 2 innate lymphoid cell
MCER	Mapped conformational epitope region
MHC II	Major histocompatibility complex class II
MIN	Minimized structural-analysis protocol
MLER	Mapped linear epitope region
PC	Principal component
PCA	Principal component analysis
PDB	Protein Data Bank
PI	Protrusion index
pLDDT	Predicted local distance difference test
QSAR	Quantitative structure-activity relationship
QSPR	Quantitative structure-property relationship
RAW	Non-minimized structural-analysis protocol
RMSD	Root-mean-square deviation
RSA	Relative solvent accessibility
SASA	Solvent-accessible surface area
sIgE	Specific immunoglobulin E
SPADE	Surface comparison-based Prediction of Allergenic Discontinuous Epitopes
SPT	Skin prick test
T cell	T lymphocyte

Th2	Type 2 helper T cell
TSLP	Thymic stromal lymphopoietin
UniProtKB	UniProt Knowledgebase
VH	Heavy-chain variable domain
VL	Light-chain variable domain
WHO/IUIS	World Health Organization and International Union of Immunological Societies
YASARA	Yet Another Scientific Artificial Reality Application

1 INTRODUCTION

1.1 Literature review

1.1.1 Immune basis of IgE-mediated allergy

Allergy arises when the adaptive immune system mounts an antigen-specific response against otherwise harmless environmental proteins (allergens).[1,2] An antigen is any molecule specifically recognized by adaptive immune receptors, whereas an allergen is an antigen that triggers a hypersensitivity or allergic reaction in susceptible individuals.

Allergy mediated by immunoglobulin E (IgE) develops in two stages, namely sensitization and the effector response after re-exposure. Sensitization occurs when allergens pass through the epithelial barriers and are taken up by dendritic cells. As illustrated in Figure 1-1, dendritic cells break down these allergen into peptides and display them on major histocompatibility complex class II (MHC II) molecules to naïve cluster of differentiation 4-positive T lymphocytes (CD4+ T cells).[3–5] In individuals who do not have allergies, regulatory mechanisms such as regulatory T cells ensure that tolerance is maintained towards these antigens.[3] By contrast, there is a shift towards a type 2 helper T (Th2) cell phenotype in those with allergic diseases. These Th2 cells release interleukin (IL)-4, IL-5, IL-9 and IL-13, all of which lead to type 2 inflammation and cause activation of B lymphocytes (B cells).[3,4] When activated, B cells undergo class switching to produce IgE as a result of signalling through CD40 and CD40 ligand (CD40L), together with the action of IL-4 and IL-13, and with the support of Th2 or T follicular helper cells.[3,4,6] Class switching to IgE can take place either directly from immunoglobulin M (IgM) or indirectly via immunoglobulin G (IgG) intermediates.[7]

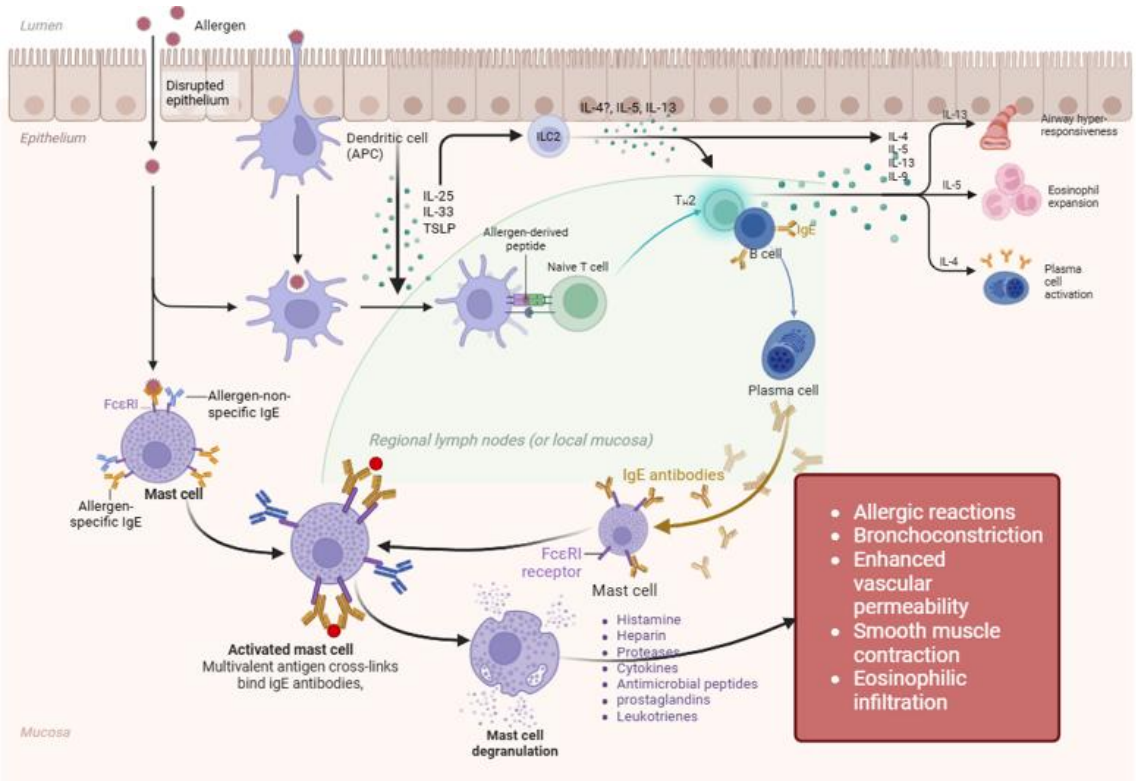

Figure 1-1. Overview of allergic sensitization and the IgE-mediated effector response.

Allergens crossing a disrupted epithelium are captured by dendritic cells and processed for presentation to naïve cluster of differentiation 4-positive (CD4+) T cells. Epithelial cytokines, including interleukin (IL)-25, IL-33, and thymic stromal lymphopoietin (TSLP), together with antigen presentation promote type 2 immune polarization, including type 2 helper T (Th2) and group 2 innate lymphoid cell (ILC2) responses. Th2-derived cytokines (IL-4, IL-5, IL-9, and IL-13) support activation of B lymphocytes (B-cells), class switching to immunoglobulin E (IgE), eosinophilic inflammation, and airway hyper-responsiveness. Plasma cells produce allergen-specific IgE, which binds with high affinity to the high-affinity IgE receptor (FcεRI) on mast cells. Upon re-exposure, an allergen that can bind more than one surface-bound IgE at the same time links IgE–FcεRI complexes on sensitized mast cells and triggers degranulation and release of mediators. Created with BioRender.com.

The resulting plasma cells produce IgE antibodies that target the allergen. These antibodies bind to high-affinity IgE receptors (FcεRI) on mast cells and basophils, preparing the body for a reaction.[3–5] The effector phase occurs after re-exposure to the allergen. If the allergen can bind more than one receptor-bound IgE at the same time, it bridges adjacent IgE–FcεRI complexes on sensitized mast cells or basophils.[3–5,7] This

triggers intracellular signals, leading to degranulation and the release of substances such as histamine, leukotrienes, proteases, and cytokines. These mediators drive the immediate features of allergic disease, including increased vascular permeability, mucosal edema, smooth muscle contraction, and recruitment of inflammatory cells.[3–5,8]

These processes establish IgE as the central antibody in immediate allergic reactions. However, the allergic response is shaped not only by IgE itself, but also by how IgE compares with other antibody isotypes, particularly IgG, which can have distinct and sometimes opposing effects.

1.1.2 IgE versus IgG roles

Mammals produce five distinct immunoglobulin isotypes: IgA, IgD, IgE, IgG, and IgM.[9,10] Among these, IgE and IgG are particularly important in allergy research and clinical management because they play distinct, and often opposing, roles in allergic disease.[11] These two isotypes differ in abundance, receptor binding, and effector functions.[12,13] IgG is the main antibody found in serum, while IgE is found in much smaller amounts.[12–16] Despite being less common, IgE is the key antibody involved in immediate allergic reactions and allergic diseases.[15,16]

As illustrated in Figure 1-2, the main difference between IgE and IgG is the structure of the heavy chain constant region. IgE contains four constant domains, whereas IgG contains three.[13,17] This difference is associated with distinct receptor binding and effector functions. IgE has a high affinity for FcεRI on mast cells and basophils and may stay bound to these cells even when there is no allergen present.[14,16] When an allergen cross-links adjacent FcεRI-bound IgE molecules, it triggers degranulation and release of inflammatory mediators such as histamine.

IgG has a different role; it mainly binds to Fc gamma receptors (FcγRs) and the neonatal Fc receptor (FcRn) and occurs mostly in a soluble form.[11,18] When it comes to allergy, allergen-specific IgG (IgG4) can act as a blocking antibody by binding to the allergen before it reaches the cell-bound IgE.[11,18] This blocking function is one mechanism associated with allergen immunotherapy and immune tolerance. However, effective blocking generally requires high IgG levels relative to IgE because FcεRI-bound IgE is already positioned on effector cells and only limited receptor cross-linking is needed to trigger activation.[16]

These differences between IgE and IgG provide an immunological framework for understanding allergic diseases. The next section moves from antibody function to clinical presentation, focusing on the co-occurrence of food allergy and asthma in childhood.

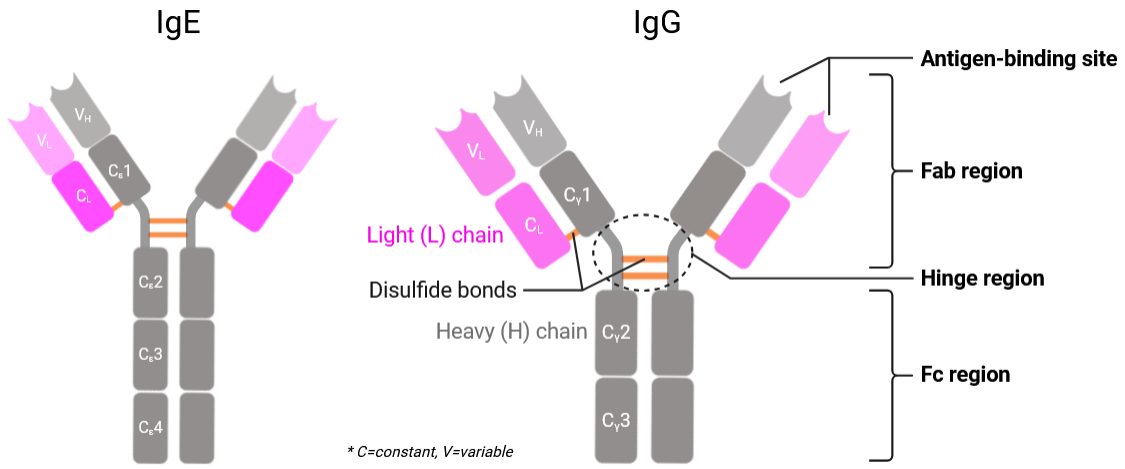

Figure 1-2. Structural organization of IgE and IgG antibodies.

Both antibodies contain two light chains and two heavy chains and are organized into antigen-binding Fab regions and an Fc region. The main structural difference is in the heavy-chain constant region: IgE contains four constant domains (Cε1–Cε4), whereas IgG contains three (Cγ1–Cγ3) and a hinge region. These structural differences contribute to distinct receptor binding and effector functions. Created with BioRender.com.

1.1.3 Clinical co-occurrence of food allergy and asthma

1.1.3.1 Epidemiology and co-occurrence of food allergy and asthma

Epidemiological studies show that food allergy and asthma often occur together in children.[4,19–26] Population-based and cohort analyses show that children with FA have higher asthma prevalence than non-allergic peers, including after adjustment for other atopic comorbidities.[8,25–32] For example, a U.S. population analysis based on serologic sensitization reported food sensitization in 27.5% of individuals with asthma versus 14.9% of non-asthmatics using panels that included milk, egg, peanut, and shrimp.[25] Prospective and cross-sectional pediatric cohorts also report increased odds of asthma among children with clinically defined FA (odds ratios ~4.9–5.3), with higher prevalence in those with multiple food allergies.[19]

Estimates further suggest that nearly half of children with asthma have at least one coexisting food allergy, and that many children with IgE-mediated FA report respiratory allergic symptoms.[4] Clinically, asthma is a documented risk factor for severe or fatal food-induced anaphylaxis[21,33,34], and FA has been associated with increased asthma morbidity in some cohorts (e.g., hospitalization and systemic steroid use).[22,32] FA has also been associated with increased risk of severe asthma exacerbations (odds ratio 5.89)[33] and poorer lung function metrics, including reduced forced expiratory flow.[32] These findings are largely observational and do not resolve whether co-occurrence reflects shared allergen-specific IgE recognition or a broader severe-atopy phenotype.

1.1.3.2 Prevalence of cow's milk allergy

IgE-mediated cow's milk allergy (CMA) is among the earliest and most prevalent food allergies in infancy and early childhood.[35–39] Diagnosis is based on a history consistent with immediate hypersensitivity supported by specific IgE (sIgE), skin prick testing (SPT),

or the double-blind placebo-controlled food challenge (DBPCFC) as reference standard.[36,38–40] Reported CMA prevalence in early childhood ranges from ~0.5% to 3% globally [39,41], with higher values reported under broader diagnostic criteria.[38,39,42,43] In Canada, physician-reported prevalence is estimated at 0.4%, lower than self-reported rates.[20] Bovine β-lactoglobulin (bovine βLG; Bos d 5) is a major allergen in cow's milk proteins (CMPs) and is recognized by IgE in about 60 to 80% of patients with CMA in some cohorts.[41,43–45] Bovine βLG is particularly relevant immunologically as it has no equivalent in human breast milk, unlike the caseins and α-lactalbumin. [38,44–47]. This makes cβLG an entirely foreign antigen to the infant's developing immune system, thereby lacking a basis for pre-existing immune tolerance and increasing its potential to trigger allergic sensitization.[48] Higher IgE reactivity to cβLG and caseins has been associated with more persistent CMA phenotypes, although sensitization levels do not reliably predict clinical reactivity or long-term outcomes.[35,49–52]

While CMA represents an important early-life FA, allergic asthma is one of the most common later respiratory manifestations of atopy. Understanding both conditions is important for examining how they may be clinically linked.

1.1.3.3 Prevalence of allergic asthma

Asthma is clinically heterogeneous, but allergic (IgE-mediated) asthma predominates in childhood and accounts for approximately 67% of cases in some longitudinal cohorts.[24,53] This phenotype differs from non-allergic asthma, which more often presents later with different inflammatory profiles.[53] Diagnosis involves a history of recurrent wheeze, cough, or dyspnea confirmed by objective measures of variable expiratory airflow limitation (spirometry) and evidence of IgE sensitization to aeroallergens [30,54,55]. In high-income regions such as North America the incidence of pediatric asthma remains high, exceeding 3,200 cases per 100,000 children in some studies.[55] On a global scale, asthma affects

more than 300 million people.[56,57] In Canada it affects about 12-25% of children.[58] Allergic asthma phenotyping emphasizes IgE reactivity to aeroallergens, including house dust mite, animal dander, and pollens.[53,59–61] Diagnostic assignment in early childhood remains challenging due to symptom overlap with other respiratory conditions and limited feasibility of objective lung function testing in young children.[4,53,55,62] These limitations affect interpretation of early allergic trajectories but do not identify allergen-specific IgE binding determinants.

With the clinical features of CMA and allergic asthma outlined, the next important question is whether these conditions are linked across childhood development. Longitudinal studies have therefore examined whether early cow's milk allergy is associated with later asthma.

1.1.3.4 Clinical association between cow's milk allergy and later asthma

Longitudinal birth cohorts report that IgE-mediated CMA in infancy is associated with increased likelihood of later asthma, including after accounting for other atopic conditions.[19,30,62] This pattern is often described within the atopic march framework, in which early FA and sensitization precede respiratory allergic disease.[4,62–64] In prospective cohorts, children with IgE-mediated CMA show higher prevalence of asthma (31% vs 13%) and rhinoconjunctivitis (66% vs 21%) at school age than non-allergic controls[63], with risk heightened by multiple food allergies or atopic dermatitis.[26] Challenge-confirmed FA at one year of age has been reported to increase asthma risk by early childhood even when FA later resolves.[26] Children with both FA and asthma show increased risk of severe outcomes, including near-fatal asthma exacerbations and food-induced anaphylaxis.[8,29,33,66]

Although these studies support a clinical association between CMA and later asthma, the explanation for the underlying molecular basis of this relationship is limited. One possible explanation is cross-reactivity, in which the same IgE antibody recognizes allergens from different sources.

1.1.3.5 Cross-reactivity as one possible molecular explanation

To interpret the association between CMA and asthma, it is important to distinguish molecular cross-reactivity from co-sensitization and clinical co-occurrence. IgE cross-reactivity occurs when the same IgE specificity recognizes different allergens.[12,65,66] In contrast, co-sensitization refers to different IgE antibodies generated against different allergens, and clinical co-occurrence is simply an epidemiological link without molecular evidence. Therefore, true IgE cross-reactivity can only be established based on evidence showing that one population of IgE molecules recognizes surfaces that are either shared or similar among allergens.[15,66]

This distinction is important for this study because CMA and allergic asthma frequently occur together, but their clinical association alone does not indicate a shared molecular mechanism. Shared IgE recognition is one possible explanation that can be evaluated by examining allergen structure and antibody binding.

1.1.4 Mechanism of allergen–antibody interactions

Antibody recognition of an allergen occurs at a noncovalent interface between an allergen epitope and an antibody paratope.[2,12,16,17,69,70] The part of the allergen that binds to the antibody is known as the epitope. The paratope, which is the antigen-binding site that complements the epitope, is found in the variable domains of the Fab region of the antibody and makes direct contact with the epitope.[12,65–67] The paratope is primarily formed by residues within the complementarity determining regions (CDR), although the

residues that contact antigen do not always match canonical CDR definitions exactly.[68–72]

Protein B-cell epitopes may be linear, comprising residues that are adjacent in the primary sequence, or discontinuous, when residues separated in sequence are brought together on the folded protein surface.[2,12] Studies that focused on the structural analysis of antibody-antigen complexes showed that protein B-cell epitopes usually involve about 15 residues in contact, although epitope size varies depending on the antigen surface and the distance criterion used to define contacts.[67]

In solved antibody–antigen complexes, the antigen residues that make direct contact with antibody residues define the structural epitope, while the corresponding antibody contact residues define the structural paratope.[2,12,16,73,74] This structural definition does not necessarily identify the subset of residues that contribute most strongly to binding energy, which is more appropriately termed the functional epitope and usually requires mutational or thermodynamic validation.[12,16,73] A representative IgE Fab–allergen complex is shown in Figure 1-3.

Allergen–antibody binding depends on three-dimensional shape and physicochemical complementarity. The interface can be stabilized by direct and water-mediated hydrogen bonds, electrostatic interactions, salt bridges, van der Waals contacts, hydrophobic burial, and aromatic interactions such as π–π and cation–π contacts (Figure 1-3).[67,75–78] The net binding energy also depends on desolvation and any conformational changes that occur during binding.Click or tap here to enter text.[12,75,78,79]

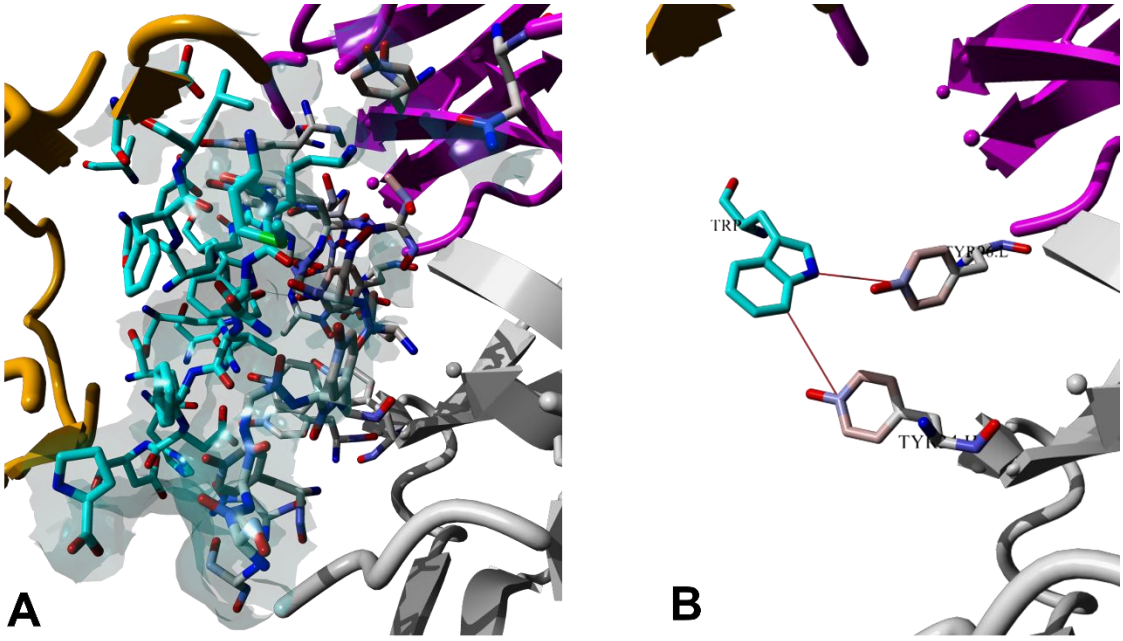

Figure 1-3. Structural basis of allergen–antibody recognition.

(A) Interface view of the allergen–Fab complex (PDB: 2VXQ), showing allergen residues within 5.0 Å of the antibody paratope. The allergen epitope is shown in cyan and the contacting Fab paratope residues are shown as sticks colored by charge, with the interface molecular surface also displayed in cyan. (B) π–π interactions. Interacting epitope and paratope residues are labeled.

1.1.5 Lipocalins as a structurally conserved allergen family

Lipocalins are small extracellular proteins (about 160–200 amino acids and approximately 18–40 kDa) that share a highly conserved tertiary structure.[6,61,80,81] Their defining structural feature is an eight-stranded antiparallel β-barrel that encloses an internal ligand-binding cavity (the calyx shown in Figure 1-4B), which accommodates small hydrophobic molecules.[6,69,80–82] As shown in Figure 1-4A, the β-barrel is typically accompanied by a C-terminal α-helix and an N-terminal 3_{10}-helix positioned against the exterior of the barrel.[6,80,81] Although their overall sequences are not very similar, lipocalins

retain a similar β-barrel core, while their surface residues and connecting loops differ [69,83–88] (Figure 1-4C).

It is important to note that a similar lipocalin fold does not always mean IgE-binding sites are shared. For example, Madhurantakam et al.[89] found that the dog allergen Can f 2 and the cat allergen Fel d 4 can both bind IgE, even though their sequences are less than 22% identical. However, the same study found no IgE cross-reactivity between Can f 2 and the mouse, horse, and rat lipocalins tested in five patient samples, even though their ligand-binding cavities are similar.[91] These findings show the importance of focusing on local features IgE-binding regions as opposed to just the overall fold.

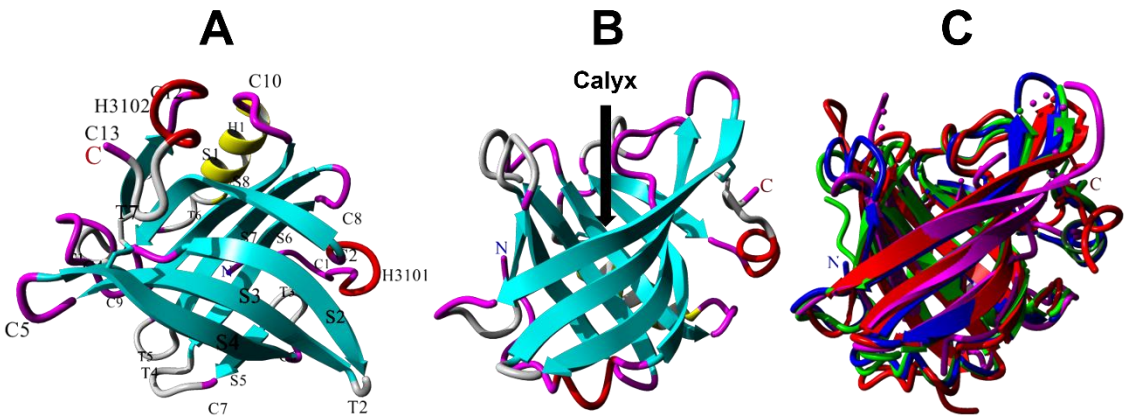

Figure 1-4. Lipocalin fold conservation.(A) Cartoon representation of the canonical lipocalin architecture, consisting of an eight-stranded antiparallel β-barrel (strands S1–S8, cyan) forming the calyx, capped by a short α-helix (H1, yellow) and two 3_{10}-helix (H3101–H3102, red). Connecting coil and turn regions are labeled (C1–C13, magenta; T1–T7, grey). The N and C termini are indicated. (B) The calyx, the internal ligand-binding cavity formed by the β-barrel, is shown. (C) Structural superposition of representative lipocalins highlighting fold conservation: *Bos d* 5 (2R56, blue), sheep β-lactoglobulin (4CK4, green), cat allergen Fel d 4 (8AMC, red), and human tear lipocalin / LCN1 (1XKI, magenta). All structures were aligned and rendered in YASARA.

1.1.6 IgE–allergen complex structures and interface constraints

Direct structural definition of IgE epitopes is limited by the low abundance of IgE (ng/mL) in serum and the rarity of allergen-specific IgE-expressing B cells (3×10^{-7} to 7×10^{-6}) in peripheral blood.[7,12,13,18,90,91] Methods such as X-ray crystallography require milligram quantities of pure and homogeneous antibody–allergen complexes, something that is difficult to achieve with IgE.[12,13,81] Human IgE responses are also polyclonal, so assays based on serum assays reflect mixtures of IgE clones instead of single defined interactions.[7,12,13,18,90,91] These experimental challenges have led to the development of approaches for isolating monoclonal IgE antibodies, including combinatorial display libraries and hybridoma-derived antibodies.[12,18,73] Although these approaches enable structural characterization of individual antibody–allergen interactions, each monoclonal antibody represents only a fraction of the polyclonal IgE repertoire *in vivo*.

Peptide scanning can find linear IgE-binding regions, but it may not show how these segments are shaped or exposed in the folded allergen.[18,66,92] Most IgE epitopes on intact allergens are conformational, comprising residues that are far apart in the sequence.[88,93–98] Mouse IgG antibodies can help point to possible IgE-binding regions, but their specificity does not always match how human IgE recognizes allergens.[3,7,17,99]

The determination of high-resolution crystal structures of IgE Fan–allergen complexes has provided the first atomic descriptions of IgE Fab–allergen interfaces and the structural basis of IgE recognition [2,12,16,17] (Figure 1-5). Structures of recombinant human IgE-derived Fabs in complex with allergens such as Bos d 5 (PDB: 2R56, Figure 1-5A) and the grass pollen allergen Phl p 2 (PDB: 2VXQ, Figure 1-5B) found that the epitopes are mostly flat, discontinuous areas made up of β-sheet surfaces.[16,17] Later, structures of Der p 2 with Fabs from naturally paired human IgE antibodies 2F10 and 4C8 showed different, non-overlapping epitopes (PDB: 7MLH, 8VK1; Figure 1-5C,D).[12] Across these structures, IgE binding is supported by localized geometric and physicochemical complementarity, but the observed interfaces span more than one epitope topology and IgE Fab–allergen orientation.[12,16,17,100]

IgE complex structures with lipocalin allergens further show that allergen conformational state can influence which surfaces are accessible. In the Can f 1 complexes, human IgE-derived Fabs 1J11 and 12F3 recognize some of the same epitope regions, but these are linked to different shapes at the N-terminal end (PDB: 9NPH/8VQF, 9NPI/9NPG; Figure 1-5F/E,H/G).[2] In the 12F3 complex, the N-terminal segment moves, allowing the heavy-chain CDR3 loop to reach into the calyx. These results demonstrate that the same lipocalin fold can support different ways for antibodies to bind.[13,16,17,77]

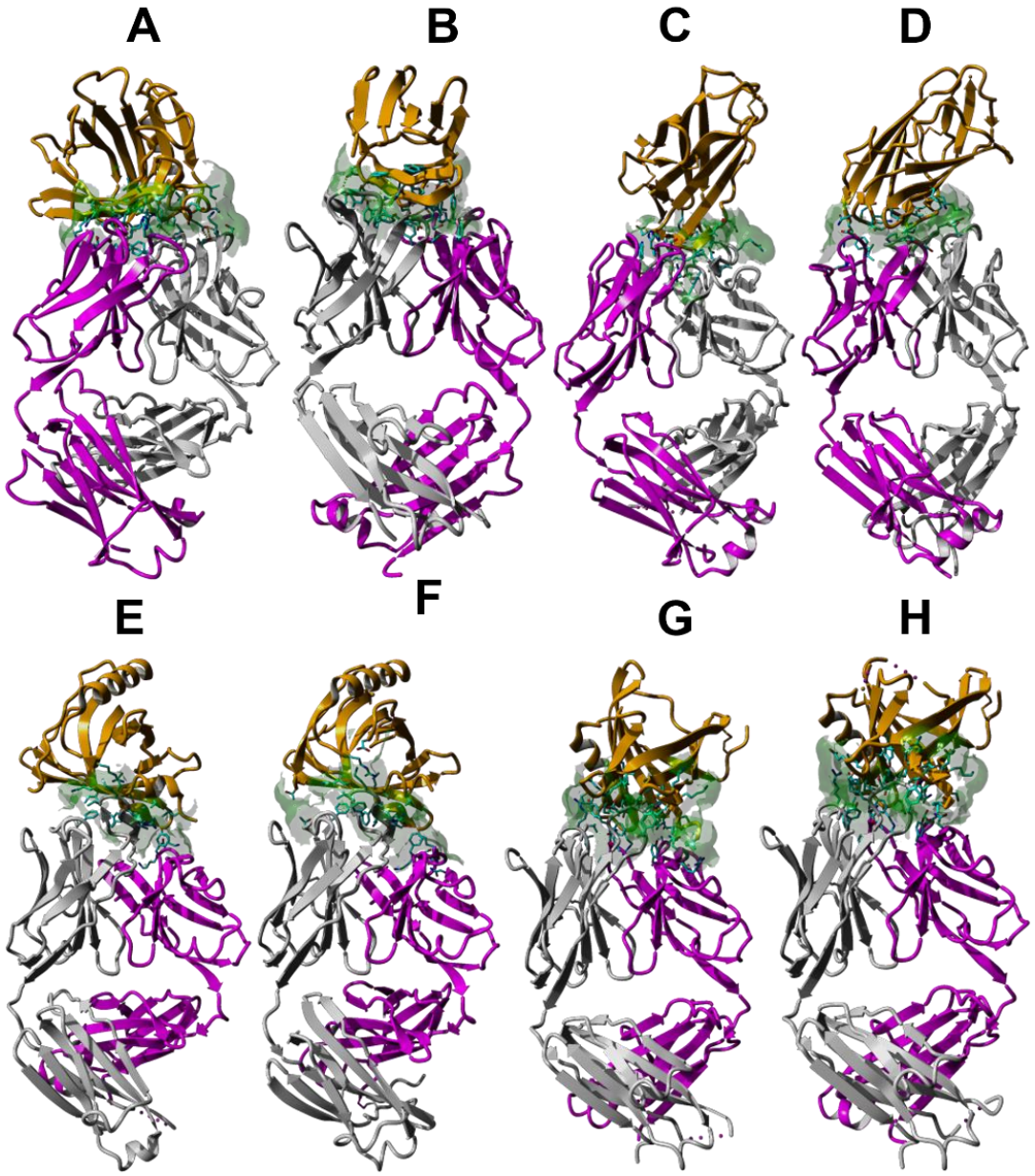

Figure 1-5. Contact surfaces in IgE–allergen complexes.

Panels show IgE-Fab bound to their allergens: (A) Bos d 5/β-lactoglobulin (PDB 2R56), (B) Phl p 2 (PDB 2VXQ), (C) Der p 2 (PDB 7MLH), (D) Der p 2 (PDB 8VK1), (E) Can f 1 (PDB 8VQF), (F) Can f 1 (PDB 9NPH), (G) Can f 1 (PDB 9NPG), and (H) Can f 1 (PDB 9NPI). Allergens are rendered as gold ribbons, while Fab light and heavy chains are shown as magenta and gray ribbons, respectively. For each complex, only interfacial residues (defined by a van der Waals surface gap cutoff of 0.5 Å) are displayed as sticks with element coloring, and the interacting region is highlighted by a semi-transparent green accessible surface. Together, these structures illustrate that IgE recognition targets diverse conformational epitopes on allergen surfaces, with variation in epitope topology and Fab–allergen orientation. All structures were rendered in YASARA.

1.1.7 Bos d 5 as a food lipocalin allergen

As mentioned earlier, bovine βLG is the most abundant whey protein and a major cow's milk allergen.[10,38,44,50,80] Structurally, bovine βLG adopts the canonical lipocalin architecture[5617,101] (Figure 1-4). Although bovine βLG can dimerize under physiological conditions, IgE binding has been reported at the monomer level, indicating that relevant IgE-binding surfaces are present on the folded monomer.[10,17,45,10410,17,45,102]

Early mapping of bovine βLG IgE-binding determinants used peptide scanning, fragments, and assays with denatured protein, which favor detection of linear epitopes.[35,45,103–109] These studies reported multiple IgE-reactive regions across the sequence, often including the C-terminus and several loops.[45] However, IgE binding to native bovine βLG as compared to denatured bovine βLG shows a strong dependence on the intact tertiary structure, and linear peptides generally fail to inhibit IgE binding to folded bovine βLG.[18,104,110] Because patient IgE is polyclonal, peptide-level patterns can also differ between individuals.[104] As a result, indirect mapping identifies candidate regions but does not define functional conformational epitopes on the native protein.[17]

The Bos d 5–IgE Fab complex structure (2R56; Figure 1-4A) provides an experimentally validated conformational epitope at atomic resolution.[17] The interface spans approximately 890 Å² and is formed mainly by β-strand segments on a relatively planar surface.[17] The epitope lies on a single monomer and is distinct from oligomerization interfaces.[17] These findings define an experimentally defined IgE-binding surface on a food-derived lipocalin, but it does not establish whether respiratory or endogenous lipocalins present analogous IgE-compatible surfaces.

Although the Bos d 5–IgE complex defines one validated conformational epitope, it remains unclear how small sequence differences on the same structural scaffold alter

IgE recognition. Closely related non-bovine β-lactoglobulins (oβLG) therefore provide a useful natural comparison set for testing how local substitutions affect an otherwise conserved allergen surface.

1.1.8 β-lactoglobulin homologs and milk-protein cross-reactivity

β-Lactoglobulins from closely related ruminants (buffalo (bβLG), sheep (sβLG), goat (gβLG)) provide a controlled model for testing how a small number of amino-acid substitutions affects IgE recognition on a conserved lipocalin scaffold. These proteins share high sequence identity with bovine βLG[40,47,111] and their structure still exhibit the same β-barrel arrangement[112,113] With these proteins it is possible to look at local sequence differences in the context of a similar overall structure.

Certain substitutions do alter the chemical characteristics of the relevant surface areas. For instance, the residue at position 130 in bovine βLG is Asp130, whereas in sβLG it is Asn130 and in gβLG it is Lys130, with the numbering referring to the mature protein.[113,114] In polyclonal IgE responses, such changes are expected to affect only IgE clones that contact the altered region, producing inhibition patterns specific to each patient in serological assays.[104,115]

Reindeer β-lactoglobulin (rβLG) provides an additional test case for high sequence identity. rβLG shares about 94% sequence identity with bovine βLG and differs by fewer than ten residues; its structure (PDB: 1YUP) is very similar to bovine βLG (RMSD ~0.45 Å)[116] Reported partial IgE cross-reactivity between bovine βLG and rβLG in patients sensitized to bovine βLG[40,116] is consistent with local substitutions altering electrostatics near a region that contacts IgE, without changing the overall β-sheet topology[17], although such mapping alone does not prove causality.

At larger phylogenetic distances, sequence divergence is associated with additional structural differences and weaker cross-reactivity. Equine βLGs show reported differences in oligomerization behavior and local β-barrel organization[85,101,117,118], and IgE cross-reactivity with bovine βLG is weak or absent in several studies[47,118]

1.1.9 Structural and physicochemical descriptors of IgE epitopes

Experimentally determined IgE Fab–allergen complexes provide a basis for selecting structural descriptors that characterize IgE-binding surfaces. These complexes provide a basis for comparing surface patches that are capable of binding to IgE.[2,12,16,17,68,75] In this context, a patch is a local, spatially defined region of the allergen surface.[119–121]

The surface features that are relevant consist of solvent-accessible surface area (SASA), relative solvent accessibility (RSA), protrusion, electrostatic surface potential (ESP), and residue composition.[69,76,99,119,122–124] These properties indicate whether a candidate region is solvent exposed, presents a comparable local shape, and possesses chemical features that could support antibody binding. Interface-level descriptors provide a quantitative measure of complex stability, including buried surface area (BSA) and the density of hydrogen bonds and van der Waals contacts.[13,18,75,76,123,125] Furthermore, the interface is stabilized by a network of non-covalent interactions, including ionic salt bridges, hydrophobic burial, and aromatic contacts such as π–π stacking and cation–π interactions, which are critical for high-affinity recognition.[7,13,75,77,126] Together, these descriptors provide complementary information on the structural and physicochemical basis of epitope–paratope interactions in experimentally determined or modeled complexes.

Epitope presentation can also be affected by oligomerization, ligand binding, conformational state, post-translational modifications (PTMs), and differences between

recombinant and native allergen forms.[127–130] Therefore, although descriptor analysis provides a standardized framework for comparing candidate IgE-binding surfaces across homologous proteins in this thesis, experimental IgE-binding and inhibition assays remain essential for confirming functional cross-reactivity and clinical relevance.

1.1.10 IgE cross-reactivity in lipocalins: evidence and molecular interpretation

1.1.10.1 Experimental evidence

IgE cross-reactivity is commonly investigated using competitive inhibition immunoassays, including enzyme-linked immunosorbent assay (ELISA) inhibition.[15,66] For example, in a standard assay, patient serum is preincubated with allergen A, after which residual IgE binding to an immobilized target allergen B is measured. Inhibition of IgE binding to B by A indicates that at least a proportion of serum IgE antibodies recognize epitopes shared between the two allergens, whereas detection of IgE binding to both allergens in separate assays demonstrates co-sensitization.[15,66] Homologous inhibition serves as a positive control for assay performance, whereas reciprocal inhibition experiments determine whether shared IgE recognition is observed in both directions.[15,66]

Partial inhibition may reflect that only a subset of IgE antibodies recognizes shared epitopes, differences in antibody affinity or avidity, or the presence of allergen-specific epitopes absent from the corresponding inhibitor. Quantitative inhibition is also influenced by experimental factors, including inhibitor concentration and molar excess, target immobilization, protein purity, isoform composition, glycosylation, protein folding, and the use of natural or recombinant allergens.[66]

Mammalian lipocalins include several well-characterized examples of IgE cross-reactivity.[131] Representative examples are summarized in Table 1-1. Early component-based studies demonstrated limited IgE cross-reactivity between Can f 1 and human tear

lipocalin (LCN1), Can f 1 and Can f 2, and Equ c 1 and Mus m 1.[129,131] Subsequent studies identified additional cross-reactive relationships, particularly among Can f 6, Fel d 4, and Equ c 1, as well as between Can f 1 and Fel d 7.[82,129,132–136] Additional experimentally supported examples include Can f 2/Fel d 4, Can f 4/bovine odorant-binding protein, Cav p 6/Fel d 4/Can f 6, the cockroach lipocalins Bla g 4 and Per a 4, and bovine and reindeer β-lactoglobulins[89,131,135,137–139] (Table 1-1). Cross-reactive urinary allergens have also been demonstrated in mouse and rat urine extracts, although these experiments were performed using whole extracts rather than purified Mus m 1 and Rat n 1.[140]

Where reciprocal inhibition has been examined, the degree of inhibition may differ between assay orientations, suggesting that shared and allergen-specific IgE populations coexist. Interpretation of inhibition data also requires consideration of methodological differences among studies, including allergen preparation, assay format, and inhibitor concentration, all of which may influence the magnitude of the measured response.[15]

Table 1-1. Representative IgE cross-reactivity between lipocalin proteins.

Lipocalin pair	IgE cross-reactivity observed	Reference(s)
Can f 6/Fel d 4	Reciprocal but patient-dependent inhibition	82,132,133
Can f 6/Equ c 1	Strong inhibition in selected patients; extent varied among sera	132,134
Fel d 4/Equ c 1	Reciprocal partial inhibition	132,135
Can f 1/Fel d 7	Partial, asymmetric and patient-dependent inhibition	141
Can f 1/Can f 2	Low or asymmetric inhibition	131,135
Can f 1/human LCN1[a]	Limited and asymmetric inhibition	131
Can f 2/Fel d 4	Partial and patient-dependent inhibition despite low sequence identity	89,135
Can f 4/bovine 23-kDa OBP[b]	Nearly complete inhibition by Can f 4; reciprocal inhibition was not tested	137
Equ c 1/Mus m 1	Weak and predominantly unidirectional inhibition	131
Cav p 6 / Fel d 4; Can f 6	Partial and asymmetric inhibition	138
Bla g 4/Per a 4	Partial reciprocal inhibition despite low sequence identity	139
Bos d 5/rβLG[c]	Partial inhibition by reindeer β-lactoglobulin; reciprocal inhibition was not tested	142

Only primary studies that examined IgE cross-inhibition using patient sera and isolated recombinant or natural lipocalins were included. Reciprocal inhibition indicates inhibition in both directions but does not imply an equivalent degree of inhibition. Results are described qualitatively because assay format, inhibitor concentration, protein preparation and serum selection differed among studies. Allergen names follow World Health Organization/International Union of Immunological Societies ((WHO/IUIS) nomenclature.[143]

[a] Human LCN1 is the endogenous human lipocalin-1, also known as tear lipocalin, and is included as a cross-reactive homolog rather than as a recognized allergen.

[b] No official allergen designation has been assigned; the descriptive protein name is therefore used.

[c] rβLG is a Bos d 5 homolog without a separate official allergen designation.

Abbreviations: IgE, immunoglobulin E; LCN1, lipocalin-1; rβLG, reindeer β-lactoglobulin; OBP; odorant-binding protein.

1.1.10.2 Molecular interpretation

The available experimental evidence suggests that IgE cross-reactivity between lipocalins arises from conserved surface features. Cross-reactive allergens are thought to present exposed conformational regions with sufficiently similar shape and physicochemical properties to permit recognition by the same IgE antibody, or by overlapping populations of polyclonal IgE antibodies.[18,66,70] Although lipocalins share a conserved β-barrel scaffold, this structural framework does not necessarily preserve the surface geometry, electrostatic potential, hydrophobicity, conformational flexibility, or residue interactions that determine antibody binding.[3,69,88] This distinction helps explain why cross-reactivity is selective instead of it being a general property of all conserved protein families.

1.1.11 Human lipocalins: immune tolerance and self-reactivity

Human lipocalins include lipocalin-1 (LCN1; tear lipocalin) and lipocalin-2 (LCN2; neutrophil gelatinase-associated lipocalin; Figure 1-4C).[144,145] Although their amino-acid sequences vary considerably, these proteins share the characteristic eight-stranded β-barrel of the lipocalin fold.[1,80,82] Some human lipocalins also show sequence and structural similarity to mammalian respiratory lipocalin allergens. However, evidence for IgE recognition of endogenous human lipocalins is limited.

The best documented example is the relationship between Can f 1 and human LCN1 [131]. These proteins share approximately 61% sequence identity.[1,82,131] Saarelainen et al. found low levels of IgE specific to LCN1 in seven patients who were allergic to dogs, six of whom also had IgE specific to Can f 1.[131] Competitive ELISA inhibition demonstrated partial, asymmetric IgE cross-reactivity, with Can f 1 producing greater inhibition of IgE binding to LCN1 than vice versa (Table 1-1). These findings indicate that Can f 1 and

LCN1 share IgE-binding determinants, although the study did not establish the primary sensitizing allergen or the clinical significance of LCN1-reactive IgE.[6,131]

Subsequent work showed that Can f 1- and LCN1-specific $CD4^+$ T-cell responses were low and broadly comparable in allergic and non-allergic individuals.[146] These observations suggest that differences in T-cell antigenicity alone are unlikely to account for the contrasting immunological behavior of Can f 1 and the endogenous protein LCN1. More broadly, IgE autoreactivity to a variety of human proteins has been described in patients with atopic dermatitis, particularly in those with severe disease or elevated total IgE concentrations.[7,147–150] However, the reported autoantigens encompass diverse protein families, and current evidence does not indicate that endogenous human lipocalins are frequent targets of autoreactive IgE.

Ultimately, the available evidence suggests that human lipocalins are useful reference proteins for examining structural features shared across the lipocalin family while also illustrating that structural similarity alone is insufficient to predict allergenicity or IgE cross-reactivity. Although limited IgE cross-reactivity has been demonstrated between Can f 1 and LCN1, endogenous human lipocalins are generally considered within the context of immunological tolerance rather than as clinically relevant allergens.

1.1.12 Computational approaches for IgE epitope and cross-reactivity analysis

1.1.12.1 Cross-reactivity screening based on sequence

Codex Alimentarius and European Food Safety Authority (EFSA) guidance recommend sequence similarity searches as an initial step in allergenicity assessment. A newly expressed protein is considered to have the potential for IgE cross-reactivity with a known allergen when it shares more than 35% sequence identity over a segment of at least 80 amino acids.[151,152] This criterion is widely used because it is rapid, transparent, and

readily applied across allergen databases. However, it is intended as a conservative screening threshold as opposed to evidence of allergenicity or clinically relevant cross-reactivity.

The main limitation is that sequence similarity is only a proxy for antibody recognition and does not indicate whether conserved residues are exposed on the protein surface, form a conformational IgE epitope, or adopt a similar three-dimensional orientation. This limitation is important for lipocalins because mammalian lipocalins often retain a conserved fold despite low sequence identity, and IgE cross-reactivity has been reported for some lipocalin pairs but not for others.[89,131,151] Sequence screening is therefore useful for initial risk assessment and for detecting clear homologues.

To address these limitations, sequence-based approaches have increasingly been complemented by methods that incorporate allergen structure, particularly because many IgE epitopes are conformational.

1.1.12.2 Epitope predictors based on antigen structure

A second class of methods predicts B-cell epitopes from the three-dimensional structure of a single antigen. Tools such as DiscoTope and SEPPA (Spatial Epitope Prediction of Protein Antigens) analyze antigen structure to identify likely conformational epitope patches based on surface exposure, residue properties, and spatial clustering.[153,154] Because these methods evaluate only the antigen and do not incorporate the sequence or structure of a specific antibody, their predictions cannot prove that the same IgE would bind to corresponding surfaces on different allergens.[153,154] Furthermore, most general B-cell epitope prediction methods have been developed and benchmarked using diverse antibody–antigen datasets rather than datasets focused specifically on IgE–allergen complexes.[154]

1.1.12.3 Allergen surface comparison methods

Some computational methods developed for allergy research compare allergen surfaces directly. SPADE (Surface comparison-based Prediction of Allergenic Discontinuous Epitopes) predicts candidate discontinuous IgE epitopes by comparing the surfaces of structurally related allergens and incorporating IgE cross-reactivity data.[70] It uses structural similarity, including surface shape and physicochemical properties, to identify surface regions that are shared by cross-reactive allergens but not by allergens with weak or absent cross-reactivity.[70]

Cross-React is another structure based method for predicting allergen cross-reactivity. It compares the amino acid composition and solvent accessible properties of a known or predicted epitope with surface patches on other allergens.Click or tap here to enter text.[2,12,16,17,67,732,12,16,17,67,73] The method is based on the idea that allergens may cross react when they contain accessible surface patches with similar composition and spatial organization.[155]

These methods are more directly relevant to allergy than sequence-based approaches or predictors that examine a single allergen because they compare allergens at the surface level, where conformational IgE epitopes are presented.[70,155] However, they generally evaluate antigen surfaces independently of a defined antibody–antigen interface, leaving molecular basis of antibody recognition unresolved.

1.1.12.4 Analysis and modeling of antibody–antigen interfaces

A further class of computational approaches involves prediction of epitopes from antibody–antigen complexes. For example, EpiPred combines geometric matching with an antibody–antigen interface scoring function to predict epitope regions and rank docking poses.[156] SEPPA-mAb (Spatial Epitope Prediction of Protein Antigens or

monoclonal antibodies) determines candidate epitopes by using the structure of the antigen and that of a particular monoclonal antibody.[157] In neither case is it necessary for the two partners to have already been resolved in a complex.

Other tools, such as AppA and Arpeggio, analyze existing antibody–antigen complexes by identifying contact residues, interfacial waters, hydrogen bonds, hydrophobic interactions, van der Waals contacts, and ionic interactions.[158,159] General interface-analysis approaches can also quantify buried surface area, contact residues, and interaction fingerprints, which are often more directly related to antibody binding.[159,160] These methods are useful because they describe the physical contacts formed across the interface.[156–160]

Complex models may be produced through template placement or docking when an experimentally determined antibody–antigen complex is available.[161] In template placement, structural alignment is used to place the query protein in relation to its partner in an experimentally determined complex. This differs from *de novo* docking, where the relative orientation of the two binding partners is searched without imposing a specific experimental complex as the template.[161] In antibody–antigen modeling, methods such as RosettaDock, SnugDock, and HADDOCK (High Ambiguity Driven protein–protein Docking) illustrate how interface information, including known or predicted paratope and epitope residues, can guide docking and local refinement of complex models.[162–164]

For IgE cross-reactivity, this strategy is particularly valuable when the template is an experimentally determined IgE Fab–allergen complex. The Fab defines the reference paratope and binding orientation, allowing homologous allergens to be superposed into the same position for comparison. Unlike allergen surface comparison methods, which identify similar exposed surface patches[69,70,155], template-based modeling evaluates

whether a homologous allergen preserves a compatible surface within the experimentally observed IgE-binding geometry. For lipocalins, whose conserved fold is often maintained despite relatively low sequence identity, this provides a framework for evaluating candidate cross-reactive surfaces under the structural constraints of a known IgE interface.

1.1.13 Unresolved questions

The reviewed literature leaves three main unresolved questions. Longitudinal and clinical studies associate IgE-mediated CMA with asthma.[23,30,62,165–167] However, these associations do not prove that milk and aeroallergens are recognized by the same IgE . Moreover, although inhibition studies can indicate cross-reactivity it cannot determine the specific antibody clones or the sites at which they bind.[82,131] Lastly, the currently available IgE Fab–allergen complex structures define important atomic-level features of IgE recognition[2,12,16,17,73], but these interface features have not been systematically applied to compare candidate IgE-binding surfaces across food, respiratory, and human lipocalins.

1.2 Study rationale, goal, and objectives

1.2.1 Rationale

The IgE Fab–Bos d 5 complex provides an experimentally determined interface for structural analysis of the lipocalin protein family.[17] These proteins are suitable for comparison because they share a common fold while differing in exposed residues and in the available evidence for IgE recognition.[1,61,69,80,84,110,131,168,169] The relevant question is whether these differences are reflected in mapped Bos d 5 linear IgE-binding regions and comparative models constructed using its Fab.

This thesis develops and applies a workflow with two linked analyses. The epitope analysis workflow maps Bos d 5 linear IgE-binding regions, as reported in the literature, and the Bos d 5 conformational epitope onto aligned query proteins, after which structural and physicochemical descriptors are calculated. The interface analysis involves positioning each query protein in the orientation of the template Fab–allergen and then computing the descriptors for the modelled interface as well as those for the epitope and paratope defined by contact cutoff. The key aspect of this study is the combination and application of these analyses to a group of lipocalin proteins consisting of β-lactoglobulin homologues, aeroallergens, and human lipocalins and to prioritize candidate proteins and surface regions for subsequent experimental evaluation.

1.2.2 Goal

The immediate goal is to develop and apply a comparative structural workflow to determine how closely β-lactoglobulin homologues, aeroallergens, and human lipocalins retain the measured epitope, paratope and interface features of Bos d 5. The long term goal is to use these comparisons to guide experimental investigation of IgE cross-reactivity between food and respiratory allergens.

1.2.3 Objectives

1. Develop and evaluate a general comparative structural workflow
   1.1. Develop a reproducible comparative structural workflow that (i) maps reference allergen epitope regions onto structurally aligned homologous proteins and computes standardized structural and physicochemical descriptors for the reference and mapped regions, and (ii) uses an experimentally determined IgE Fab–allergen complex as a structural template to generate comparative IgE Fab–allergen models and compute corresponding epitope, paratope, and interface descriptors for the reference and comparative complexes.
   1.2. Evaluate the robustness, interpretability, scope, and limitations of this workflow.
2. Apply the workflow to Bos d 5 and homologous lipocalins
   2.1. Assess whether lipocalin aeroallergens exhibit epitope-region and IgE Fab–allergen interface descriptors similar to those of Bos d 5, consistent with structural features that could support IgE cross-reactivity.
   2.2. Assess whether non-bovine ruminant β-lactoglobulin homologues exhibit the greatest similarity to Bos d 5 in epitope-region and comparative IgE Fab–allergen interface descriptors, consistent with evolutionary conservation within the ruminant β-lactoglobulin subgroup.
   2.3. Evaluate whether endogenous human lipocalins serve as tolerated comparators by exhibiting lower similarity to Bos d 5 epitope-region and comparative IgE Fab–allergen interface descriptors despite conservation of the lipocalin fold.

2 METHODS

2.1 Structural dataset assembly and preprocessing

2.1.1 Structural dataset

Structural data were assembled in three steps. Lipocalin allergens and their isoallergen-level UniProt accessions[170] were collected from the WHO/IUIS Allergen Nomenclature database.[143] Bos d 5 was included because its WHO/IUIS entry is listed as β-lactoglobulin, a member of the lipocalin family.[69,80,84,110] Additional β-lactoglobulin homologs representing the non-bovine ruminant, non-ruminant ungulate, carnivoran and marsupial comparison groups were retrieved from UniProt through the UniProt REST API.[170] Primary UniProtKB accessions were used as the stable protein identifiers throughout the study.

One representative structure was selected for each protein accession. Candidate experimental structures were identified from UniProtKB[170] and records in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB).[171] Models from the AlphaFold Protein Structure Database (AFDB) were considered only when no suitable experimental structure was available.[172,173] When multiple experimental structures were available for a given accession, the representative structure was selected by prioritizing experimental entries and then considering experimental method, sequence coverage relative to the UniProt entry, chain completeness, and resolution.[174] In parallel, a comparison set of endogenous human lipocalins was assembled from reviewed human UniProtKB entries and screened using the same structure-selection criteria. Experimental IgE Fab–allergen complexes were identified in the RCSB PDB[171] using the

curated allergen list. Candidate complexes were retained only when the antigen chain corresponded to a curated allergen, the entry represented an antibody–allergen complex, and both Fab heavy- and light-chain components were present. Unless otherwise stated, only Fab fragments derived from human IgE antibodies were retained.

The WHO/IUIS Allergen Nomenclature Database and RCSB PDB records were last searched on 15 June 2026. Sequence records and annotations corresponded to UniProtKB release 2026_02, accessed on 15 June 2026, and predicted structures were downloaded from AlphaFold Protein Structure Database (AFDB), accessed on 15 June 2026. The final dataset is summarized in Table 3-1–Table 3-4.

2.1.2 Structural dataset preprocessing

Representative structures were preprocessed to standardize protein chains prior to analysis. For AFDB[172,173] models, structures were preprocessed before use by defining mature-chain boundaries from UniProtKB[170] feature annotations, and then trimming the models with Biopython[175] to those boundaries while removing low-confidence regions based on per-residue AlphaFold predicted local distance difference test (pLDDT) scores.[173]

2.2 Workflow development

2.2.1 Method overview

We developed a comparative structural method comprising two workflows for analyzing homologous proteins: epitope mapping and IgE Fab–antigen complex workflow (Figure 2-1; Figure 2-2). The epitope-mapping workflow used a reference allergen with annotated epitope regions and structurally aligned homologous query proteins as inputs and generated mapped epitope regions with corresponding structural and physicochemical descriptors as outputs. The IgE Fab–allergen workflow used an experimentally

determined IgE Fab–allergen complex and homologous query proteins as inputs, and generated comparative IgE Fab–query allergen models with corresponding epitope, paratope, and interface descriptors as outputs. The workflow was implemented in Python[176] through the YASARA Python[177] interface, enabling automated structural alignment, complex construction, and downstream descriptor extraction.

2.2.2 Structure loading and cleanup

Structures were loaded in YASARA (Yet Another Scientific Artificial Reality Application) with the original coordinate frame preserved by disabling the centering parameter. Before structural alignment and descriptor calculation, crystallographic waters and other non-protein molecules were removed so that subsequent measurements were restricted to protein–protein interactions.[178,179] During complex construction, structures were imported with YASARA's correction option enabled to regularize coordinate-file inconsistencies.[177] For descriptor extraction from already prepared complexes, this correction step was not repeated. Hydrogen atoms were added and protonation states were assigned consistently at pH 7.4 before calculation of descriptors. This standardization is important since it is necessary to define the protonation states at in order to carry out electrostatic calculations[79,180], and the addition and optimization of hydrogen atom positions are necessary to determine reliable interaction geometries.[78,181]

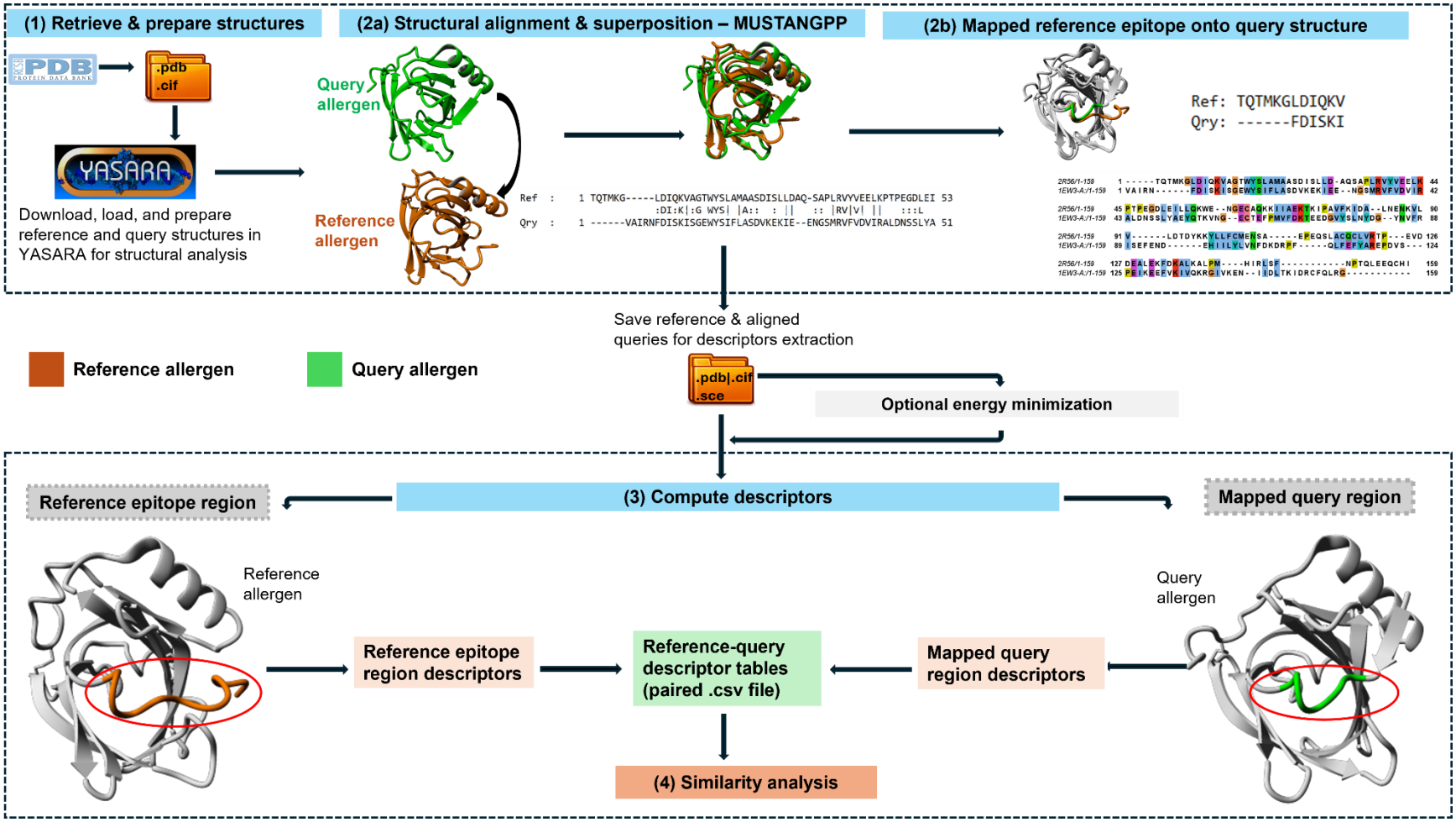

Figure 2-1. Overview of the epitope mapping workflow.

(1) Reference and query structures are retrieved and prepared. (2a) Each query protein is structurally aligned and superposed onto the reference allergen. (2b) Residue correspondence from the structural alignment is used to map reference epitope onto the query structure. (3) Structural and physicochemical descriptors are calculated for the reference epitope region and for each corresponding mapped query region. (4) Paired reference–query descriptor tables are used for similarity analysis.

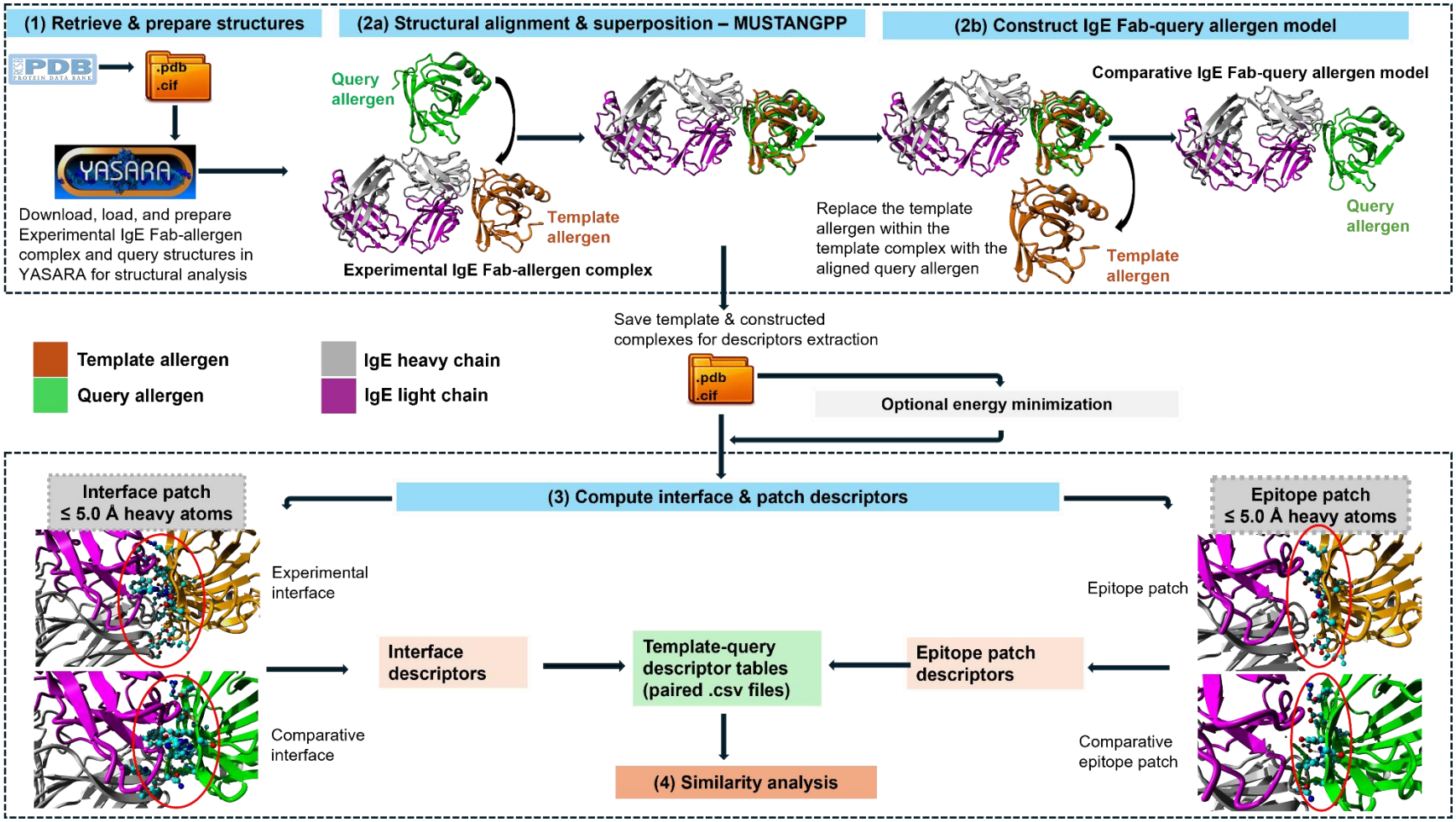

Figure 2-2. Overview of the template-based comparative structural workflow.

(1) Structures are retrieved and prepared. An experimentally determined IgE Fab–allergen complex is selected as the template complex, and the bound allergen chain is defined as the template allergen. (2a) Each query allergen is structurally aligned and superposed onto the template allergen. (2b) The fitted query allergen replaces the template allergen in the coordinate frame of the bound Fab, generating a template derived IgE Fab–query allergen model. (3) Patch and interface descriptors are calculated for the experimental template complex and for each template derived model. (4) The resulting descriptor tables are used for paired template–query similarity analysis.

2.2.3 Representative assembly selection

Some of the experimentally determined IgE Fab–allergen complexes contained more than one copy of the complex within the asymmetric unit (ASU).[160] In such cases, a single Fab–allergen assembly was identified from the protein-chain composition and heavy-atom contact topology using a 5.0 Å cutoff. This distance threshold has been used in past studies to identify noncovalent interactions and to define interfacial residues.[72,119] To avoid redundancy within the PDB entry[182], only one assembly was retained.

2.2.4 Fab heavy-chain, light-chain, and allergen-chain assignment

Chain roles were assigned from chain length and contact topology to address the inconsistent annotation of chain identities and numbering often found in PDB entries.[183] Antibody-like chains were defined as chains of 201–350 residues, whereas non-antibody-like chains were classified as candidate allergen chains.[184] In three-chain complexes, if exactly two chains met the antibody-length criterion, those chains were assigned as the Fab pair and the remaining chain as the allergen; otherwise, Fab and allergen roles were assigned from contact topology. When additional protein chains were present, the Fab pair was defined as the contacting pair of antibody-length chains showing the strongest mutual contact. Within the Fab pair, the chain forming the greater number of heavy-atom contacts with the allergen was assigned as the heavy chain and the other as the light chain. This choice of this approach is informed by structural studies showing that the heavy chain often contributes a larger fraction of antibody–antigen contacts and buried surface area than the light chain.[13,67,178,185]

2.2.5 Structural alignment, epitope mapping, and Fab–query allergen model construction

Structural alignment was used for two related purposes: mapping reference epitope annotations onto homologous query structures and constructing Fab–query allergen models from experimentally determined IgE Fab–allergen complexes.

For epitope mapping, each query allergen was structurally aligned to the reference allergen using YASARA's implementation of the Multiple Structural Alignment Algorithm (MUSTANG) developed by Konagurthu et al.[177,186] Unless otherwise stated, the following parameters were used: maximum distance cutoff, 5.0 Å; maximum angle cutoff, 90°; minimum aligned segment length, 5 residues; gap-opening penalty, 10; gap-extension penalty, 2; and terminal overhang, 0. Residue correspondence from the structural alignment was used to identify query residues aligned to the reference epitope positions. These aligned query residues formed the mapped query regions used for subsequent descriptor calculation.

For interface analysis, each query allergen was structurally aligned to the template allergen from an experimentally determined IgE Fab–allergen complex using the same MUSTANGPP procedure.[177,186] Alignment quality was recorded using root-mean-square deviation (RMSD), sequence identity, and the number of aligned Cα residue pairs. These metrics were used to quantify the structural and sequence similarity between the template allergen and each query allergen.

The resulting rigid-body superposition was then used to position the query allergen in the template-defined binding position. The template allergen was replaced *in situ* by the superposed query allergen within the experimental IgE Fab–allergen complex, thereby generating a constructed IgE Fab–query allergen complex. This procedure generated a comparative model derived from the template by placing the query allergen in the

paratope context of the reference Fab. The approach is analogous to template-based protein–protein complex modeling, in which an experimentally determined complex provides the relative binding orientation for comparative analysis or model construction.[7,187–189] The resulting constructed complexes were saved for subsequent descriptor extraction and comparative analysis.

The curated IgE Fab–allergen complexes (Table 3-1) were used to evaluate the template-based workflow. As identity controls, each template complex was also reprocessed using its own allergen chain as the query structure. These controls were used to determine whether structural alignment, complex construction, and descriptor extraction altered the measured interface properties in the absence of allergen substitution.

2.2.6 Two-protocol workflow

Two descriptor-calculation protocols were implemented. In the RAW protocol, interface descriptors were calculated directly from the protein structures after standardized structural preparation in YASARA.[177] This preparation included protonation-state assignment at pH 7.4, hydrogen-bond optimization, AMBER14[190] force-field assignment, construction of a periodic simulation cell, and treatment of long-range Coulomb interactions using the Particle Mesh Ewald (PME) algorithm.[191]

In the MIN protocol, the same prepared protein structures were subjected to an additional energy-minimization step before descriptor calculation. Minimization was performed after solvation (explicit TIP3P[192] water molecules) and charge-neutralization of the periodic cell, with protonation states assigned during the preparation stage at pH 7.4. The minimization procedure was used to reduce steric strain introduced during template-based structural replacement before descriptor calculation.[178,180,193]

Descriptor outputs were retained according to protocol-completion and quality-control criteria. For each attempted template-complex/query-allergen comparison, a descriptor row was included only when the relevant protocol completed successfully and produced usable descriptor outputs. For paired RAW–MIN analyses, the matched analysis set was defined a *priori* as the subset of comparisons for which both RAW descriptors and MIN descriptors were available. Coverage was summarized at two levels: protocol-specific coverage and matched RAW–MIN coverage.

2.2.7 Regions and interface definition

Mapped epitope regions and comparative interface regions were defined using structural correspondence derived from structural alignment or atomic contact criteria.

For the epitope-mapping workflow, mapped query regions were defined by residue correspondence obtained from structural alignment between the reference allergen and each homologous query allergen. Query residues aligned to the reference epitope positions constituted the mapped query region.

For each experimental IgE Fab–allergen template complex and each comparative Fab–query allergen model, contact residues were defined using a 5.0 Å heavy-atom cutoff. A residue was assigned to a contact patch when at least one of its heavy atoms was within 5.0 Å of any heavy atom of the binding partner.[67,119,194,195] Hydrogen atoms were excluded from patch definition. This criterion was applied uniformly to all experimental template complexes and comparative models.

In the experimental IgE Fab–allergen template complex, allergen residues that contacted the Fab defined the epitope, whereas Fab residues satisfying the same criterion defined the paratope patch. Together, these residue sets defined the experimental Fab–allergen interface. In the comparative Fab–query allergen models, query allergen residues within

5.0 Å of the Fab defined the comparative epitope, whereas Fab residues satisfying the same criterion defined the comparative paratope. These residue sets defined the comparative Fab–query allergen interface generated under the template binding geometry.

A 5.0 Å heavy-atom cutoff is a standard structural criterion for defining antibody–antigen contact residues and related protein–protein interface contacts.[72,125,196] The same cutoff was used for all template and query comparisons to allow direct comparison of patch and interface descriptors.

Because the method development stage included structurally diverse template–query pairings, some comparative models produced no query allergen residues within the selected contact distance after structural superposition. In these cases, no comparative epitope or comparative interface could be defined, and the corresponding models were excluded from downstream descriptor calculation.

2.2.8 Region and interface descriptors

Standardized structural and physicochemical descriptors were calculated for every structural region defined by the workflow. Descriptor definitions were specified *a priori* and applied uniformly to all experimental structures and comparative models processed under the RAW and MIN protocols.

For the epitope-mapping workflow, descriptors were calculated for both the reference epitope region and the corresponding mapped query region identified by structural alignment.

For the comparative interface workflow, region descriptors were calculated separately for the epitope and paratope in both the experimental template complexes and the

comparative Fab–query allergen models. Interface descriptors were calculated for the corresponding Fab–allergen and comparative Fab–query allergen interfaces.

Region descriptors characterized residue composition and structural and physicochemical properties of the defined regions, whereas interface descriptors quantified contacts and interaction properties across the antibody–antigen interface. Together, the region and interface descriptors provided a standardized basis for comparing mapped epitope regions and comparative antibody–antigen binding sites across homologous proteins.

2.2.9 Structural and physicochemical descriptor extraction

Structural and physicochemical descriptors were extracted from mapped epitope regions and comparative antibody–antigen binding sites generated by the workflow. Descriptor calculations were performed using YASARA[177], in Biopython[175], and custom Python, as described below.

2.2.9.1 Hydrogen bonds

Hydrogen bonds were identified using YASARA's atom-level listing function (ListHBoAtom) with a minimum energy threshold of 6.25 kJ/mol. Hydrogen-bond descriptors included total hydrogen-bond count, total and mean interaction energy, mean donor–acceptor distance, and interaction directionality based on whether the allergen residue acted as the hydrogen donor or acceptor

2.2.9.2 Hydrophobic, ionic, π–π, and cation–π interactions

Non-hydrogen-bond interactions were enumerated using YASARA's ListIntAtom function. Hydrophobic contacts were identified using a 5.0 Å cutoff, consistent with graph-based clustering methods used to map hydrophobic patches in antibody–antigen interfaces.[119] Ionic interactions (salt bridges) were defined using a stricter 4.0 Å cutoff,

aligning with standard geometric criteria for oppositely charged side-chain functional groups.[79] Interactions involving aromatic rings were calculated with a 5.0 Å cutoff for both π-π and cation-π interactions, a threshold previously validated for characterizing ring-stacking and charge-ring interactions in immune complexes.[77,119]. To ensure only intermolecular contacts were counted, neighbor atoms connected by covalent bonds were excluded, and a hydrophobic occlusion filter was disabled by default to ignore hydrophobic interactions that are obstructed by other atoms.[177]

2.2.9.3 Buried surface area

Interface size was quantified as buried solvent-accessible surface area, calculated from the loss of solvent-accessible surface area upon complex formation using a 1.4 Å probe radius. Solvent-accessible surface area was measured with YASARA surface calculations implementing an accessible-surface definition based on the Lee and Richards algorithm.[197] For each complex, solvent-accessible surface area (SASA) was first measured for the allergen and the Fab in isolation and then remeasured in the full complex environment. Buried surface area (BSA) was calculated as

$$\Delta SASA = \left(SASA_{allergen,\ isolated} + SASA_{Fab,\ isolated}\right) - SASA_{complex} \qquad \text{Equation 2-1}$$

Solvent-accessible area and buried area are standard structural summaries for protein interfaces.[125,198,199]

2.2.9.4 Aromatic residue enrichment descriptors

Aromatic residues are frequently prominent at antibody-binding interfaces, making aromatic enrichment a useful descriptive feature of interface composition.[77,119,126,200,201] Aromatic residues (Tyr, Trp, Phe, and His) make major contributions to antibody–antigen recognition by supporting hydrophobic, hydrogen-bonding, cation–π, and π–π interactions.[75,77,126] Previous studies on structural analyses of antibody–antigen

interfaces showed that aromatic residues are commonly found at the binding sites, although enrichment on the antigen side is generally more modest than on the antibody side.[16,76]

Aromatic residue enrichment was calculated to quantify whether interface-associated residue patches were compositionally enriched in aromatic residues relative to their corresponding parent structural regions. A residue was classified as aromatic if it contained at least one atom matching YASARA's *Aromatic* atom selector. For any patch P, the aromatic-residue fraction was defined as

$$f_P = \frac{n_{\mathrm{arom}}^P}{n_{\mathrm{tot}}^P} \qquad \text{Equation 2-2}$$

where n_{arom}^P is the number of aromatic residues in the patch and n_{tot}^P is the total number of residues in the patch. For the corresponding background region B, the aromatic-residue fraction was defined as

$$f_B = \frac{n_{\mathrm{arom}}^B}{n_{\mathrm{tot}}^B} \qquad \text{Equation 2-3}$$

where n_{arom}^B and n_{tot}^B are the numbers of aromatic and total residues in the relevant full background region, respectively. Local aromatic enrichment was then calculated as

$$E_{\mathrm{arom}}^{\mathrm{local}} = \frac{f_P}{f_B} \qquad \text{Equation 2-4}$$

where B could correspond to the full allergen chain or full Fab, depending on the patch being analyzed. Values greater than 1 indicate enrichment of aromatic residues within the defined region (epitope, paratope, or interface) relative to its background (whole allergen or Fab), whereas values below 1 indicate depletion.

In addition to these local patch descriptors, a two-sided interface aromatic enrichment descriptor was also calculated. Interface residues were identified independently on both interacting partners using a heavy-atom contact criterion. The observed number of aromatic interface residues across both partners was then compared with the number expected from the aromatic-residue fractions of the corresponding background regions. This yielded an interface-level aromatic enrichment descriptor that reflects whether the combined interacting surface is more aromatic than expected from the compositions of the two partners.

This metric is conceptually analogous to the occurrence-propensity approach of Ramaraj et al.[125], in that it evaluates whether aromatic residues are overrepresented in the Fab-contacting patch relative to a broader compositional background. In this study, however, the background reference was the full allergen chain or Fab chains.

2.2.9.5 Sequence-derived physicochemical descriptors

Molecular weight, theoretical isoelectric point, and grand average of hydropathy (GRAVY) were calculated from the amino acid composition of the Fab-facing allergen patch using the *ProteinAnalysis* module in Biopython. Biopython.[175] These descriptors were retained as compact summaries of patch composition and physicochemical properties for downstream analyses. Such properties have been used as informative features in epitope analysis and prediction, although conformational epitopes are often only modestly distinguishable from the surrounding non-epitope protein surface.[202]

2.2.9.6 Residue-composition descriptors

Patch residue composition for each contact patch was quantified as the number of hydrophobic, polar, positively charged, negatively charged, total charged, and glycine residues, using the residue-class definitions implemented in the YASARA selection

grammar. These descriptors were used to represent the local physicochemical composition of the patch, which is relevant to antibody recognition and potential allergen cross-reactivity at exposed epitope surfaces.[67] Glycine count was included as an auxiliary compositional descriptor because glycine lacks a side chain beyond hydrogen and is associated with increased local backbone flexibility.[203]

2.2.9.7 Relative solvent accessibility

Patch solvent-accessible surface area (SASA) was calculated using a 1.4 Å probe radius in YASARA.[177] Relative solvent accessibility was then obtained by normalizing residue solvent-accessible areas to amino-acid-specific maximal accessible surface areas. This normalization follows standard RSA practice, in which solvent accessibility is expressed relative to a residue-specific reference state, commonly based on Gly–X–Gly tripeptide conformations.[76,124,204,205]

If ASA_i is the solvent-accessible area of residue i and ASA_i^{max} is the corresponding reference maximum, patch RSA was calculated as

$$RSA_{patch} = \frac{\sum_{i=1}^{N} ASA_i}{\sum_{i=1}^{N} ASA_i^{max}} \qquad \text{Equation 2-5}$$

where the sum was restricted to residues with defined reference maximum values reported by Tien et al.[124] This yielded a normalized measure of solvent exposure for the Fab-facing allergen patch.

2.2.9.8 Electrostatic surface potential

Electrostatic surface potential (ESP) of a patch was calculated in YASARA after standardized structural preparation, including cleanup. Surface-analysis settings were

kept constant across all complexes, and the full molecular environment was retained during calculation. ESP was then evaluated on the solvent-accessible surface of the selected patch using YASARA's SurfESP functionality with the particle mesh Ewald (PME) [191] electrostatics method. In the present workflow, the patch-level ESP descriptor was recorded as YASARA's software-reported unit value for the selected residue set, corresponding in YASARA to the average energy, in the current energy unit, of a hypothetical +1 charge at the surface positions contributed by that unit.

2.2.9.9 Geometric protrusion

Geometric protrusion was quantified using an in-house ellipsoid-based protrusion index (PI) adapted from the protrusion framework of Thornton[122] and the later ElliPro[99] implementation. After centering the allergen atomic coordinates, an ellipsoid was fitted to the full allergen chain by singular value decomposition.

For a point with transformed coordinates $(x_j', y_j', z_j')$and ellipsoid semi-axes (a, b, c), the normalized ellipsoidal distance was calculated as

$$d_j^2 = \left(\frac{x_j'}{a}\right)^2 + \left(\frac{y_j'}{b}\right)^2 + \left(\frac{z_j'}{c}\right)^2 \qquad \text{Equation 2-6}$$

For each residue r, protrusion was then defined as the fraction of allergen atoms whose normalized ellipsoidal distance was less than or equal to that of the residue center of mass:

$$PI_r = \frac{\#\{k: d_k^2 \leq d_r^2\}}{N_{atoms}} \qquad \text{Equation 2-7}$$

The patch-level protrusion index was reported as the average of residue-level protrusion indices across all residues in the Fab-facing allergen patch:

$$PI_{patch} = \frac{1}{N_{patch}} \sum_{r=1}^{N_{patch}} PI_r \qquad \text{Equation 2-8}$$

This yielded a dimensionless measure of the extent to which the patch occupied an outward-projecting position relative to the full allergen structure.

2.2.10 Software, outputs, reproducibility, and computational environment

All structural processing, structural alignment, energy minimization, and descriptor calculations were performed in YASARA Structure (version 25.12.1) through its Python interface.[176,177] The workflow was implemented in Python 3, with sequence-derived physicochemical descriptors calculated using Biopython[175,206] and tabular outputs generated using NumPy[207] and Pandas.[208]

Structural outputs were saved as Protein Data Bank (PDB) or Crystallographic Information File (CIF) files. Descriptor outputs were written as comma-separated value (CSV) files (Supplementary Tables S1–S6), with each file corresponding to one template-complex/query-protein comparison. Reproducibility was ensured by applying identical alignment parameters, descriptor definitions, and interaction criteria throughout the workflow.

The workflow was developed under Python 3.13.2 and executed on the Fir cluster, a Slurm-managed high-performance computing system of the Digital Research Alliance of Canada. All analyses and runtime measurements were performed on AlmaLinux 9.6 compute nodes equipped with AMD EPYC 9655 96-core processors (2 sockets, 192 logical CPUs, and one thread per core). Workflow runs were submitted as Slurm array jobs and executed as one task per job using four CPU cores and 4 GB of memory. Runtime was recorded as the total elapsed wall-clock time for each workflow execution

(Figure 3-7). The workflow developed in this study is available on GitHub at https://github.com/isp5lu/comparative-structural-workflow.

2.3 Workflow validation

The workflow validation analyses described in this section were restricted to the template derived Fab–query allergen models generated by the complex construction workflow. These analyses were designed to assess whether the template based workflow could reproducibly generate and quantify comparative Fab–query allergen interfaces. The validation therefore focused on the consistency and interpretability of structural replacement, optional energy minimization, and descriptor extraction.

2.3.1 Validation descriptor set

Before descriptor reduction and similarity analysis, a subset of interface descriptors was selected to evaluate the behavior of the comparative Fab–query allergen modeling workflow. The selected descriptors were chosen to represent interface size, buried surface area, aromatic composition, and the major classes of non-covalent interactions across the antibody–allergen interface. These descriptors cover the main physical aspects of the antibody–allergen interface and were expected to be sensitive to structural replacement, definition of the contact patch, and energy minimization. The selected validation descriptors were used only to assess workflow behavior, identity-control recovery, and RAW–MIN protocol effects (Figure 3-1 - Figure 3-3). Full interface descriptor networks are provided in Supplementary Figure S3. .

2.3.2 Identity controls

Each template complex was reprocessed using its own allergen chain as the query structure. These identity controls were used to determine whether structural alignment,

IgE Fab-allergen complex construction, and descriptor extraction altered the measured interface when the template allergen is used as its own query structure (Figure 3-1).

2.3.3 Workflow protocols comparison

The RAW and MIN protocols were compared only for matched template-complex/query-allergen comparisons that completed successfully under both protocols. For each descriptor f and paired comparison i, the RAW value was denoted as $x_i = x_{i,f}^{RAW}$ and the MIN value was denoted as $y_i = x_{i,f}^{MIN}$. The paired protocol difference was calculated as

$$\Delta_{i,f}^{MIN-RAW} = y_i - x_i \qquad \text{Equation 2-9}$$

such that $\Delta_i > 0$ indicates an increase after minimization and $\Delta_i < 0$ indicates a decrease. Because several descriptor distributions were skewed, protocol differences were summarized using number of paired comparisons, the median paired difference, the interquartile range of paired differences, the mean absolute paired difference, and Spearman's rank correlation (ρ) between RAW and MIN values. Rank preservation was evaluated using Spearman rank correlation, which measures monotonic association without requiring a linear relationship and is robust to non-normality and outliers.[209] Spearman's ρ was computed by ranking RAW and MIN values independently,

$$R_i = \mathrm{rank}(x_i), S_i = \mathrm{rank}(y_i) \qquad \text{Equation 2-10}$$

and then calculating the Pearson correlation of the ranked values,

$$\rho = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}}, \qquad \text{Equation 2-11}$$

Where $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i$ and $\bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$ are the mean RAW and MIN ranks, respectively.

This quantified whether complexes with higher RAW values also tended to retain higher MIN values after minimization, independent of linear proportionality. For visualization, paired RAW–MIN values were displayed in multi-panel scatter plots with the identity line ($y = x$), and panel summaries included the number of finite paired observations (n), the median signed difference,

$$\tilde{\Delta} = \mathrm{median}(\Delta_i), \qquad \text{Equation 2-12}$$

the interquartile range,

$$\mathrm{IQR} = Q_3(\Delta_i) - Q_1(\Delta_i), \qquad \text{Equation 2-13}$$

the number of outliers,

$$n_{\mathrm{out}} = \sum_{i=1}^{n} 1\,[\Delta_i < Q_1 - 1.5 \times \mathrm{IQR} \text{ or } \Delta_i > Q_3 + 1.5 \times \mathrm{IQR}], \qquad \text{Equation 2-14}$$

and Spearman's rank correlation coefficient (ρ) (Figure 3-2). Distributions of Δ_i were additionally visualized using boxplots with a horizontal reference line at $\Delta = 0$ to indicate no change after minimization; boxplot summaries included the median, IQR, non-outlier whisker range, and the number of outliers for each descriptor (Figure 3-3).

2.3.4 Model generation and matched RAW–MIN comparison set

Output was assessed to determine how consistently the comparative Fab–query allergen workflow produced completed descriptor rows from the intended input design (Figure 3-5). The input design consisted of the experimental IgE Fab–allergen template complexes paired with the lipocalin query structures. Each attempted comparison corresponded to one template complex and one query structure.

RAW and MIN resulting outputs were summarized separately for each template complex. Failed model generation was recorded at the template-complex/query-allergen level and

classified according to the workflow step responsible for loss of output. For the MIN protocol, unavailable descriptor records were attributed to cases in which energy minimization failed to produce a converged structure. Paired RAW–MIN analyses were restricted to a matched comparison set. A comparison was included in this matched set only if descriptor outputs were available for the same template-complex/query-allergen pair under both RAW and MIN protocols. Comparisons missing from either protocol were excluded from paired RAW–MIN analyses.

2.3.5 Structural alignment quality and fold compatibility

Structural alignment quality was assessed to determine whether each query allergen could be reliably positioned within the binding orientation defined by each experimental IgE Fab–allergen complex.

For each template–query comparison, the alignment output was first checked to determine whether residue-level structural correspondence had been established between the template allergen and the query allergen. A comparison was considered to meet the structural-alignment inclusion criterion when the alignment returned at least one aligned Cα residue pair. Comparisons that reported zero RMSD together with zero aligned Cα pairs were excluded from the alignment-quality summary, because no residue-level structural correspondence was established and the RMSD value was therefore not interpretable. For each template complex and protocol, the number and percentage of available comparisons meeting this criterion were reported. Among comparisons that met the criterion, alignment quality was summarized using Cα RMSD, the number of aligned Cα pairs, and sequence identity, reported as the median and interquartile range. MIN comparisons for which energy minimization failed to converge were excluded from both the percentage calculation and the alignment-quality

summaries because they did not produce usable workflow outputs (Figure 3-6, Table 3-6).

2.4 Workflow application

2.4.1 Study design and analysis overview

The structural comparison workflow described above was applied to compare Bos d 5 with a set of query lipocalins. The analysis evaluated the extent to which query proteins preserved structural and physicochemical descriptors observed in Bos d 5 IgE binding regions and in the Bos d 5–IgE Fab interface. Descriptor similarity was used as an indicator of resemblance to Bos d 5 epitope region or interface features and as the basis for prioritizing candidates for experimental IgE-binding or inhibition studies.

2.4.2 Biological comparison groups

Query lipocalins were assigned a *priori* to biologically defined comparison groups before statistical analysis (Table 3-7). Bos d 5 (bovine βLG) was used as the reference allergen because it is an established cow's-milk food allergen (WHO/IUIS[143]) and the only milk protein with an experimentally resolved IgE complex.[17]

Non-bovine ruminant βLGs were used as positive controls (or close homologous comparators) because they are expected to retain high structural similarity within the ruminant β-lactoglobulin subgroup. Milk cross-reactivity is well documented, including partial direct inhibition for bovine versus reindeer βLG and broad cow–goat/sheep/buffalo cross-recognition in inhibition and immunoblot studies.[47,142,210–212] Non-ruminant ungulate βLGs were included as intermediate lineage comparators, preserving βLG identity while increasing phylogenetic distance from Bos d 5. Carnivoran and marsupial βLGs served as distant βLG homologs to test whether retention of the βLG scaffold alone is sufficient to maintain compatibility with the binding interface of the template Fab.

Respiratory lipocalin allergens (aeroallergens) constituted the primary hypothesis group because lipocalins comprise many major mammalian inhalant allergens linked to rhinitis

and asthma.[1,6,127,131,168,213,214] Endogenous human lipocalins were included as tolerated comparators because they share the lipocalin fold but are not typical targets of clinically relevant IgE responses. Human lipocalin-1, LCN1, has been used as a homologous human comparator for animal lipocalin allergens.[168,215]

2.4.3 Bos d 5 reference regions and Fab–allergen interface

2.4.3.1 Literature-reported linear IgE-binding regions

Bos d 5 linear IgE-binding regions were curated from published epitope-mapping studies and used as reference linear epitope annotations. Because some reported regions overlapped or were fully contained within larger regions, a minimally redundant region set was defined for the primary analysis (Table 3-8, Figure 3-8). Regions completely contained within longer reported regions were retained for supplementary analyses but excluded from the primary linear-region dataset (Table S3. 8).

2.4.3.2 Bos d 5 conformational epitope defined by Fab contact residues

The Bos d 5 conformational epitope was defined using the 2R56 IgE Fab–Bos d 5 complex. Epitope residues reported in the original structural study were retained for annotation purposes (Figure 3-9, Table 3-9). For descriptor extraction, the Bos d 5 epitope was defined using the standardized heavy atom contact criterion applied throughout the workflow. The epitope therefore consisted of Bos d 5 residues with at least one heavy atom within the defined contact cutoff of any Fab heavy atom in the experimental complex. Because structural preparation and energy minimization could alter residue contacts, RAW and MIN epitope residue sets were defined and analyzed separately where applicable.

2.4.3.3 Bos d 5–IgE Fab interface used as the template framework

The experimentally determined IgE Fab–Bos d 5 complex (PDB: 2R56) was used as the template for comparative Fab–query allergen modelling. The Bos d 5 allergen served as the template antigen, and the bound IgE Fab defined the reference antibody orientation. Comparative Fab–query allergen models were generated using the workflow described in Section 2.2.5, and region and interface descriptors were subsequently calculated according to Sections 2.2.7 and 2.2.8.

2.4.4 Application datasets and units of comparison

2.4.4.1 Linear and conformational epitope region mapping datasets

The epitope region mapping datasets contained one row for each query structure and each Bos d 5 reference region. Separate datasets were generated for the literature-reported linear IgE-binding regions and for the Bos d 5 conformational epitope defined from the Fab contact criterion.

For each query and reference region pair, the datasets included residue correspondence and mapping quality metrics, local sequence identity, local RMSD, and structural and physicochemical descriptors calculated for both the Bos d 5 reference region and the corresponding mapped query region.

2.4.4.2 Comparative Fab–allergen complex model datasets

The comparative Fab–query allergen model datasets contained one row for each query allergen positioned within the binding orientation defined by the Bos d 5–IgE Fab template complex.

For each comparative model, standardized descriptor sets were calculated separately for the comparative epitope, the comparative paratope, and the comparative Fab–query allergen interface. For analyses integrating all three structural regions, these descriptor

sets were combined into a composite descriptor set, representing the complete antibody–antigen binding site within the comparative model.

2.4.5 Mapping quality assessment and descriptor transformation for epitope region mapping datasets

Mapping quality was assessed for each Bos d 5 reference region and query protein pair using metrics generated by the epitope region mapping workflow (Figure 3-10 and Figure 3-11). These metrics included the number of aligned residues, reference region coverage, gap fraction, local sequence identity, and local RMSD. Coverage was defined as the proportion of residues in the Bos d 5 reference region with structurally aligned counterparts in the query structure.

For descriptor comparisons, mapped query region descriptors were transformed into Δ descriptors, representing descriptor differences relative to the corresponding structurally aligned Bos d 5 reference region. For each descriptor j, query allergen q, and Bos d 5 reference region r, the descriptor difference relative to the Bos d 5 was calculated as:

$$\Delta D_{q,r,j} = D_{q,r,j}^{mapped} - D_{ref,r,j}^{aligned} \qquad \text{Equation 2-15}$$

where $D_{q,r,j}^{mapped}$ is the descriptor value calculated from the mapped query residue patches, and $D_{ref,r,j}^{aligned}$ is the descriptor value calculated from the corresponding structurally aligned subset of the Bos d 5 reference region. Positive and negative Δ values therefore represent deviations from the matched Bos d 5 reference subset for the same descriptor.

Descriptors calculated from the complete Bos d 5 reference region were retained as contextual reference descriptors but were not used to calculate delta values. These full-reference descriptors were used only to describe the complete reference epitope region and were not included in reference-relative distance calculations unless otherwise stated.

Two epitope region mapping datasets were generated. The first dataset included Δ descriptors for all mapped regions, regardless of coverage, and was used to assess the effect of retaining lower coverage mappings. The second dataset retained only mappings with at least 60% structural coverage of the corresponding Bos d 5 reference region. This coverage filtered dataset was used as the primary dataset for distance comparisons and multivariate analyses.

For each Bos d 5 reference region, a matched Bos d 5 reference row was retained as the reference profile. In Δ descriptor space, this row represented a zero deviation profile and served as the reference anchor for standardization, visualization, and distance calculations. The minimally redundant set of reported linear IgE binding regions was used for the primary analysis.

2.4.6 Descriptor preprocessing and redundancy assessment

Descriptor redundancy was assessed separately for each analysis block because epitope region mapping datasets and comparative Fab–query allergen model datasets represented different structural units. Redundancy was assessed independently for linear IgE binding region descriptors, conformational epitope descriptors, and comparative composites. RAW and MIN datasets were assessed separately.

2.4.6.1 Correlation-based redundancy assessment

Pairwise descriptor associations were evaluated using Spearman's rank correlation coefficient implemented in SciPy. Spearman correlation was used because descriptor relationships were not assumed to be linear and because several descriptors were counts, bounded variables, or non-normally distributed. Spearman's coefficient measures monotonic association between two variables after conversion to ranks. It was calculated as the Pearson correlation coefficient between ranked descriptor values:

$$\rho_s(x,y) = \frac{\sum_{i=1}^{n}(R(x_i) - \bar{R}_x)\left(R(y_i) - \bar{R}_y\right)}{\sqrt{\sum_{i=1}^{n}(R(x_i) - \bar{R}_x)^2}\sqrt{\sum_{i=1}^{n}\left(R(y_i) - \bar{R}_y\right)^2}} \qquad \text{Equation 2-16}$$

where $R(x_i)$ and $R(y_i)$ are the ranks of descriptors x and y for observation i, $\bar{R}_x$ and $\bar{R}_y$ are the corresponding mean ranks, and n is the number of retained observations in the relevant dataset. Absolute correlations, $|\rho_s|$, were used to identify descriptor groups carrying similar information.

2.4.6.2 Descriptor clustering and representative descriptor selection

Descriptor relationships were represented as correlation networks.[216] Each node represented one descriptor, and an edge was drawn between two descriptors when the absolute Spearman correlation exceeded the selected threshold. Network construction and visualization were performed using NetworkX 3.6[217] and Matplotlib 3.10.5.[218]

Correlation networks were inspected to identify groups of redundant descriptors. Thresholds of the dimensionless Spearman rank coefficient |ρs| = 0.70–0.90 were used to examine broader redundancy structure, whereas |ρs| ≥ 0.90 was used to identify strongly correlated or near-duplicate descriptors. Within each correlated descriptor group, one representative descriptor was retained. Representative descriptors were selected based on biological interpretability, relevance to epitope or interface structure, and avoidance of repeated measurement of the same physical property (Figure S3. -S3.5).

This filtering step was used to reduce descriptor redundancy before standardization, PCA, and distance-based analyses. Correlation-based descriptor preselection is commonly used in molecular-descriptor and quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) workflows to reduce redundancy and improve interpretability of downstream models.[219,220]

2.4.7 Standardization

Descriptor values were standardized using the *StandardScaler* implementation in scikit-learn[221] prior to analysis because retained descriptors were measured on different numerical scales. Standardization was performed separately for each descriptor set and protocol, with RAW and MIN descriptor matrices processed independently.

For the primary analyses, descriptor means and standard deviations were estimated using only query structures. The Bos d 5 reference (bovine βLG) was excluded when fitting the standardization parameters. The resulting transformation was subsequently applied to all observations, including the Bos d 5 reference row, thereby placing both query structures and the reference profile in the same standardized descriptor space.

For each descriptor j, the non-reference query rows were used to estimate the mean μ_j and standard deviation s_j, and each value was transformed as:

$$z_{ij} = \frac{x_{ij} - \mu_j}{s_j} \qquad \text{Equation 2-17}$$

where μ_j and s_j were estimated from the query structures.

For analyses of mapped linear epitope regions (MLERs), standardization was performed separately for each Bos d 5 reference region. Feature means and standard deviations were estimated from the corresponding query regions, and the matched Bos d 5 reference region was transformed using the same parameters.

For mapped conformational epitope and composite datasets, standardization was performed separately for each descriptor set and protocol. RAW and MIN datasets were standardized independently.

2.4.8 Principal component analysis

2.4.8.1 PCA of reduced descriptor sets

Principal component analysis (PCA) was performed using scikit-learn[222] to examine the structure of the reduced descriptor matrices and to generate lower-dimensional representations for comparisons based on distance. PCA was applied separately to each descriptor set and protocol after descriptor standardization.

For exploratory analyses, scree plots and PCA biplots were examined to assess the variance captured by the retained descriptors and to identify the principal descriptor combinations contributing to variation among query structures (Figure 3-12-Figure 3-14; Table 3-11-Table 3-12; Figures S3.6-3.26). PCA was used as a diagnostic tool to evaluate whether the reduced descriptor sets retained interpretable patterns of variation after removal of highly redundant descriptors.

2.4.8.2 PCA model fitting and component retention

PCA models were fitted separately for each descriptor set, including mapped epitope region descriptors, comparative epitope descriptors, paratope descriptors, and interface descriptors.

For each descriptor set, the PCA model was fitted using only the standardized query structures. The Bos d 5 reference row was excluded during model fitting and subsequently projected into the fitted PCA space. Consequently, both the standardization parameters and PCA coordinate system were defined by the query structures, while Bos d 5 served as an external reference.

PCA was performed with whitening enabled. In scikit-learn[222], whitening produces uncorrelated component scores with unit variance and was used to ensure that Euclidean

distance calculations were not dominated by principal components explaining larger proportions of variance.

For each descriptor set b, the number of retained principal components, k_b, was selected as the smallest number of components required to reach a cumulative explained-variance threshold of 95%:

$$k_b = \min\left\{m: \sum_{i=1}^{m} r_{b,i} \geq 0.95\right\} \qquad \text{Equation 2-18}$$

where $r_{b,i}$ is the explained-variance ratio of principal component i in descriptor set b (Table 3-12).

2.4.8.3 Euclidean distance calculations based on PCA

Distances between each query structure and the Bos d 5 reference were calculated separately for each descriptor set using the retained whitened PCA scores.

For each query structure q, the PCA distance to the matched bovine βLG reference was calculated using the first k_b whitened PCA scores:

$$d_{b,\mathrm{PCA}}(q, \text{bovine } \beta\mathrm{LG}) = \| \mathbf{t}_{q,b,1:k_b} - \mathbf{t}_{\text{bovine } \beta\mathrm{LG},b,1:k_b} \|_2 \qquad \text{Equation 2-19}$$

where $\mathbf{t}_{q,b,1:k_b}$ is the retained whitened PCA-score vector for query structure q in set b, and $\mathbf{t}_{\text{bovine } \beta\mathrm{LG},b,1:k_b}$ is the corresponding score vector for the bovine βLG reference.

Euclidean distance in PCA score space was used as a measure of descriptor-space similarity, with smaller distances indicating greater similarity to the Bos d 5 reference. The mapped epitope region distance was analyzed separately because it was derived solely from allergen structural descriptors and did not incorporate Fab–allergen contact information.

2.4.8.4 Visualization and downstream analyses

PCA score plots and biplots were used for visualization of relationships in descriptor space among query structures and the Bos d 5 reference. Nearest-neighbor rankings and group-level statistical analyses were performed using the PCA distances specific to each set (Figure 3-12-Figure 3-20).

2.4.9 Nearest-neighbor ranking analysis

To identify query structures that were most similar to Bos d 5, query structures were ranked according to their distance from the corresponding Bos d 5 reference profile. Rankings were generated separately for each descriptor set and protocol.

The descriptor sets comprised literature-reported mapped linear epitope descriptors (MLERs), mapped conformational epitope descriptors (MCER), comparative epitope descriptors, comparative paratope descriptors, comparative interface descriptors, and the comparative composite descriptor set. For each ranking, the query identifier, biological group, distance value, and rank position were recorded. Rankings were based on ascending PCA distance values (Figure 3-15-Figure 3-16; Table 3-13-Table 3-15; Figures S3.27-S3.35; Tables S3.1-S3.7).

2.4.10 Statistical testing of group differences

2.4.10.1 Pairwise permutation tests

Planned group comparisons were performed using pairwise permutation tests implemented in SciPy.[222] Tests were conducted separately within each descriptor set, epitope region (or matched mapped region), and protocol so that comparisons were restricted to matched descriptor spaces (Table 3-16-Table 3-17; Figure 3-17-Figure 3-20; Figures S3.36-S3.42).

For each comparison, the test statistic was defined as the difference in median distance between groups:

$$T_{\text{obs}} = \text{median}(d_y) - \text{median}(d_x) \qquad \text{Equation 2-20}$$

where d_x and d_y are the Bos d 5 distance values for groups x and y, respectively.

Two-sided permutation tests were performed using 100,000 resamples and a fixed random seed. For independent sample tests, observations were pooled and randomly reassigned to groups while preserving the observed group sizes. When the total number of possible label permutations did not exceed the specified number of resamples, the exact null distribution was evaluated; otherwise, p-values were estimated from randomized permutation sampling.

Planned comparisons included aeroallergens, non-bovine ruminant β-lactoglobulins, endogenous human lipocalins, and intermediate β-lactoglobulin groups. To account for multiple testing, p-values from planned pairwise comparisons were adjusted using the Benjamini–Hochberg false-discovery-rate (FDR) procedure implemented in Statsmodels.[223]

3 RESULTS

3.1 Structural dataset

3.1.1 Structural dataset

The curated structural dataset comprised an experimental IgE Fab–allergen template-complex set and three comparison sets of representative lipocalin protein structures (Table 3-1 – 3-4). The template-complex set contained eight experimentally determined IgE Fab–allergen complexes representing four allergens: Bos d 5 (n=1), Phl p 2 (n=1), Der p 2 (n=2), and Can f 1 (n=4) (Table 3-2). This set was structurally heterogeneous and included one β-lactoglobulin complex[17], one Phl p 2 complex[16], two Der p 2 complexes[12,73], and four Can f 1[2] complexes.

The curated WHO/IUIS lipocalin-allergen set comprised 23 isoallergen-level entries (Table 3-2). Of these, 14 were represented by experimentally determined PDB structures and 9 were represented only by AlphaFoldDB models. The reviewed human lipocalin reference set comprised 18 proteins (Table 3-3), including 13 with experimental structures and 5 represented only by AlphaFoldDB models. The supplementary β-lactoglobulin comparison set comprised of 18 entries (Table 3-4), including the official Bos d 5.0101 (variant B) allergen entry and additional non-bovine β-lactoglobulin homologs; 7 entries had experimental structures and 11 were represented only by AlphaFoldDB models.

Table 3-1. Experimentally determined IgE Fab–allergen complexes.

WHO/IUIS allergen name	Species	WHO/IUIS biochemical name	Isoallergen variant	UniProt accession	PDB ID	Resolution (Å)	Allergen protein family	Epitope nature	Complex (allergen-IgE antibody)
Bos d 5	*Bos domesticus (Bos taurus)*	Beta-lactoglobulin	Bos d 5.0101	P02754	2R56	2.80	Lipocalin	Planar; flat β-sheet surface	nBos d 5-rFab hIgG1 D1; IgE-derived variable regions[17]
Phl p 2	*Phleum pratense (Timothy)*	Grass group II/III	Phl p 2.0101	P43214	2VXQ	1.90	Expansin family	Planar; flat β-sheet surface	rPhl p 2-rFab hIgG1 huMab2; IgE-derived variable regions[16]
Der p 2	*Dermatophagoides pteronyssinus (European house dust mite)*	NPC2 family; MD-2-related lipid recognition (ML) domain containing protein	Der p 2.0103	P49278	7MLH	2.10	NPC2 family	Hydrophobic C-terminal region	rDer p 2-nFab hIgE mAb 2F10[12]
Der p 2	*Dermatophagoides pteronyssinus (European house dust mite)*	NPC2 family; MD-2-related lipid recognition (ML) domain containing protein	Der p 2.0103	I2CMD6	8VK1	3.05	NPC2 family	N-terminal and surface loop	rDer p 2-nFab hIgE mAb 4C8[73]
Can f 1	*Canis familiaris (C. lupus familiaris)*	Lipocalin	Can f 1.0101	O18873	8VQF	3.11	Lipocalin	Surface binding; N-terminus blocks HCDR3 access to cavity	nCan f 1-nFab hIgE mAb 1J11[2]
Can f 1	*Canis familiaris (C. lupus familiaris)*	Lipocalin	Can f 1.0101	O18873	9NPG	3.12	Lipocalin	Can f 1 variant; 12F3 binding preserved	rCan f 1-C100S - nFab hIgE mAb 12F3[2]
Can f 1	*Canis familiaris (C. lupus familiaris)*	Lipocalin	Can f 1.0101	O18873	9NPH	2.59	Lipocalin	Can f 1 variant; 1J11 binding preserved	rCan f 1 - C100S - nFab hIgE mAb 1J11[2]
Can f 1	*Canis familiaris (C. lupus familiaris)*	Lipocalin	Can f 1.0101	O18873	9NPI	1.85	Lipocalin	Calyx/cavity; HCDR3 inserts into hydrophobic cavity	rCan f 1 - nFab hIgE mAb 12F3[2]

Primary UniProtKB accessions and PDB structure identifiers are reported. Database releases and access dates are provided in Sections 2.1.1 and 2.1.2. PDB, Protein Data Bank

Table 3-2. Lipocalin allergens assembled from the WHO/IUIS database.

Allergen name	Isoallergen variant	UniProt accession ID	Species	Representative PDB	Representative structure source	Resolution (Å)	PDB coverage	Mean AFDB pLDDT
Bla g 4	Bla g 4.0101	P54962	*Blattella germanica*	4N7C	Experimental PDB	1.75	0.97	
Bos d 2	Bos d 2.0103	Q28133	*Bos domesticus (Bos taurus)*	4WFV	Experimental PDB	1.40	0.91	
Bos d 5	Bos d 5.0101	P02754	*Bos domesticus (Bos taurus)*	6NKQ	Experimental PDB	2.30	1.00	
Can f 1	Can f 1.0101	O18873	*Canis familiaris (C. lupus familiaris)*	8AEH	Experimental PDB	3.00	0.97	
Can f 2	Can f 2.0101	O18874	*Canis familiaris (C. lupus familiaris)*	8AEI	Experimental PDB	1.74	0.97	
Can f 4	Can f 4.0101	D7PBH4	*Canis familiaris (C. lupus familiaris)*	AF-D7PBH4-F1	AlphaFoldDB			89.43
Can f 6	Can f 6.0101	H2B3G5	*Canis familiaris (C. lupus familiaris)*	8AEJ	Experimental PDB	1.84	0.99	
Cav p 1	Cav p 1.0101	A0A484HM70	*Cavia porcellus*	8A0D	Experimental PDB	3.69	0.95	
Cav p 1	Cav p 1.0201	A0A484HRI4	*Cavia porcellus*	AF-A0A484HRI4-F1	AlphaFoldDB			90.72
Cav p 2	Cav p 2.0101	P83508	*Cavia porcellus*	AF-P83508-F1	AlphaFoldDB			92.11
Cav p 3	Cav p 3.0101	F0UZ12	*Cavia porcellus*	AF-F0UZ12-F1	AlphaFoldDB			90.29
Cav p 6	Cav p 6.0101	S0BDX9	*Cavia porcellus*	AF-S0BDX9-F1	AlphaFoldDB			90.63
Equ c 1	Equ c 1.0101	Q95182	*Equus caballus*	1EW3	Experimental PDB	2.30	0.85	
Equ c 1	Equ c 1.0102	A0A3Q2HX90	*Equus caballus*	AF-A0A3Q2HX90-F1	AlphaFoldDB			87.26
Fel d 4	Fel d 4.0101	Q5VFH6	*Felis domesticus (F. catus)*	8AMC	Experimental PDB	2.95	1.00	
Fel d 7	Fel d 7.0101	E5D2Z5	*Felis domesticus (F. catus)*	8EPV	Experimental PDB	2.19	0.88	
Mes a 1	Mes a 1.0101	Q9QXU1	*Mesocricetus auratus*	AF-Q9QXU1-F1	AlphaFoldDB			91.16
Mus m 1	Mus m 1.0101	P02762	*Mus musculus*	1I06	Experimental PDB	1.90	1.00	
Mus m 1	Mus m 1.0102	P11589	*Mus musculus*	2NND	Experimental PDB	1.60	0.99	
Ory c 4	Ory c 4.0101	U6C8D6	*Oryctolagus cuniculus*	AF-U6C8D6-F1	AlphaFoldDB			90.12
Per a 4	Per a 4.0102	Q1M0Y5	*Periplaneta americana*	3EBW	Experimental PDB	2.80	0.89	
Phod s 1	Phod s 1.0101	S5ZYD3	*Phodopus sungorus*	AF-S5ZYD3-F1	AlphaFoldDB			95.09
Rat n 1	Rat n 1.0101	P02761	*Rattus norvegicus*	2A2U	Experimental PDB	2.50	1.00	

Primary UniProtKB accessions and PDB or AlphaFold DB structure identifiers are reported. Database releases and access dates are provided in Sections 2.1.1 and 2.1.2. PDB, Protein Data Bank; pLDDT, predicted Local Distance Difference Test.

Table 3-3. Reviewed human lipocalin proteins retrieved from UniProtKB

UniProt	Gene	Protein name	Representative structure source	Representative PDB	PDB coverage	Resolution (Å)	AFDB model ID	AFDB pLDDT mean
P09466	PAEP	Glycodelin	Experimental PDB	4R0B	0.94	2.45		
P80188	LCN2	Neutrophil gelatinase-associated lipocalin	Experimental PDB	3CBC	1.00	2.17		
O95445	APOM	Apolipoprotein M	Experimental PDB	2YG2	0.91	1.70		
P19652	ORM2	Alpha-1-acid glycoprotein 2	Experimental PDB	7OUB	0.95	1.82		
P05090	APOD	Apolipoprotein D	Experimental PDB	2HZQ	0.92	1.80		
Q9NY56	OBP2A	Odorant-binding protein 2a	Experimental PDB	4RUN	0.95	2.60		
P02753	RBP4	Retinol-binding protein 4	Experimental PDB	4O9S	1.00	2.30		
P07360	C8G	Complement component C8 gamma chain	Experimental PDB	2RD7	0.91	2.15		
P41222	PTGDS	Prostaglandin-H2 D-isomerase	Experimental PDB	4ORR	1.00	1.40		
P02760	AMBP	Protein AMBP	Experimental PDB	4ES7	0.57	2.00		
P02763	ORM1	Alpha-1-acid glycoprotein 1	Experimental PDB	3KQ0	0.96	1.80		
P31025	LCN1	Lipocalin-1	Experimental PDB	1XKI	0.92	1.80		
Q9NPH6	OBP2B	Odorant-binding protein 2b	AlphaFoldDB				AF-Q9NPH6-F1	86.98
P62502	LCN6	Epididymal-specific lipocalin-6	AlphaFoldDB				AF-P62502-F1	79.73
Q6JVE9	LCN8	Epididymal-specific lipocalin-8	AlphaFoldDB				AF-Q6JVE9-F1	79.96
Q6JVE6	LCN10	Epididymal-specific lipocalin-10	AlphaFoldDB				AF-Q6JVE6-F1	72.06
Q8WX39	LCN9	Epididymal-specific lipocalin-9	AlphaFoldDB				AF-Q8WX39-F1	85.27
Q6UWW0	LCN15	Lipocalin-15	Experimental PDB	2XST	0.88	1.63		

Primary UniProtKB accessions and PDB or AlphaFold DB structure identifiers are reported. Database releases and access dates are provided in Sections 2.1.1 and 2.1.2. PDB, Protein Data Bank; pLDDT, predicted Local Distance Difference Test.

Table 3-4. Other beta-lactoglobulin proteins sourced from UniProtKB.

UniProt	Gene	Organism name	Representative structure source	Representative PDB	PDB coverage	Resolution (Å)	AFDB model ID	AFDB pLDDT mean
*P02754	LGB	*Bos taurus*	Experimental PDB	6NKQ	1.00	2.30		
*B5B0D4		*Bos taurus*	Experimental PDB	5IO6	0.91	2.85		
P02755	LGB	*Bubalus bubalis*	AlphaFoldDB				AF-P02755-F1	91.08
P33685	LGB1	*Canis lupus familiaris*	AlphaFoldDB				AF-P33685-F1	90.54
P02756	LGB	*Capra hircus*	Experimental PDB	4TLJ	0.90	1.17		
P13613	LGB1/2	*Equus asinus*	AlphaFoldDB				AF-P13613-F1	89.65
P19647	LGB1/2	*Equus asinus*	AlphaFoldDB				AF-P19647-F1	84.50
P02758	LGB1/2	*Equus caballus*	Experimental PDB	3KZA	0.90	2.00		
P07380	LGB1/2	*Equus caballus*	AlphaFoldDB				AF-P07380-F1	82.61
P33687	LGB1	*Felis catus*	AlphaFoldDB				AF-P33687-F1	89.65
P21664	LGB2	*Felis catus*	AlphaFoldDB				AF-P21664-F1	90.92
P33688	LGB3	*Felis catus*	AlphaFoldDB				AF-P33688-F1	89.61
P11944	LGB	*Macropus giganteus*	AlphaFoldDB				AF-P11944-F1	86.20
Q29614	LGB	*Notamacropus eugenii*	AlphaFoldDB				AF-Q29614-F1	84.047
P67976		*Ovis aries*	Experimental PDB	4CK4	0.90	1.12		
P67975	LGB	*Ovis aries musimon*	AlphaFoldDB				AF-P67975-F1	94.72
Q00P86		*Rangifer tarandus tarandus*	Experimental PDB	1YUP		2.10		
P04119		*Sus scrofa*	Experimental PDB	1EXS	0.90	2.39		
Q29146	LGB	*Trichosurus vulpecula*	AlphaFoldDB				AF-Q29146-F1	83.14

*Bos d 5.0101 (UniProt P02754) is the official WHO/IUIS bovine β-lactoglobulin allergen entry. P02754 is the reviewed UniProt sequence for β-lactoglobulin variant B, whereas the additional bovine entry B5B0D4 is a distinct Bos taurus β-lactoglobulin accession represented here by PDB 5IO6; because the Allergome database[224] cross-references B5B0D4 as Bos d 5.0102 (variant A), it was retained as a separate bovine β-lactoglobulin. Primary UniProtKB accessions and PDB or AlphaFold DB structure identifiers are reported. Database releases and access dates are provided in Sections 2.1.1 and 2.1.2. PDB, Protein Data Bank; pLDDT, predicted Local Distance Difference Test.

3.2 Workflow development and validation

The method validation was designed to assess whether the template-based workflow could reproducibly generate and quantify comparative Fab–query allergen interfaces. The validation therefore evaluated four practical aspects of workflow performance: reproduction of the experimental template interface when the template allergen was used as its own query, the effect of RAW and MIN protocols on descriptor values, successful descriptor output generation across the full template–query comparison design, and the structural interpretability of the resulting comparative IgE Fab-query allergen modles in relation to fold compatibility.

3.2.1 Identity controls confirmed reproducibility of the reconstruction workflow

Identity-control comparisons were used to assess whether the experimental complex reconstruction workflow altered interface descriptors when no query allergen was introduced. For each of the eight experimental IgE Fab–allergen template complexes, the template allergen was reprocessed as its own query structure under both the RAW and MIN protocols.

Across all eight template complexes, descriptors of the reconstructed complex were identical to the corresponding descriptors of the experimental template complex at the reported precision. This agreement was observed under both RAW and MIN protocols and was consistent across all evaluated descriptor classes, including interface residue counts, ΔSASA, interface aromatic enrichment, hydrogen-bond descriptors, ionic contacts, hydrophobic contacts, π–π interactions, and cation–π interactions. For every descriptor, the maximum absolute difference between the reconstructed identity-control value and the corresponding experimental template value was 0.00 (Figure 3-1; Table 3-5).

These results showed that structural alignment, template-complex reconstruction, patch definition, and descriptor extraction did not introduce measurable descriptor deviation when the template allergen was reprocessed as the query structure. The identity controls therefore confirmed internal reproducibility of the reconstruction and descriptor-extraction workflow under fixed input structures and fixed protocol conditions.

Although identity-control reconstruction preserved descriptor values within each protocol, absolute descriptor values differed between RAW and MIN processing for several template complexes. This indicated that energy minimization could alter interface descriptor magnitudes even when the allergen identity was unchanged. Subsequent analyses therefore evaluated RAW–MIN differences using matched comparisons.

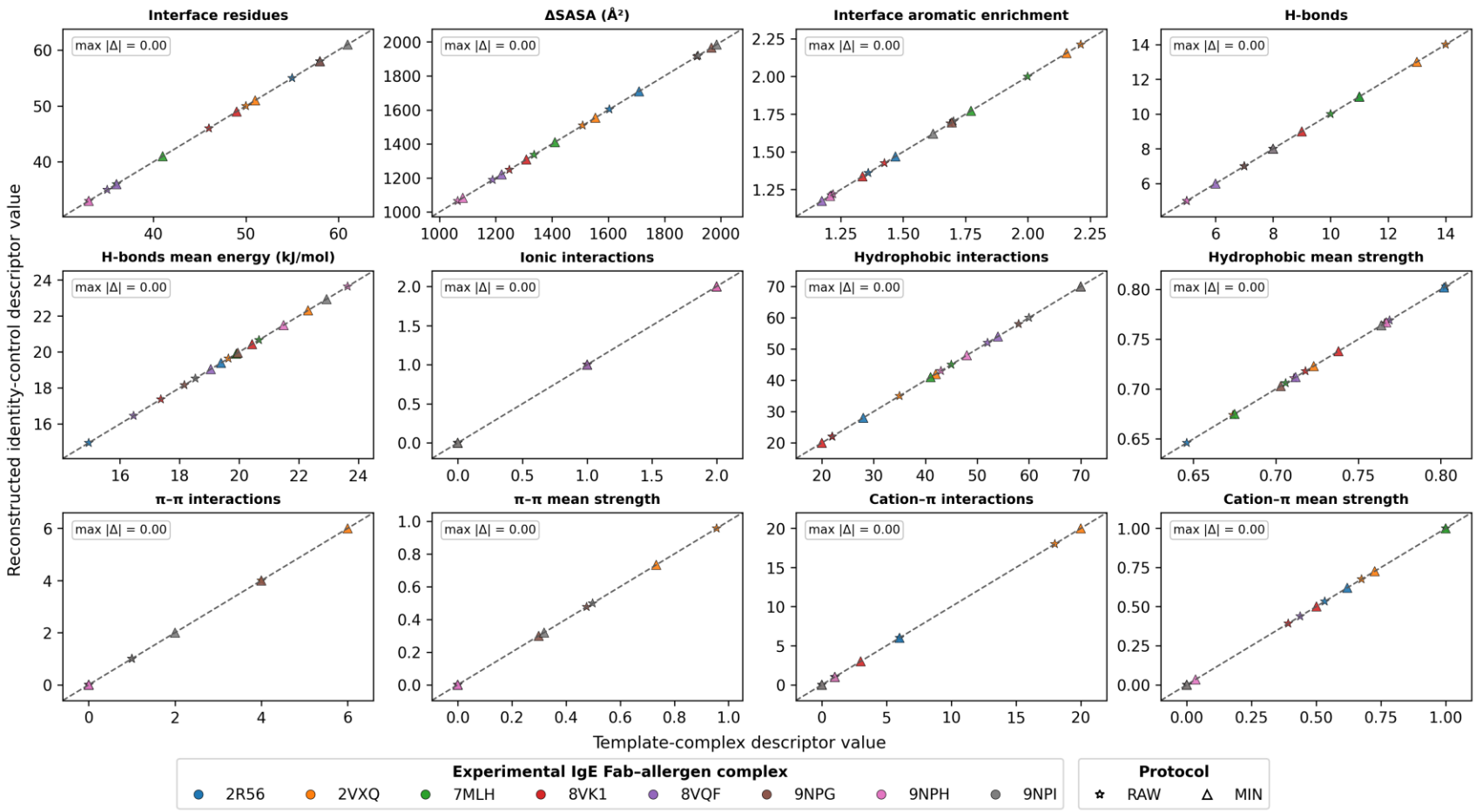

Figure 3-1. Concordance between experimental and reconstructed IgE Fab–allergen complexes.

Each experimental IgE Fab–allergen complex was reprocessed using its own allergen chain as the query structure under the RAW and MIN protocols. The x-axis shows descriptor values measured from the experimental template complex, and the y-axis shows the corresponding descriptor values measured from the reconstructed identity-control complex. Colors indicate template complexes, while marker shapes indicate protocol. The dashed line denotes exact equality between experimental and reconstructed descriptor values. Across all descriptors and protocols, the maximum absolute within-protocol difference was 0.00 at the reported precision, indicating that the reconstruction and descriptor-extraction workflow preserved template-composite values in the absence of allergen substitution.

Table 3-5. Identity-control descriptor values under RAW and MIN protocols.

PDB	Protocol	Interface residues	ΔSASA (Å²)	Interface aromatic enrichment	H-bonds	H-bonds ME (kJ/mol)	Ionic interactions	Hydrophobic interactions	Hydrophobic mean strength	π–π interactions	π–π mean strength	Cation–π interactions	Cation–π mean strength	Max |Δ|
2R56	MIN	58/58	1709.24/1709.24	1.47/1.47	11/11	19/19	2/2	28/28	0.80/0.80	0/0	0.00/0.00	6/6	0.62/0.62	0.00
2R56	RAW	55/55	1603.75/1603.75	1.36/1.36	7/7	15/15	0/0	22/22	0.65/0.65	0/0	0.00/0.00	6/6	0.53/0.53	0.00
2VXQ	MIN	51/51	1554.02/1554.02	2.15/2.15	13/13	22/22	1/1	42/42	0.72/0.72	6/6	0.73/0.73	20/20	0.73/0.73	0.00
2VXQ	RAW	50/50	1508.91/1508.91	2.21/2.21	14/14	20/20	0/0	35/35	0.67/0.67	4/4	0.96/0.96	18/18	0.68/0.68	0.00
7MLH	MIN	41/41	1410.32/1410.32	1.77/1.77	11/11	20/20	0/0	41/41	0.68/0.68	0/0	0.00/0.00	1/1	1.00/1.00	0.00
7MLH	RAW	36/36	1336.48/1336.48	2.00/2.00	10/10	21/21	0/0	45/45	0.71/0.71	0/0	0.00/0.00	1/1	1.00/1.00	0.00
8VK1	MIN	49/49	1308.72/1308.72	1.34/1.34	9/9	20/20	2/2	20/20	0.74/0.74	0/0	0.00/0.00	3/3	0.50/0.50	0.00
8VK1	RAW	46/46	1248.43/1248.43	1.43/1.43	8/8	17/17	1/1	22/22	0.72/0.72	0/0	0.00/0.00	1/1	0.39/0.39	0.00
8VQF	MIN	36/36	1221.11/1221.11	1.17/1.17	6/6	19/19	1/1	54/54	0.71/0.71	0/0	0.00/0.00	0/0	0.00/0.00	0.00
8VQF	RAW	35/35	1188.92/1188.92	1.21/1.21	5/5	16/16	1/1	52/52	0.77/0.77	0/0	0.00/0.00	1/1	0.44/0.44	0.00
9NPG	MIN	58/58	1966.39/1966.39	1.70/1.70	8/8	20/20	0/0	70/70	0.70/0.70	4/4	0.30/0.30	0/0	0.00/0.00	0.00
9NPG	RAW	58/58	1914.93/1914.93	1.69/1.69	7/7	18/18	0/0	58/58	0.77/0.77	1/1	0.48/0.48	0/0	0.00/0.00	0.00
9NPH	MIN	33/33	1082.99/1082.99	1.21/1.21	8/8	21/21	2/2	48/48	0.77/0.77	0/0	0.00/0.00	1/1	0.03/0.03	0.00
9NPH	RAW	33/33	1064.95/1064.95	1.22/1.22	5/5	24/24	1/1	43/43	0.71/0.71	0/0	0.00/0.00	0/0	0.00/0.00	0.00
9NPI	MIN	61/61	1984.78/1984.78	1.62/1.62	8/8	23/23	0/0	70/70	0.76/0.76	2/2	0.32/0.32	0/0	0.00/0.00	0.00
9NPI	RAW	58/58	1919.08/1919.08	1.70/1.70	8/8	19/19	0/0	60/60	0.80/0.80	1/1	0.50/0.50	0/0	0.00/0.00	0.00

Descriptor values are reported for each experimental IgE Fab–allergen template complex and its corresponding reconstructed identity-control complex. Max |Δ| denotes the maximum absolute difference between paired experimental template values and reconstructed identity control values across the descriptor set. Reported descriptors include interface residue count, ΔSASA, interface aromatic enrichment, hydrogen-bond count, hydrogen-bond energy, ionic-contact count, hydrophobic-contact count, π–π interaction count, cation–π interaction count, and the strength of each type of intermolecular interaction. ΔSASA is reported in Å². Aromatic enrichment and interaction strengths are reported as ratios.

3.2.2 Effect of protocol-level minimization

Matched RAW and MIN comparisons were used to determine whether the optional minimization step altered descriptor values in the absence of allergen substitution. This analysis was performed across the eight identity-control complexes, with the protocol difference defined as Δ = MIN − RAW (Figures 3-2 and 3-3). Although identity-control reconstruction preserved descriptor values within each protocol, RAW and MIN processing produced different absolute descriptor values for several interface features. Thus, minimization introduced protocol-level differences even when the template allergen was retained as its own query structure.

Across broader interface-size and burial descriptors, minimization generally increased descriptor magnitudes while largely preserving the relative ordering of template complexes. Interface residues increased after minimization, with a median paired difference of $\tilde{\Delta}$ = +2 residues, IQR = 2.2, and strong RAW–MIN concordance (ρ = 0.98; Figures 3-2A and 3-3A). The corresponding Δ distribution was entirely non-negative within the whisker range [0, 5], indicating that minimization either preserved or increased the interface residues across complexes. Buried surface area also increased consistently under MIN, with $\tilde{\Delta}$ = +56 Å², IQR = 26 Å², and complete rank concordance across complexes (ρ = 1.00; Figures 3-2B and 3-3B). Its paired Δ distribution was fully positive within the whisker range [18, 110] Å². Hydrophobic-contact counts also showed a positive overall shift, with $\tilde{\Delta}$ = +5.5, IQR = 6.8, and high RAW–MIN concordance (ρ = 0.92; Figures 3-2G and 3-3G). However, the Δ distribution spanned both negative and positive values [−4, 12], indicating that the increase was not uniform across all complexes.

Interface aromatic enrichment changed only modestly and not strictly in a positive direction. The median paired difference was slightly negative ($\tilde{\Delta}$ = −0.048), with IQR = 0.076, two outliers, and high RAW–MIN rank concordance (ρ = 0.95; Figures 3-2C and

3-3C). The non-outlier Δ range [−0.088, 0.006] indicated a small downward shift for most complexes, although two complexes showed larger deviations. Hydrophobic mean strength was centered close to zero, with $\tilde{\Delta}$ = −0.0055, IQR = 0.094, and weak negative rank concordance between RAW and MIN values (ρ = −0.21; Figures 3-2H and 3-3H). Thus, although hydrophobic-contact counts generally increased after minimization, hydrophobic mean strength did not show a consistent response that preserve the rank.

The descriptors that depend more directly on local atomic geometry showed greater protocol sensitivity. Hydrogen-bond counts increased modestly overall, with $\tilde{\Delta}$ = +1, IQR = 0.75, and moderate RAW–MIN concordance (ρ = 0.77; Figures 3-2D and 3-3D). The Δ range was [0, 1.5], but three outliers were detected, indicating that a subset of complexes underwent larger hydrogen-bond changes after minimization. Hydrogen-bond mean energy showed a positive median shift but lower concordance, with $\tilde{\Delta}$ = +2.6 kJ/mol, IQR = 2.2 kJ/mol, and ρ = 0.57 (Figures 3-2E and 3-3E). Its Δ distribution spanned [−2.1, 4.4] kJ/mol, indicating heterogeneous energetic responses among complexes. Ionic interactions were sparse and changed modestly, with $\tilde{\Delta}$ = +0.5, IQR = 1, a Δ range of [0, 2], and limited RAW–MIN concordance (ρ = 0.60; Figures 2-2F and 2-3F).

Aromatic interaction descriptors showed mixed protocol sensitivity. π–π interaction counts had a median paired difference of zero, with IQR = 1.2, and very high RAW–MIN concordance (ρ = 0.99; Figures 3-2I and 3-3I). The Δ distribution ranged from [0, 3], indicating that most complexes were unchanged while a subset gained π–π interactions after minimization. π–π mean strength also had $\tilde{\Delta}$ = 0, with IQR = 0.18, and complete rank concordance (ρ = 1.00; Figures 3-2J and 3-3J). Its Δ range [−0.22, 0] suggested unchanged or slightly reduced mean strength after minimization. Cation–π interaction counts changed modestly, with $\tilde{\Delta}$ = 0, IQR = 1.2, a Δ range of [−1, 2], and moderate RAW–MIN concordance (ρ = 0.79; Figures 3-2K and 3-3K). Cation–π mean strength

showed a small positive median shift, with $\tilde{\Delta}$ = +0.017, IQR = 0.060, and ρ = 0.83, but one outlier reflected a larger negative change in one complex (Figures 3-2L and 3-3L).

Overall, minimization produced measurable protocol-level shifts in several descriptors, especially ΔSASA, interface-residue counts, hydrogen-bond descriptors, hydrophobic-contact counts, and selected aromatic interaction features. Some descriptors retained strong RAW–MIN rank concordance despite changes in magnitude, particularly ΔSASA, interface-residue counts, hydrophobic-contact counts, and π–π interaction counts. Other descriptors, especially hydrogen-bond mean energy, ionic contacts, and hydrophobic mean strength, were more sensitive to local geometric changes introduced by minimization. Therefore, RAW and MIN outputs were analyzed as matched but distinct protocol states in subsequent analyses and were not considered interchangeable measurements.

The structural basis of these RAW–MIN differences is illustrated by the 2R56 identity control (Figure 3-4**)**. In this example, minimization preserved the overall template-defined placement of the allergen within the Fab while producing only small adjustments at the interface. The main protocol-dependent changes were limited to contact-boundary reassignment and refinement of interaction geometry, consistent with the increase in hydrogen-bond count from 7 in the RAW protocol to 11 under the MIN protocol . More generally, this pattern supports the interpretation that minimization acts primarily as a relaxation step that can alter threshold-sensitive interaction counts, especially hydrogen bonds, while leaving the broader interface footprint largely unchanged. This is consistent with the general role of energy minimization in relieving steric strain and improving local stereochemical geometry during structural refinement.[225–227]

Overall, minimization produced the expected effect, increasing burial and hydrophobic packing, while leaving the relative behavior of the broader contact and burial descriptors largely unchanged.[228] Within this identity allergen-query control validation set, the additional minimization step therefore offered limited added value for comparative analyses driven mainly by interface descriptors[226,227], whereas descriptors that depend on exact atomic geometry, particularly hydrogen bonds and to a lesser extent ionic contacts, remained more protocol-sensitive.[226–228]

3.2.3 Protocol choice implications for subsequent analyses

Together, the identity-control and protocol-comparison results show that the workflow was internally reproducible under fixed inputs and fixed processing conditions. In the absence of allergen substitution, identity controls reproduced the compact descriptor panel exactly within protocol, whereas the additional minimization step changed some absolute descriptor values without altering the broader ordering of complexes for the main interface-size and packing descriptors.

Accordingly, both RAW and MIN were carried forward for downstream analyses. RAW provided the reference representation closest to the deposited experimental complexes, whereas MIN served as a complementary, relaxed representation for testing protocol dependence. Using both protocols allowed robust comparative interpretation of results.

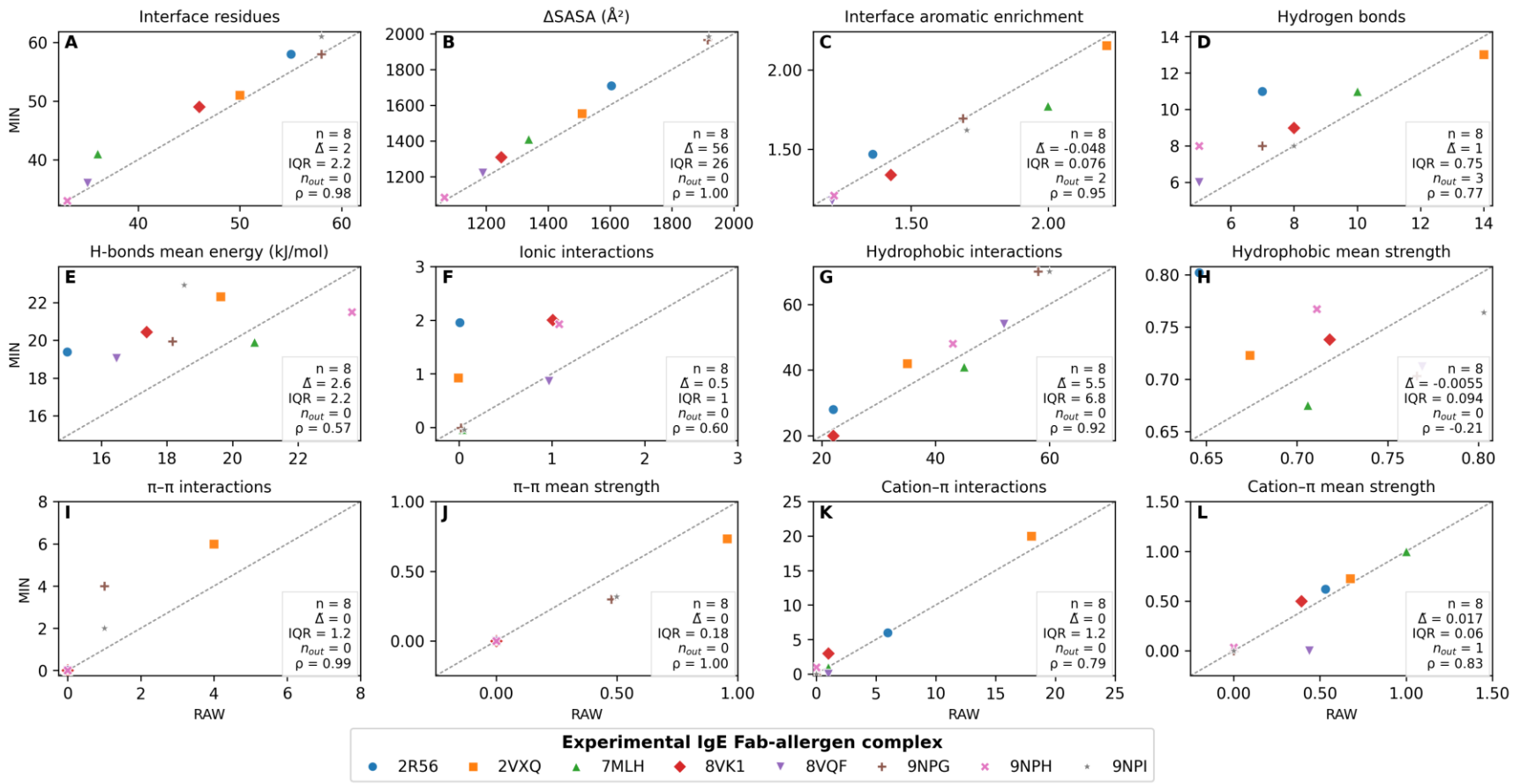

Figure 3-2. Protocol sensitivity of interface descriptors.

Matched RAW and MIN descriptor values were compared across the eight identity-control IgE Fab–allergen template complexes. Each panel shows one interface descriptor, with RAW values on the x-axis and MIN values on the y-axis. The dashed diagonal indicates equality between protocols (y=x); points above the line indicate an increase after minimization, whereas points below the line indicate a decrease. Colors and symbols identify the experimental IgE Fab–allergen template complexes. Inset statistics summarize the paired RAW–MIN comparison for each descriptor, where n is the number of template complexes, $\tilde{\Delta}$ is the median paired difference calculated as MIN − RAW, IQR is the interquartile range of the paired differences, n_{out} is the number of outliers in the paired differences, and ρ denotes Spearman rank concordance between RAW and MIN values across template complexes.

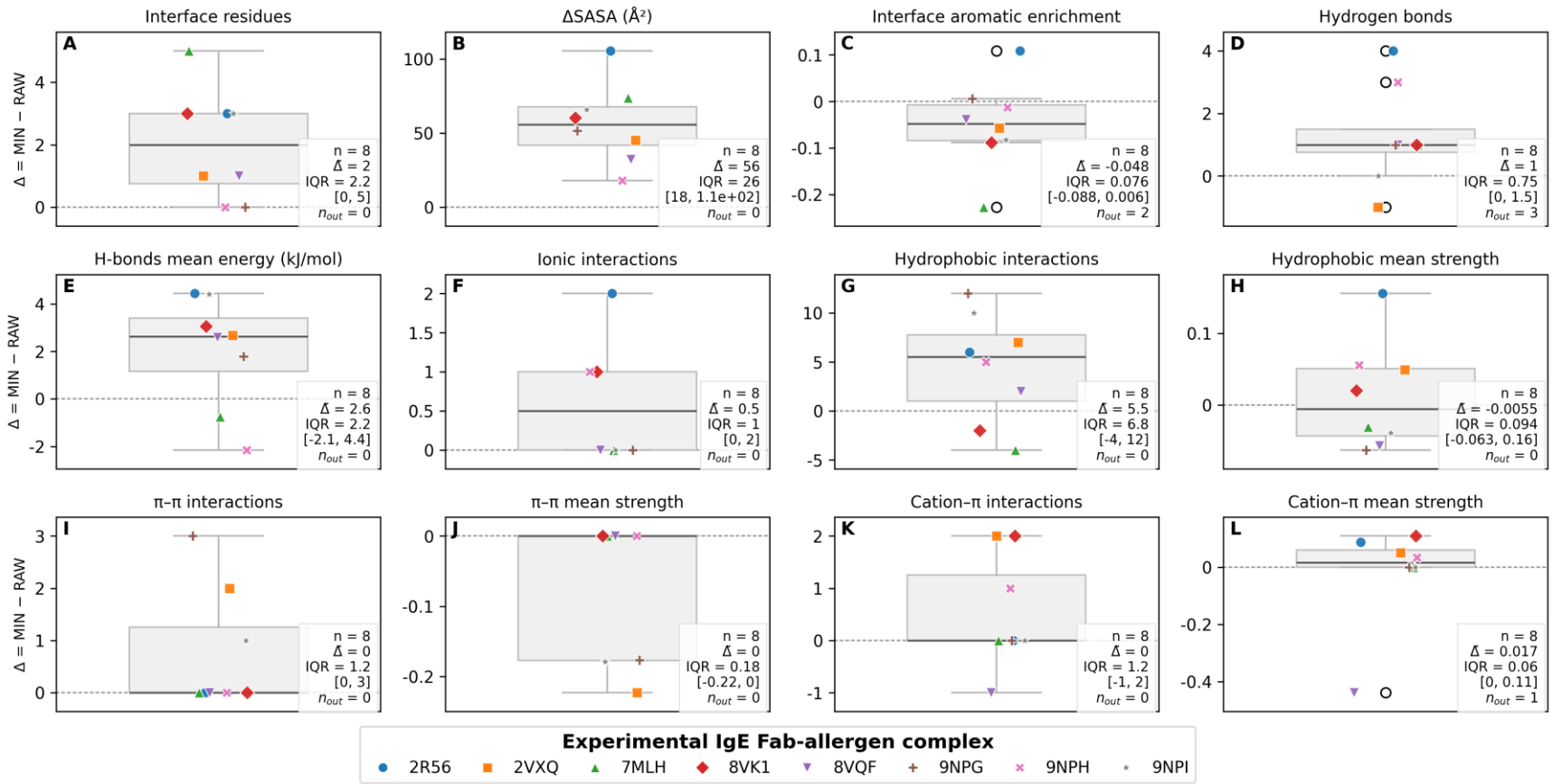

Figure 3-3. Distribution of protocol-level descriptor changes after minimization.

Paired descriptor differences were calculated for each identity-control complex as Δ = MIN − RAW. Each panel shows the distribution of RAW–MIN differences for one interface descriptor across the eight experimental IgE Fab–allergen template complexes. Boxplots summarize the paired Δ distribution, with individual complexes overlaid as colored symbols. The dashed horizontal line marks Δ = 0, corresponding to no protocol-dependent change. Positive values indicate descriptor increases after minimization, whereas negative values indicate decreases. Inset statistics report the number of template complexes, central Δ value, interquartile range, range, and the number of outlying points (n_{out}) for each descriptor.

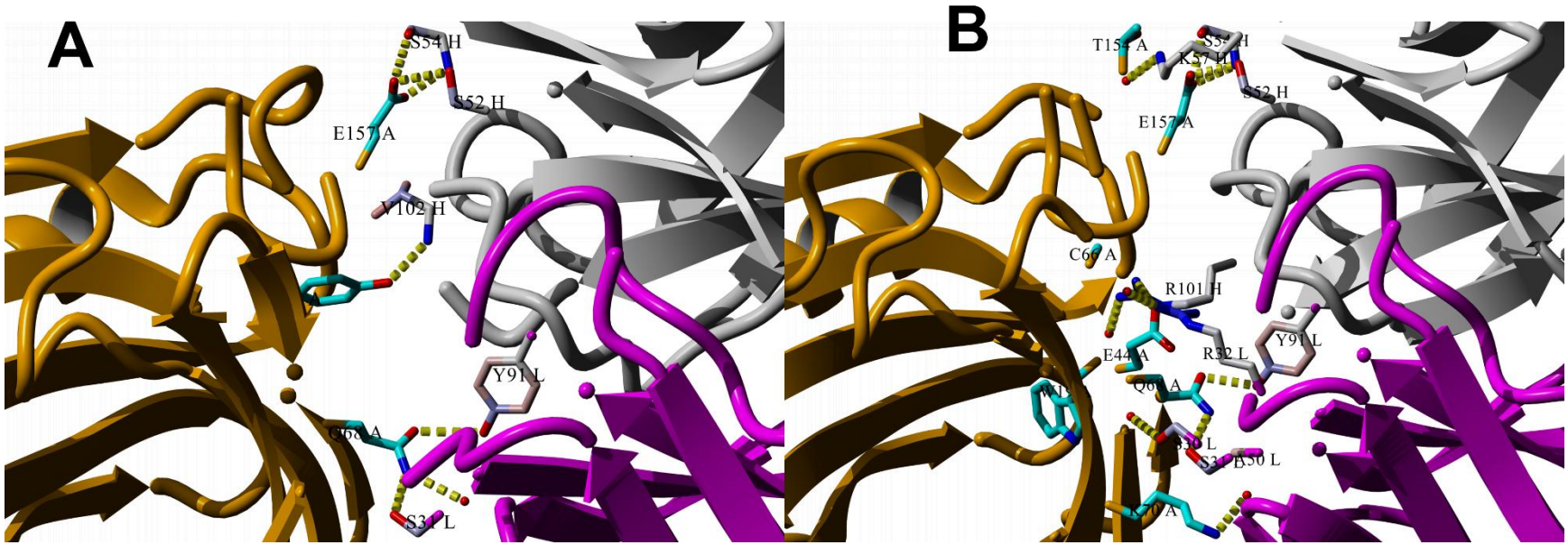

Figure 3-4. Structural effects of energy minimization on the IgE Fab–Bos d 5 interface.

(A) Unminimized (RAW) and minimized (MIN) complexes. Epitope residues involved in hydrogen bonding are colored in cyan, whereas paratope residues are colored by charge. Intermolecular hydrogen-bond networks in the RAW (A) and MIN (MIN) complexes, respectively, show an increase in the number of hydrogen bonds from 7 in RAW (A) to 11 in MIN (B), driven by geometric refinement and contact-threshold reassignment.

3.2.4 Model generation and matched RAW-MIN comparison set

The workflow output was assessed to determine whether the workflow produced usable descriptor outputs across the full template–query design. The full comparison design comprised eight template complexes and 58 lipocalin query structures, corresponding to 464 possible template-complex/query-allergen comparisons per protocol (Figure 3-5). Under the RAW protocol, descriptor values were obtained for all 464 comparisons. Under the MIN protocol, descriptor values were obtained for 455/464. The nine missing MIN comparisons reflected cases in which energy minimization failed to yield a converged structure.

Missing MIN outputs were limited to a small subset of template complexes. One query output was missing for 8VK1, 9NPH, and 9NPI, four outputs were missing for 2VXQ, and two query outputs were missing for 9NPG. Complete MIN coverage was retained for the 8VQF, 2R56, and 7MLH template complexes. Thus, missingness was localized to specific template-complex/query-allergen combinations.

All paired RAW–MIN analyses were restricted to the matched comparisons successfully measured under both protocols. Because RAW output was complete, the matched analysis set was determined by MIN convergence and comprised 455/464 comparisons. By template complex, matched output ranged from 54/58 to 58/58 lipocalin-allergen queries (Figure 3-5). This matched set was used for subsequent paired RAW–MIN analyses.

Overall, the output analysis showed that the workflow can handle many comparisons efficiently. The RAW protocol produced descriptor outputs for the complete design, whereas the MIN protocol introduced a small number of non-converged cases due to

numerical instability. These failures were limited and interpretable, supporting the use of the remaining matched comparisons for downstream protocol comparisons.

3.2.5 Structural alignment quality and fold compatibility

Structural alignment quality was evaluated because placement of each query allergen within the binding position defined by the template complex depends on structural correspondence between the template allergen and the query allergen. This analysis was performed after workflow quality control and therefore distinguishes unavailable comparisons from comparisons that were available but did not establish residue-level structural correspondence.

Unavailable comparisons corresponded to cases in which no usable MIN output was generated because energy minimization failed to converge. In contrast, comparisons that were available for analysis but returned RMSD = 0.00 Å with zero aligned Cα pairs under the selected YASARA MUSTANGPP alignment criteria were deemed not to meet the structural-alignment inclusion criterion and are therefore not true zero-RMSD structural superpositions.

Structural alignment quality was compared between the RAW and MIN datasets for each template complex (Figure 3-6). The proportion of comparisons with at least one aligned Cα pair was highest for the lipocalin template complexes. The Bos d 5 template 2R56 and the Can f 1 templates 8VQF, 9NPG, 9NPH, and 9NPI showed complete or near-complete inclusion in both protocols. In RAW, all five lipocalin templates had 58/58 comparisons with at least one aligned Cα pair. In MIN, 2R56 and 8VQF retained complete inclusion with 58/58 comparisons, whereas 9NPH and 9NPI had 57/57 comparisons and 9NPG had 56/56 comparisons among available MIN outputs. Among included comparisons, these lipocalin templates showed consistently low median Cα RMSD

values, generally around 1.5–1.8 Å, and high median aligned Cα pair counts of approximately 100–130 residues. Their median sequence identities over aligned residues were also higher than those of the non-lipocalin templates, clustering around approximately 20–25%.

By contrast, the structurally distinct non-lipocalin templates showed weaker correspondence with the lipocalin query set. The Phl p 2 template 2VXQ had the lowest proportion of included comparisons, with 37/58 in RAW and 33/54 in MIN. The Der p 2 templates 7MLH and 8VK1 also showed reduced inclusion, with 42/58 and 43/58 comparisons in RAW, respectively, and 41/58 and 39/57 comparisons in MIN, respectively. These three templates had much shorter aligned regions than the lipocalin templates, with median aligned Cα pair counts generally below 30 residues, and they also showed lower median sequence identity over aligned residues. Although their median Cα RMSD values were not always much higher than those of the lipocalin templates, these RMSD values apply only to the subset of comparisons that returned at least one aligned Cα pair and were therefore considered in conjunction with the inclusion percentage and aligned length.

Comparisons with zero aligned Cα pairs were concentrated entirely among the non-lipocalin templates 2VXQ, 7MLH, and 8VK1, as expected for template allergens that do not share the lipocalin fold. No such cases were observed for the lipocalin templates 2R56, 8VQF, 9NPG, 9NPH, and 9NPI in either RAW or MIN. Because comparisons with RMSD = 0.00 Å and zero aligned Cα pairs did not establish residue-level structural correspondence, they were counted as not meeting the structural-alignment inclusion criterion in Figure 3-6A and were excluded from the Cα RMSD, aligned-length, and sequence-identity summaries in Figure 3-6B–D to avoid interpretation of failed structural correspondences as valid zero-RMSD alignments.

Overall, the structural alignment analysis showed that template-based replacement was best supported when the template allergen and query allergen shared the lipocalin fold. The reduced inclusion percentages, shorter aligned regions, and lower sequence identities observed for 2VXQ, 7MLH, and 8VK1 define the practical boundary of applying the lipocalin query set to non-lipocalin template allergens. Thus, the workflow could be applied across all template complexes, but the structural interpretability of the resulting Fab–query allergen models depended strongly on fold compatibility between the template allergen and the query allergen.

3.2.6 Workflow outputs, reproducibility, and computational execution

The workflow produced traceable outputs for each successfully completed comparison between a template complex and a query allergen. Constructed IgE Fab–query allergen complexes were saved as coordinate files, and descriptor values were written as tabular outputs for each comparison (Table S2. 1-S2.6). The same alignment settings, chain assignment rules, contact criteria, and descriptor definitions were applied across all comparisons.

Runtime was recorded for each template complex batch (Figure 3-7). RAW outputs were generated rapidly, requiring approximately 16 to 25 seconds per template complex. Completion of the full RAW and MIN workflow required 632 to 824 seconds. This increase was expected because the MIN protocol included additional structure preparation and energy minimization before descriptor extraction. Overall, RAW provided a fast workflow for descriptor extraction, whereas MIN provided a relaxed structural representation at greater computational cost.

3.2.7 Scope and limitations of the workflow

The validation analyses show that the workflow is reproducible and scalable for structural comparison based on template replacement. Identity controls confirmed that the workflow did not introduce artificial descriptor deviations when the template allergen was used as its own query.

The output analysis showed high output retention across the full comparison design, with missing MIN outputs limited to cases where minimization failed to produce a converged structure. The main limitation is that the workflow measures preservation of a paratope–epitope interface defined by the template. Interpretation is strongest when the template allergen and query allergen share a conserved fold. The workflow also showed that RAW and MIN protocols are not interchangeable because minimization changed several descriptor values, especially descriptors sensitive to local geometry and contact formation.

Therefore, the workflow is a controlled structural comparison method for comparative Fab–query allergen interfaces within a defined template context.

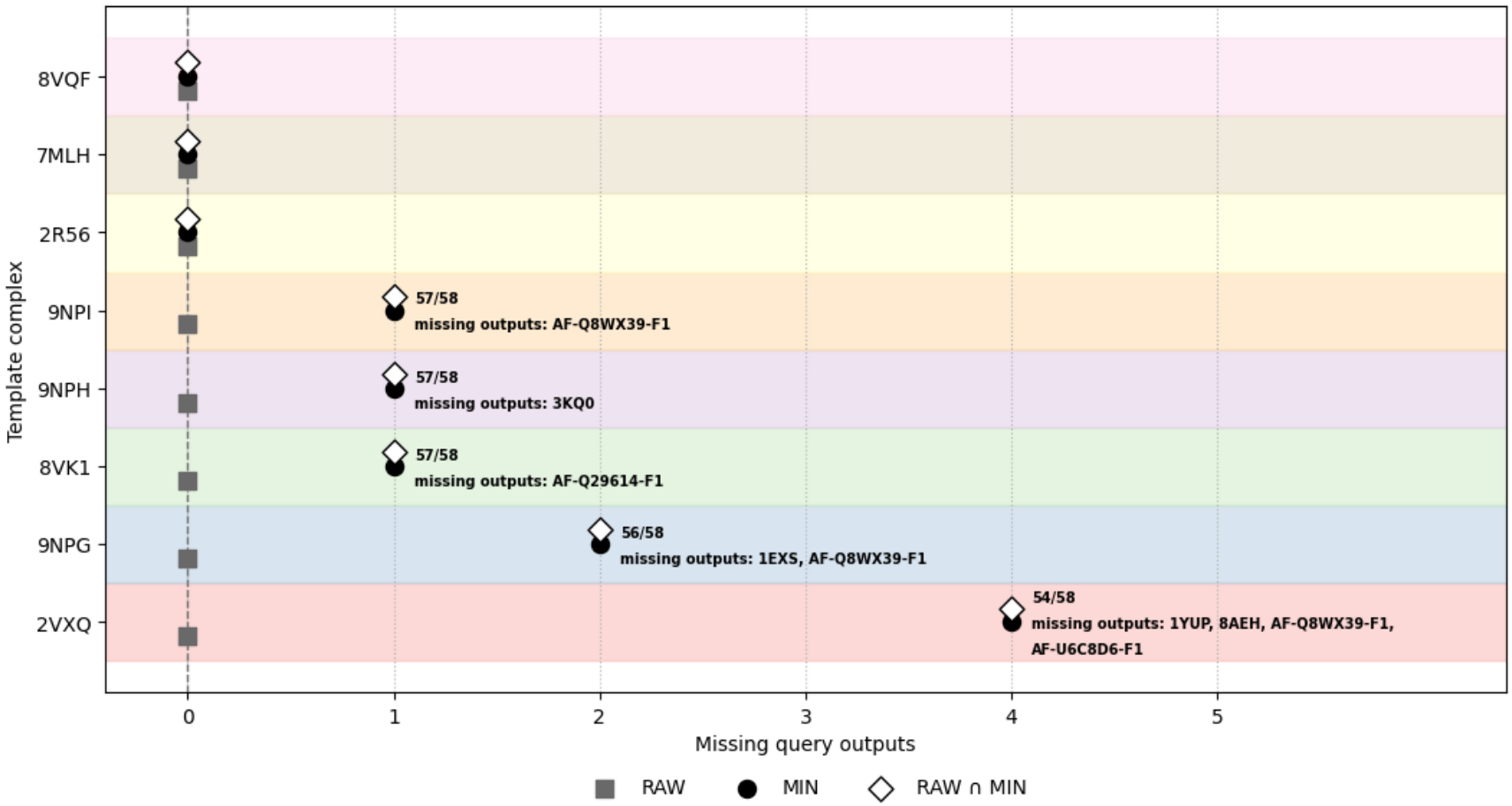

Figure 3-5. Workflow output by template complex and protocol.

Each experimental IgE Fab–allergen template complex was paired with the input lipocalin query structures and processed under the RAW and MIN protocols. Points show the number of template-query comparisons that did not produce completed descriptor rows for RAW, MIN, and the matched RAW ∩ MIN set. The matched set includes only comparisons with descriptor rows available under both protocols. Labels identify incomplete matched cases and list the corresponding query structures. RAW produced completed descriptor rows for all template-query comparisons, whereas incomplete matched cases were driven by missing MIN outputs.

Table 3-6. Structural alignment validity and quality by template complex.

Protocol	Template complex	Outputs	Comparisons with ≥1 aligned Cα pair	Comparisons with <1 aligned Cα pair	Median Cα RMSD (Å)	Median aligned Cα pairs	Median sequence identity (%)
MIN	2R56	58	100.00	0	1.72	129	23.66
MIN	2VXQ	54	61.11	21	1.85	25	10.00
MIN	7MLH	58	70.69	17	1.80	29	10.26
MIN	8VK1	57	68.42	18	1.77	28	6.25
MIN	8VQF	58	100.00	0	1.63	104	22.28
MIN	9NPG	56	100.00	0	1.86	96	22.22
MIN	9NPH	57	100.00	0	1.70	105	22.00
MIN	9NPI	57	100.00	0	1.81	97	22.64
RAW	2R56	58	100.00	0	1.64	129	24.87
RAW	2VXQ	58	63.79	21	1.72	21	8.33
RAW	7MLH	58	72.41	16	1.85	28	10.39
RAW	8VK1	58	74.14	15	1.94	29	6.82
RAW	8VQF	58	100.00	0	1.62	105	21.55
RAW	9NPG	58	100.00	0	1.78	99	22.58
RAW	9NPH	58	100.00	0	1.66	106	22.22
RAW	9NPI	58	100.00	0	1.71	99	23.23

Structural-alignment quality was summarized for RAW and MIN outputs by template complex. Comparisons with ≥1 aligned Cα pair indicates the percentage of available outputs that produced at least one aligned Cα pair under the selected MUSTANGPP alignment parameters. Comparisons with <1 aligned Cα pair had RMSD = 0.00 Å and zero aligned Cα pairs. Median RMSD, aligned Cα pair count, and sequence identity were calculated from valid structural superpositions only. Non-converged MIN outputs were excluded from MIN denominators.

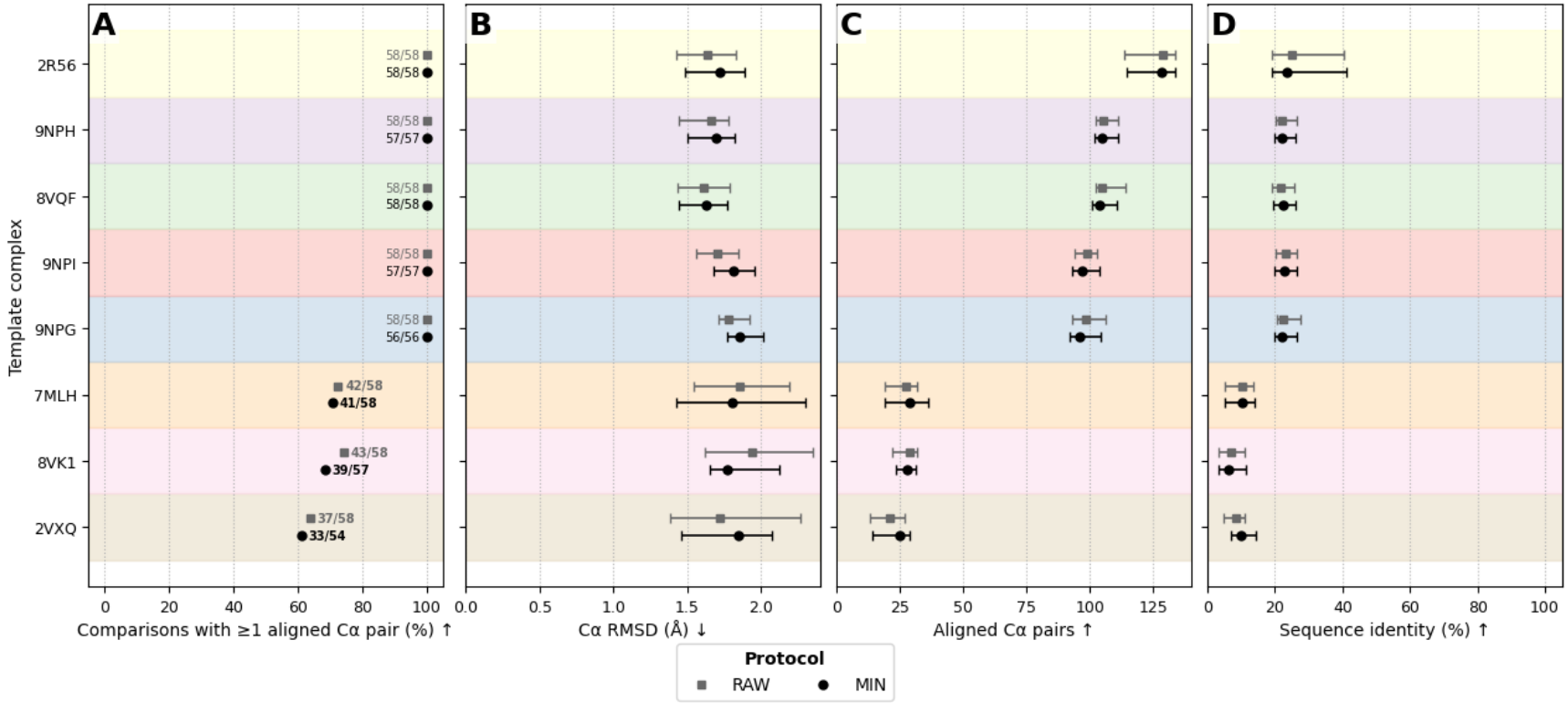

Figure 3-6. Structural alignment quality by template complex.

Structural-alignment outputs were compared between the RAW and MIN datasets for each IgE Fab–allergen template complex. A: Percentage of available template–query comparisons that returned at least one aligned Cα residue pair under the selected MUSTANGPP alignment parameters. Labels indicate the number of comparisons meeting this criterion over the number of available comparisons. B–D: Cα RMSD, number of aligned Cα residue pairs, and sequence identity among comparisons that returned at least one aligned Cα residue pair. Sequence identity was calculated over the aligned residues. Non-converged MIN outputs were considered unavailable workflow outputs and were excluded from the MIN denominators. Comparisons reporting RMSD = 0.00 Å together with zero aligned Cα pairs were considered not to meet the structural-alignment inclusion criterion; these comparisons were counted as not meeting the criterion in A and excluded from B–D because no residue-level structural correspondence was established. Points indicate medians, and horizontal bars indicate interquartile ranges. Higher values in A, C, and D indicate a greater proportion of included comparisons, a larger aligned region, and higher sequence identity over aligned residues, respectively. Lower values in B indicate lower Cα RMSD and therefore greater structural similarity between the template allergen and the aligned query structure.

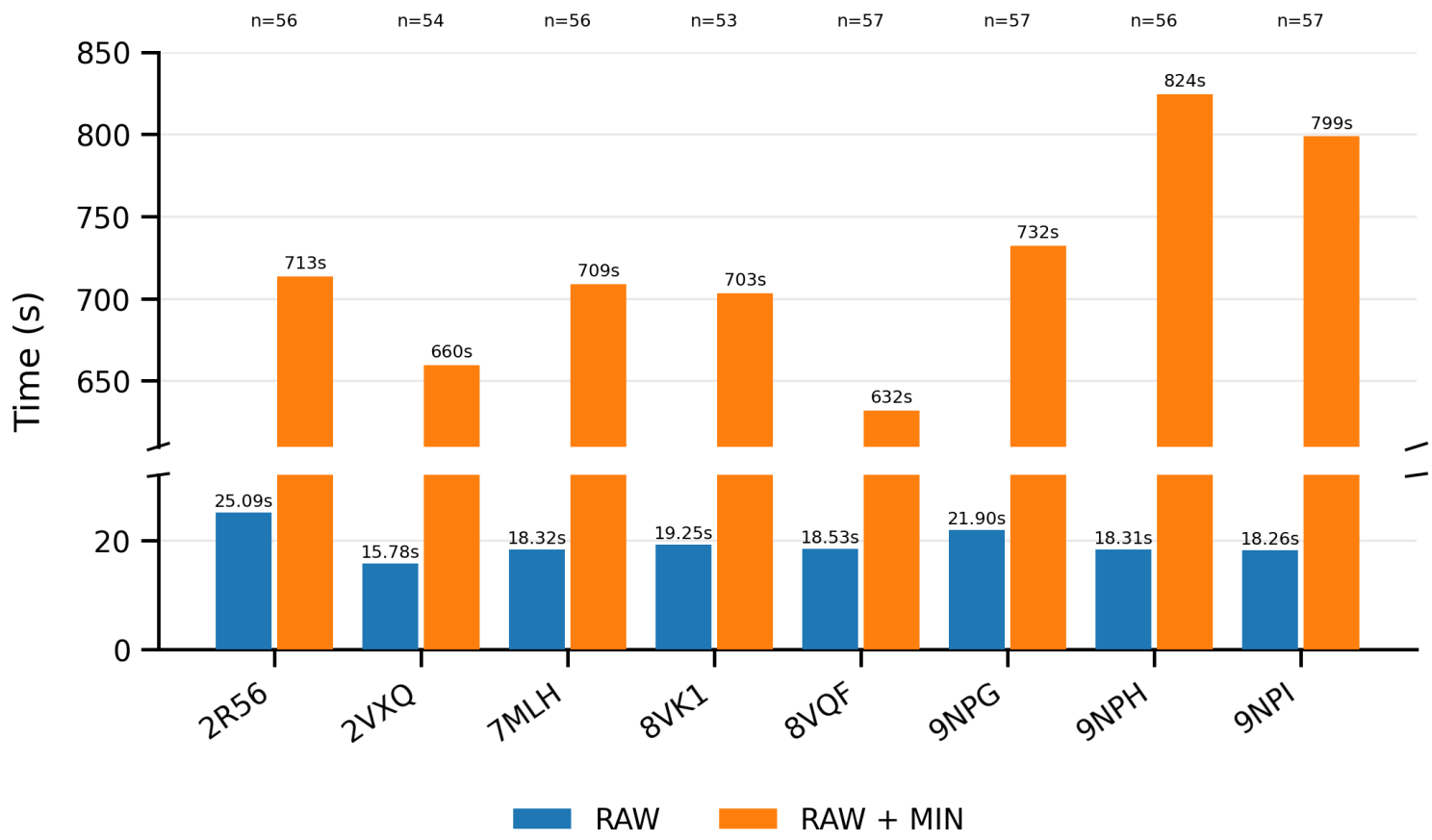

Figure 3-7. Runtime of workflow execution by template complex.

Wall-clock time was recorded for each template-complex batch. Blue bars show the time required to run the RAW protocol alone, which produced descriptor outputs for all query comparisons. Orange bars show the total runtime for the combined RAW + MIN workflow, including RAW descriptor generation, structure preparation, energy minimization, and descriptor extraction. Values above bars indicate elapsed time in seconds. Labels at the top indicate the number of matched RAW–MIN comparisons available for each template complex, defined as comparisons for which both RAW and converged MIN descriptor outputs were obtained. Because RAW outputs were generated for all queries, matched counts below 57 reflect MIN output cases where energy minimization failed to yield a converged structure. The broken y-axis is used because RAW runtimes were much shorter than combined RAW + MIN runtimes.

3.3 Workflow application

3.3.1 Composition of the lipocalin comparison dataset

The Bos d 5 application dataset comprised 59 lipocalin structures, including one Bos d 5 reference structure and 58 non-reference query structures. The query structures were assigned to five biological comparison groups: non-bovine ruminant β-lactoglobulins, non-ruminant ungulate β-lactoglobulins, distant β-lactoglobulin homologs, aeroallergens, and human lipocalins.

Aeroallergens formed the largest query group, with 23 structures, followed by human lipocalins with 18 structures. The β-lactoglobulin comparison set included five non-bovine ruminant β-lactoglobulins, five non-ruminant ungulate β-lactoglobulins, and seven distant β-lactoglobulin homologs. The group assignments and their roles in the application analysis are summarized in Table 3-7.

Table 3-7. Lipocalin query groups used for comparison with Bos d 5

Group	Number of lipocalin structures	Purpose
Bos d 5 reference	1	Reference allergen
Non-bovine ruminant βLGs	5	Positive control (close homologs)
Non-ruminant ungulate βLGs	5	Intermediate comparators
Distant βLG homologs	7	Fold retention comparators
Aeroallergens	23	Primary hypothesis group
Human lipocalins	18	Tolerated (non-allergenic) comparators

3.3.2 Bos d 5 reference IgE-binding regions used for comparison

The curated literature set contained 20 reported Bos d 5 linear IgE-binding regions. Several regions overlapped or were contained within longer reported regions. For the primary analysis, a minimally redundant set of 12 regions was selected (Table 3-8; Table S3. 8).

The selected linear IgE-binding regions covered multiple parts of the Bos d 5 sequence, including N-terminal, central, and C-terminal segments. When mapped onto the Bos d 5 structure, these regions were distributed across both loop-associated and β-strand-associated surface areas (Figure 3-8).

The contact-defined Bos d 5 conformational epitope was derived from the 2R56 IgE Fab–Bos d 5 complex using the standardized heavy-atom contact criterion. The contact-defined residue sets overlapped with the author-reported epitope from the experimental complex (Figure 3-9). The RAW contact-defined epitope omitted some author-reported residues and included a small number of additional residues satisfying the contact criterion. The MIN contact-defined epitope showed a similar overall pattern but differed slightly from the RAW set. RAW and MIN structural-epitope definitions were therefore retained separately in downstream analyses.

Table 3-8. Minimally redundant Bos d 5 linear IgE-binding regions.

No.	Mature chain	P02754 precursor	Reported sequence	Reference(s)
2	1-16	17-32	LIVTQTMKGLDIQKVA	106,108
3	9–29	25-45	GLDIQKVAGTWYSLAMAASDI	109
4	31-60	47-76	LLDAQSAPLRVYVEELKPTPEGDL EILLQK	45,108,109
7	56-70	72-86	ILLQKWENGECAQKK	106
8	58-77	74-93	LQKWENGECAQKKIIAEKTK	105
9	67-86	83-102	AQKKIIAEKTKIPAVFKIDA	108
10	76-90	92-106	TKIPAVFKIDALNEN	106,109
12	95-113	111-129	LDTDYKKYLLFCMENSAEP	103,107,109
15	102-124	118-140	YLLFCMENSAEPEQSLVCQCLVR	45
16	121-135	137-151	CLVRTPEVDDEALEK	109,229
18	127-156	143-172	EVDDEALEKFDKALKALPMHIRLS FNPTQL	106,108
20	149-162	165-178	LSFNPTQLEEQCHI	45

P02754 residues 1-16 form the signal peptide. The mature chain corresponds to precursor residues 17-178. Original region identifiers are retained. Fully contained regions are excluded from the primary set; partial overlaps are retained.

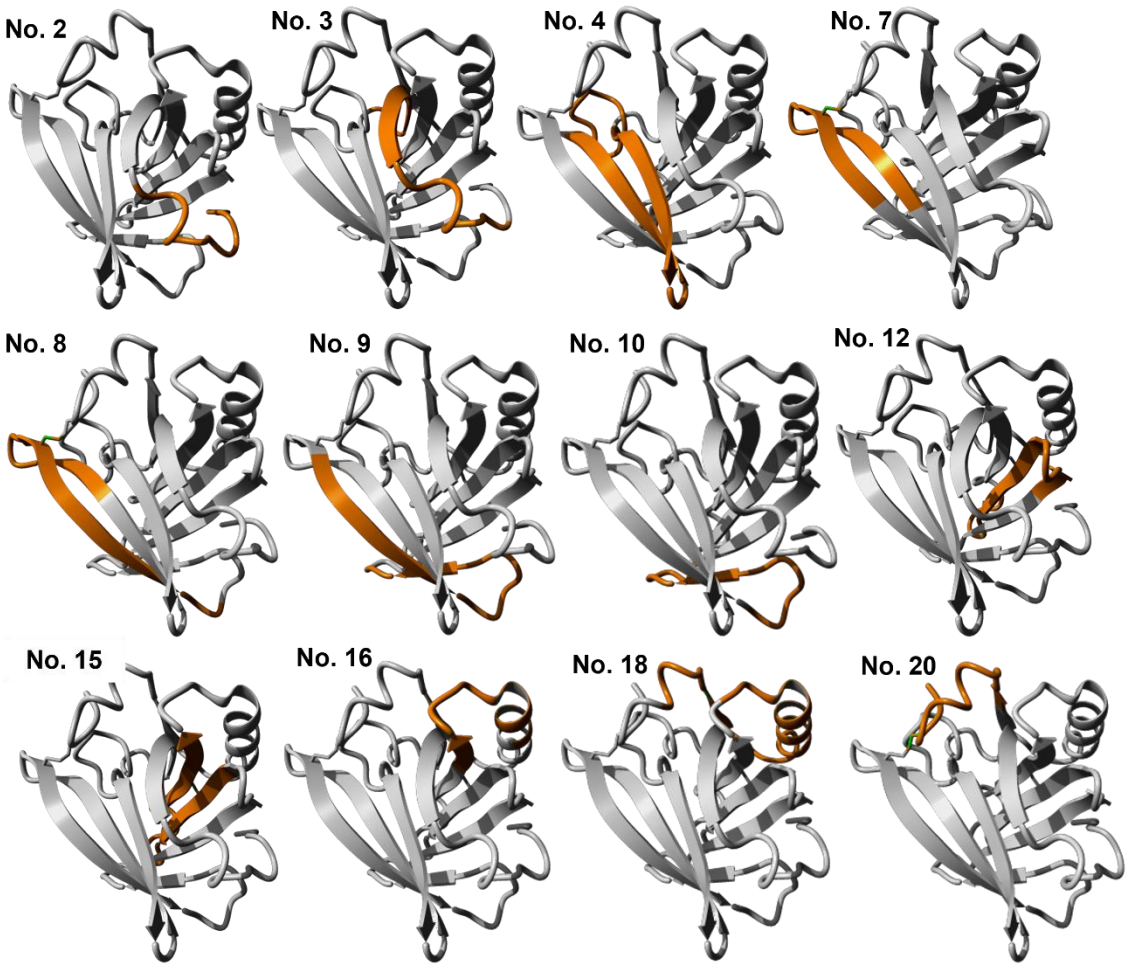

Figure 3-8. Bos d 5 linear IgE-binding regions mapped onto the Bos d 5 structure.

Selected minimally redundant literature-reported Bos d 5 linear IgE-binding regions were mapped onto the Bos d 5 reference structure. The Bos d 5 backbone is shown in grey, and mapped linear regions are highlighted in orange. The selected regions span multiple surface-exposed segments of the β-lactoglobulin fold, including loop, β-strand, and terminal regions.

Table 3-9. Residue composition of the Bos d 5 conformational epitope.

Epitope	Epitope position	Epitope sequence	Reference
Literature	18-20, 43-45, 47, 55, 57, 59, 65-70, 125-127, 153-154, 156-159, 161-162	T18, W19, Y20, V43, E44, E45, K47, E55, L57, Q59, E65, C66, A67, Q68, K69, K70, T125, P126, E127, P153, T154, L156, E157, E158, Q159, H161, I162	(Niemi et al., 2007)
RAW	18-20, 43-45, 47, 55, 57, 59, 65-70, 125-127, 153-155, 156-159, 162	T18, W19, Y20, V43, E44, E45, K47, E55, L57, Q59, E65, C66, A67, Q68, K69, K70, T125, P126, E127, P153, T154, **155Q**, L156, E157, E158, Q159, **H161**, I162	Defined using YASARA based on ≤ 5 Å cutoff
MIN	18-21, 43-45, 47, 55, 57, 59, 65-70, 125-127, 153-155, 156-159, 162	T18, W19, Y20, **21S**, V43, E44, E45, K47, E55, L57, Q59, E65, C66, A67, Q68, K69, K70, T125, P126, E127, P153, T154, **155Q**, L156, E157, E158, Q159, **H161**, I162	Defined using YASARA based on ≤ 5 Å cutoff

Residue positions and identities are shown for the original author-reported Bos d 5 conformational epitope and for the contact-defined epitope extracted from the 2R56 complex using the standardized YASARA contact criterion (≤ 5 Å) under the RAW and MIN protocols. Residues shared between definitions are shown in plain text. Residues absent from the contact-defined epitope relative to the author-reported epitope are highlighted in red, and residues newly included in the contact-defined epitope are bolded.

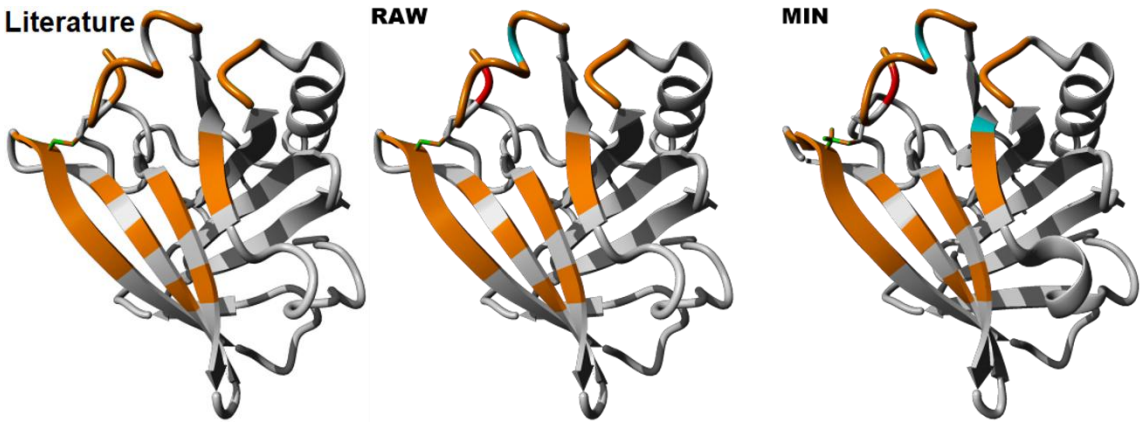

Figure 3-9. Bos d 5 conformational epitope mapped on Bos d 5 structure.

The Bos d 5 structure in the 2R56 IgE Fab–allergen complex is shown as a grey cartoon, with epitope residues highlighted. Literature: original author-reported conformational epitope from the experimental 2R56 complex; RAW and MIN: contact-defined epitope derived using the standardized YASARA contact criterion (≤ 5 Å) under the RAW and MIN protocols, respectively. Residues shared with the author-reported epitope are shown in orange. Residues present in the author-reported epitope but absent from the contact-defined epitope are shown in red, whereas residues newly included in the contact-defined epitope but not reported by the original authors are shown in cyan.

3.3.3 Mapping quality of mapped epitope region comparisons

Mapping quality varied across the selected Bos d 5 linear IgE-binding regions (Figure 3-10). After applying the predefined ≥60% reference-region coverage threshold, most regions retained a moderate-to-high proportion of query mappings. Several regions were retained for most query structures in both RAW and MIN datasets, including epitope No. 3, No. 8, No. 9, No. 10, No. 15, and No. 16. Epitope No. 15 showed the highest retention, with 58/58 RAW mappings and 56/57 MIN mappings retained. In contrast, epitope No. 2 and epitope No. 4 showed lower retention, while epitope No. 20 showed the lowest retention, with only 6/58 RAW mappings and 6/57 MIN mappings retained.

Coverage distributions also differed among regions. Several central regions had coverage values concentrated above the 60% threshold, whereas epitope No. 2, No. 4, and especially No. 20 showed lower or more variable coverage. Epitope No. 20

corresponds to the C-terminal segment of Bos d 5 and showed poor coverage across most query structures. Descriptor comparisons for this region were therefore based on a small, retained subset. Local Cα RMSD values among retained mappings were generally low to moderate for most regions, although high-RMSD outliers were observed. The coverage-versus-RMSD plot showed that mappings passing the coverage threshold could still differ in local structural agreement. This is consistent with the use of RMSD as a structural-alignment metric calculated over aligned residue pairs, where lower RMSD indicates closer structural correspondence.

Mapping quality was also evaluated for the contact-defined Bos d 5 mapped conformational epitope region (MCER) (Figure 3-11). The RAW and MIN reference epitope sets comprised 27 and 28 residues, respectively. After applying the ≥60% coverage threshold, 45/58 RAW mappings and 48/57 MIN mappings were retained. Most retained mappings clustered near or above the coverage threshold in both datasets, although several low-coverage mappings were excluded. For the conformational epitope, local Cα RMSD values were generally low to moderate among retained mappings, but several high-RMSD outliers were present, particularly in the RAW dataset.

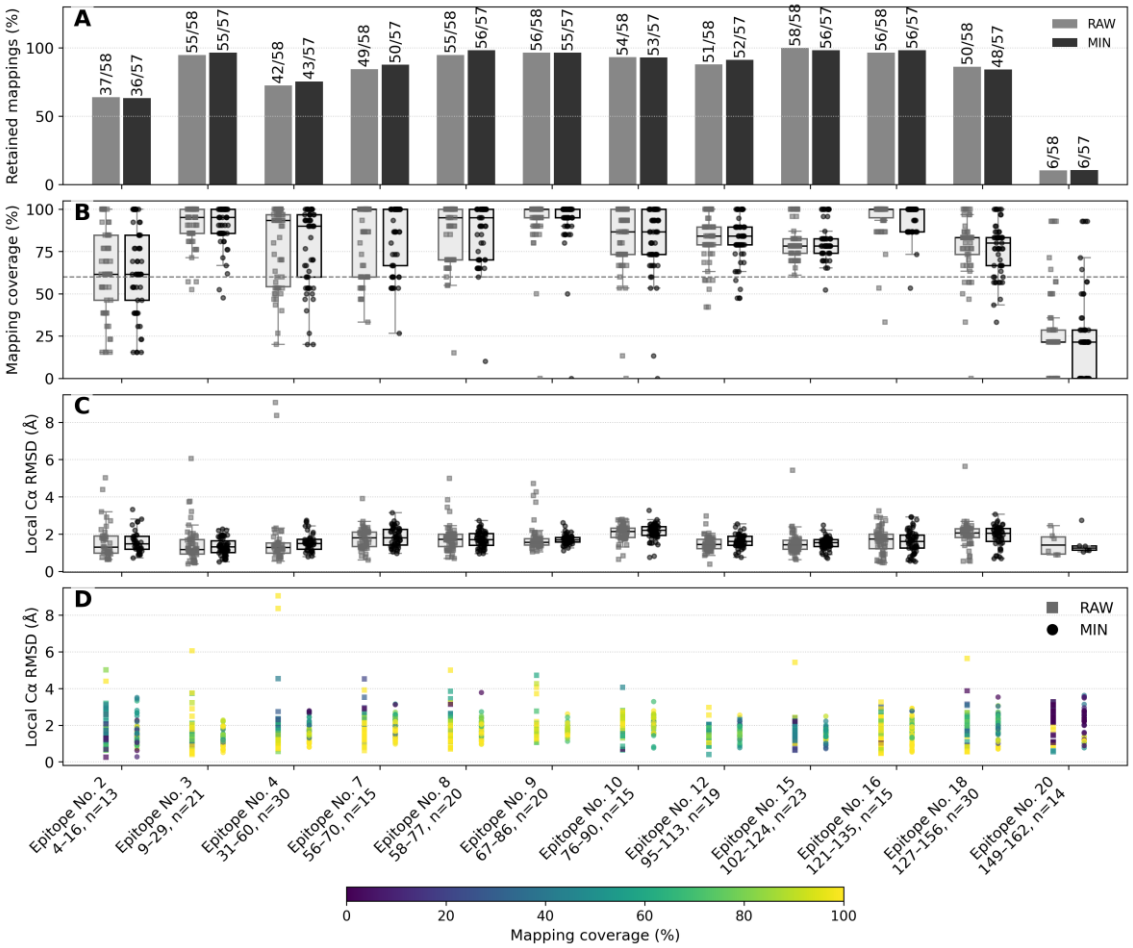

Figure 3-10. Mapping quality for Bos d 5 MLERs across query lipocalins.

Mapping quality was assessed for the selected Bos d 5 mapped linear epitope regions (MLERs) in the RAW and MIN mapped epitope region mapping datasets. (A) Number of RAW and MIN mappings retained after applying the ≥60% reference-region coverage threshold. (B) Distribution of mapping coverage values for each Bos d 5 linear IgE-binding region. (C) Local Cα RMSD distributions for mapped regions. (D) Relationship between mapping coverage and local Cα RMSD. Epitope numbers correspond to the minimally redundant Bos d 5 linear IgE-binding regions used in the primary analysis. RAW and MIN denote protocol-specific outputs of mapped regions.

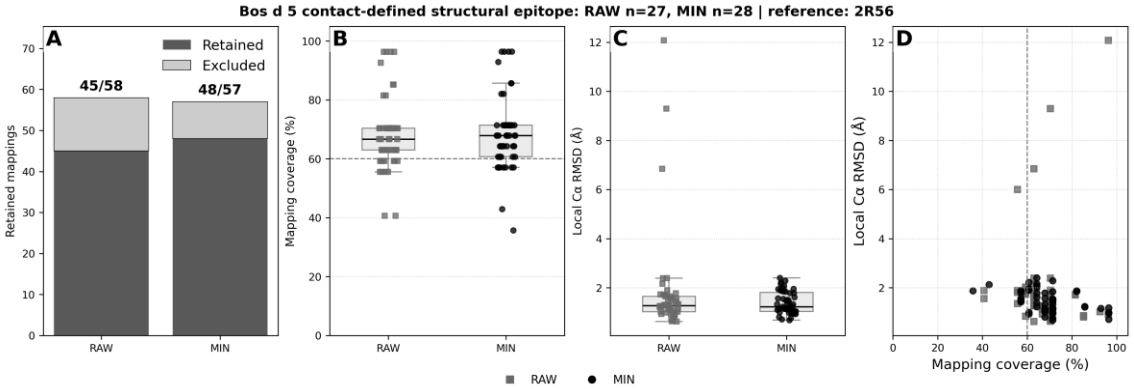

Figure 3-11. Mapping quality for the contact-defined Bos d 5 MCER.

Mapping quality was assessed for the protocol-specific Bos d 5 contact-defined mapped conformational epitope region (MCER) in the RAW and MIN mapped epitope region mapping datasets. The reference epitope contained 27 residues in RAW and 28 residues in MIN because Fab-contact residues were defined after protocol-specific structural preparation. (A) Number of RAW and MIN mappings retained or excluded after applying the ≥60% reference-region coverage threshold. (B) Mapping coverage distributions. (C) Local Cα RMSD distributions. (D) Relationship between mapping coverage and local Cα RMSD. The RAW and MIN reference epitope sets contained 27 and 28 residues, respectively. RAW and MIN denote protocol-specific outputs of mapped regions.

3.3.4 Descriptor correlation structure and redundancy

Descriptor correlation networks showed clear redundancy within several descriptor sets before PCA-based distance analysis. RAW and MIN datasets produced broadly similar correlation structures, indicating that the main descriptor relationships were preserved after minimization. The complete correlation networks are provided in Supplementary Figures S3.1–S3.5.

Across descriptor sets, recurrent correlated modules were observed for surface-area and exposure descriptors, charge and electrostatic descriptors, aromaticity-related descriptors, secondary-structure fractions, and interface-interaction descriptors. In epitope and paratope descriptor sets, SASA-related variables and mapped or contact-residue counts frequently clustered together. Charge descriptors also overlapped with

broader electrostatic descriptors, while aromatic enrichment clustered with related aromaticity measures in several descriptor sets.

Interface descriptor networks showed the strongest redundancy. For individual contact types, interaction counts, densities, mean distances, mean strengths, and summed strengths often formed correlated modules. This pattern was observed for π–π, ionic, hydrophobic, cation–π, and hydrogen-bond interaction descriptors, indicating that the full interface descriptor set contained multiple alternative measures of the same interaction classes.

Representative descriptors were retained to reduce repeated weighting of the same physical properties in downstream analyses (Table 3-10).

Table 3-10. Retained structural and physicochemical descriptors.

Category	Descriptor	Abbreviation	Units	Definition / implementation
Interface	Buried solvent-accessible surface area	ΔSASA	Å²	Buried solvent-accessible surface area computed as the sum of accessible-surface losses on the two partners upon complex formation.
Interface	Hydrogen bonds density	H-bond density	$Å^{-2}$	Number of hydrogen bonds divided by ΔSASA.
Interface	Ionic interactions density	Ionic density	$Å^{-2}$	Number of ionic interactions divided by ΔSASA.
Interface	Hydrophobic interactions density	Hydrophobic density	$Å^{-2}$	Number if hydrophobic interactions divided by ΔSASA.
Interface	CationPi interactions density	Cation-π density	$Å^{-2}$	Number of cation-π interactions divided by ΔSASA.
Interface	PiPi density	π-π density	$Å^{-2}$	Number of π-π interactions divided by ΔSASA.
Paratope, epitope and interface	Aromatic enrichment	Aromatic enrichment	ratio	Aromatic residue fraction of epitope or paratope or interface divided by the aromatic residue fraction in the whole allergen or Fab or complex.
Paratope and epitope	Solvent accessible surface area	SASA	Å²	Solvent-accessible surface area returned by YASARA SurfRes/Surf with standardized surface parameters.
Paratope and epitope	Relative solvent accessibility	RSA	ratio	Sum of residue solvent-accessible areas divided by the sum of residue-specific reference maximally accessible areas.
Paratope and epitope	Electrostatic surface potential	ESP	kJ/mol	YASARA SurfESP patch-level value (unit=All) on the solvent-accessible surface using PME; corresponds to the average energy of a hypothetical +1 charge at the surface positions contributed by the selected unit.
Paratope and epitope	Protrusion index	PI	dimensionless	Ellipsoid-based geometric protrusion score for the allergen patch relative to the full allergen shape.

Descriptors were retained to represent interface footprint, normalized contact chemistry, aromatic enrichment, surface exposure, electrostatic potential, and local shape while reducing redundant measurements of the same physical property.

3.3.5 Exploratory PCA of retained descriptors

PCA was performed separately for each descriptor set using both the full and retained descriptor set. The retained descriptors were used for primary analyses. The biplots and scree plots of the full descriptor set are provided in Supplementary Figures S3.6-S3.22.

3.3.5.1 PCA variance structure and biplots of mapped linear epitope regions (MLERs)

PCA was performed separately for each MLER using the retained Descriptor differences relative to Bos d 5. The first two principal components explained a substantial proportion of the variance across most mapped regions. For mapped epitopes 2–18, PC1 and PC2 together explained 67.5–75.7% of the variance in the RAW dataset and 70.1–76.3% in the MIN dataset (Table 3-11). Epitope 20 showed the highest PC1–PC2 variance capture, with 93.2% in the RAW dataset and 92.9% in the MIN dataset.

The cumulative explained variance (CEV) profiles were highly similar between RAW and MIN datasets (Figure 3-12). In the RAW dataset, the median CEV across MLERs was 42.5% for PC1, 70.6% for PC1–PC2, 90.8% for PC1–PC3, and 100.0% for PC1–PC4. The corresponding MIN values were 41.7%, 72.3%, 91.8%, and 100.0%, respectively. Thus, PC1–PC3 captured more than 90% of the median variance in both datasets, while inclusion of PC4 captured the full retained descriptor variance.

The PCA biplots showed broadly similar score distributions between the RAW and MIN datasets (Figure S3. 5-3.7). Across most MLERs, biological groups highly overlapped in PCA space. Bovine βLG was generally positioned within the central region of the score distribution, and non-bovine ruminant βLGs were usually located near the bovine βLG-associated region. Human lipocalins and aeroallergens showed broader dispersion and contributed several peripheral observations.

Descriptor loadings indicated that the main axes of variation differed among MLERs. SASA and protrusion-index differences frequently contributed to the dominant score-space orientation, whereas ESP and aromatic-enrichment differences showed more region-specific loading patterns. Epitope 4 showed a clear separation between ESP/aromatic-enrichment differences and SASA/protrusion-index differences in both RAW and MIN datasets.

Epitope 20 corresponds to the C-terminal region of Bos d 5 and was poorly represented in more distantly related lipocalin query structures. After coverage filtering, the retained MIN dataset for epitope 20 mostly contained close homologous non-bovine ruminant βLGs. Therefore, although epitope 20 showed high PC1–PC2 variance capture, it was not used to support cross-group interpretation of MLER variation.

3.3.5.2 PCA variance structure and biplots of mapped conformational epitope region (MCER)

PCA was performed on the retained Descriptor differences relative to Bos d 5 for the mapped conformational epitope region (MCER). In the RAW dataset, PC1 explained 39.1% of the variance and PC2 explained 28.1%, giving a combined PC1–PC2 variance of 67.2% (Table 3-12, Figure 3-14). In the MIN dataset, PC1 explained 41.7% and PC2 explained 27.7%, giving a combined PC1–PC2 variance of 69.4% (Table 3-12, Figure 3-14). The CEV profiles were similar between the RAW and MIN datasets (Figure S3. 7). In both datasets, PC1–PC3 captured more than 90% of the variance, while PC1–PC4 captured 100% of the retained descriptor variance. Based on the ≥95% CEV threshold, PC1–PC4 were retained for downstream analyses of the mapped conformational epitope descriptor set (Table 3-12).

The RAW and MIN PCA biplots showed broadly comparable score distributions (Figure 3-13). Biological groups overlapped in PCA space, although human lipocalins were more frequently distributed toward the upper part of the score space, whereas several non-ruminant ungulate βLGs and carnivoran/marsupial βLGs were positioned toward positive PC1. Bovine βLG was located near the central-to-lower region of the score distribution in both datasets. Descriptor loadings showed that Bos d 5-relative RSA and protrusion-index differences contributed mainly along positive PC1, while ESP differences contributed strongly along positive PC2. Protrusion-index and ESP differences were opposed along PC2 in both datasets. Aromatic-enrichment differences showed a more protocol-dependent loading pattern, with a positive PC2 contribution in the RAW dataset and a negative PC1/PC2 orientation in the MIN dataset. Overall, the mapped conformational epitope PCA showed reproducible RAW/MIN variance structure and broad overlap among biological groups.

3.3.5.3 PCA variance structure of composite descriptors

PCA was performed on retained descriptor sets derived from the comparative Fab–allergen complexes, including comparative epitope descriptors, comparative paratope descriptors, comparative interface descriptors, and the composite descriptor set. The proportion of variance represented by PC1 and PC2 differed among descriptor sets (Table 3-10). Comparative paratope descriptors showed the highest PC1–PC2 variance capture, with 75.6% in the RAW dataset and 73.7% in the MIN dataset. Comparative epitope descriptors showed intermediate PC1–PC2 variance capture, with 67.4% in the RAW dataset and 67.6% in the MIN dataset. Comparative interface descriptors showed lower PC1–PC2 variance capture, with 52.6% in the RAW dataset and 52.4% in the MIN dataset. The composite descriptor set showed the lowest PC1–PC2 variance capture, with 43.9% in the RAW dataset and 42.1% in the MIN dataset.

Cumulative explained variance analysis showed that the composite descriptor set required more principal components to reach the ≥95% variance-retention threshold than the individual descriptor sets (Figure S3. 20; Table 3-10). For the composite descriptors, PC1–PC5 captured 74.55% of variance in the RAW dataset and 73.87% in the MIN dataset. PC1–PC10 captured 95.76% and 95.52% of variance in the RAW and MIN datasets, respectively, while all 13 components captured 100% of the retained descriptor variance.

The composite descriptor biplots showed broadly overlapping biological groups in both RAW and MIN datasets (Figure 3-14). Bovine βLG and non-bovine ruminant βLGs were positioned near the central score region, whereas human lipocalins and aeroallergens showed broader dispersion and contributed several peripheral observations. The RAW and MIN score distributions were similar overall, although the MIN dataset showed a clearer spread of human lipocalins toward positive PC1.

Composite descriptor PCA loadings reflected contributions from epitope, paratope, and interface descriptors (Figure 3-14). In the RAW dataset, epitope RSA and protrusion index descriptors loaded toward positive PC1/PC2, while several interface density descriptors loaded toward negative PC1. Fab RSA and Fab protrusion-index descriptors contributed mainly along positive PC1. In the MIN dataset, antigen-side RSA, protrusion index, ESP, SASA, and hydrophobic density descriptors contributed toward positive PC1, whereas several paratope and interface descriptors loaded toward negative PC1 or negative PC2. Overall, the PCA of the composite descriptors showed that the integrated descriptor space captured more distributed variation than the individual paratope, interface, or epitope descriptor sets.

The individual descriptor sets derived from the complex showed consistent RAW/MIN behavior (Supplementary Figures S3.14-S3.17). Paratope descriptors produced the

strongest two-component representation, epitope descriptors showed intermediate compression, and interface descriptors showed more distributed variance across components. These patterns support the use of separate PCA spaces for each descriptor set and region in downstream distance analyses, with the composite descriptor set retained as an integrative representation of Fab–allergen complex variation.

Table 3-11. Variance explained by PC1–PC2 and retained PCs for the mapped linear epitope regions.

Linear epitope No.	RAW PC1 + PC2 (%)	MIN PC1 + PC2 (%)	PCs retained for 95% CEV
2	75.6	76.3	PC1-PC4
3	73.2	76.1	PC1-PC4
4	69.1	72.4	PC1-PC4
7	75.7	74.8	PC1-PC4
8	70.7	73.9	PC1-PC4
9	70.2	72.4	PC1-PC4
10	67.5	72.1	PC1-PC4
12	70.5	70.1	PC1-PC4
15	71.6	71.5	PC1-PC4
16	74.1	73.7	PC1-PC4
18	69.8	70.7	PC1-PC4
20	93.2	92.9	PC1-PC3

For each Bos d 5 MLER, the table reports the percentage of variance explained by PC1 and PC2 in the RAW and MIN datasets. The final column reports the principal components retained according to the CEV threshold used for downstream analyses.

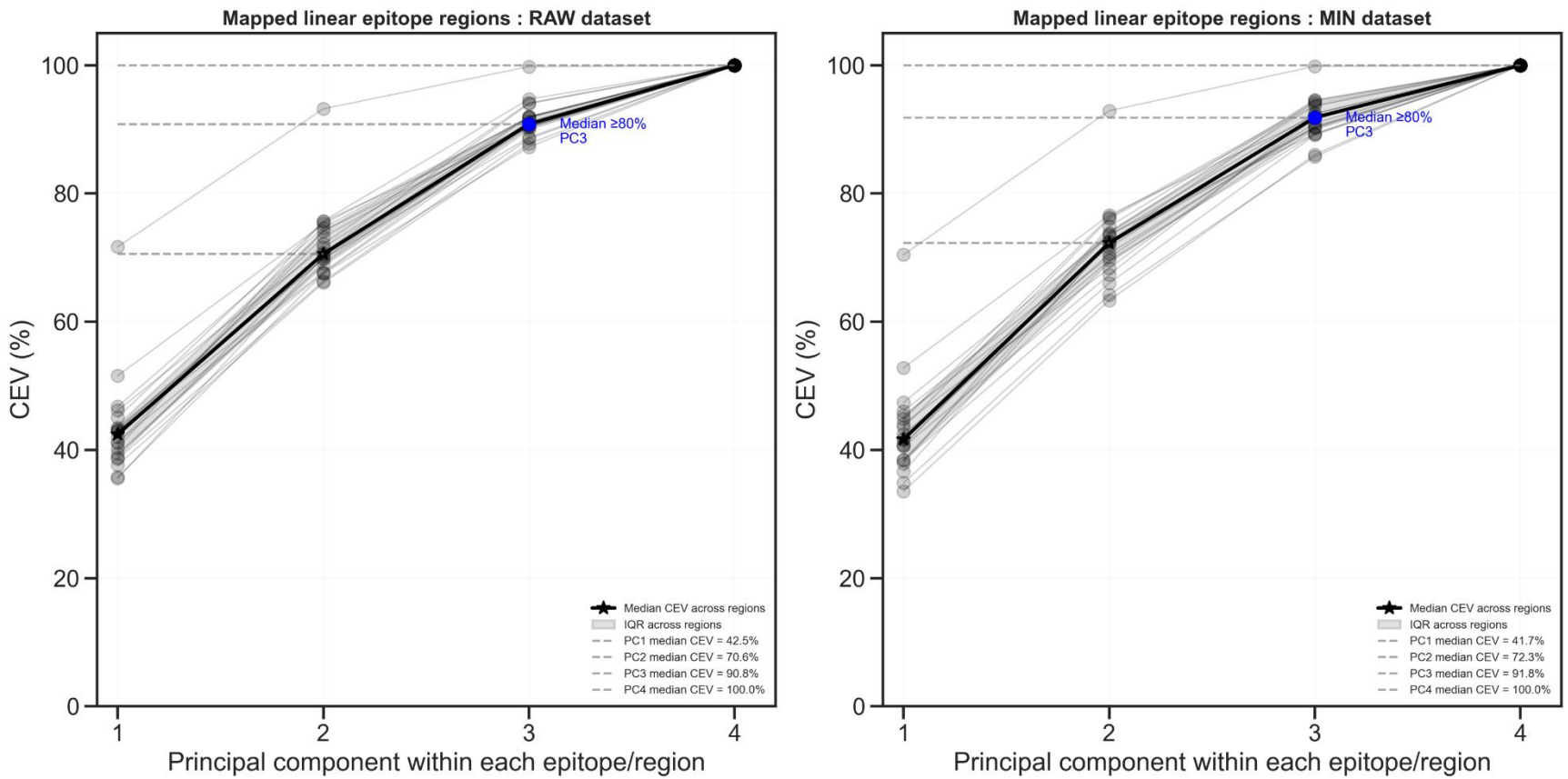

Figure 3-12. CEV of PCA models fitted to retained MLER descriptors.

Cumulative explained variance (CEV) is shown for PCA models fitted separately to each mapped linear epitope region (MLER) in the RAW and MIN datasets. Grey lines represent individual mapped regions, and the black line indicates the median CEV across regions. In both datasets, PC1–PC3 captured more than 90% of the median variance, while PC1–PC4 captured 100% of the retained descriptor variance.

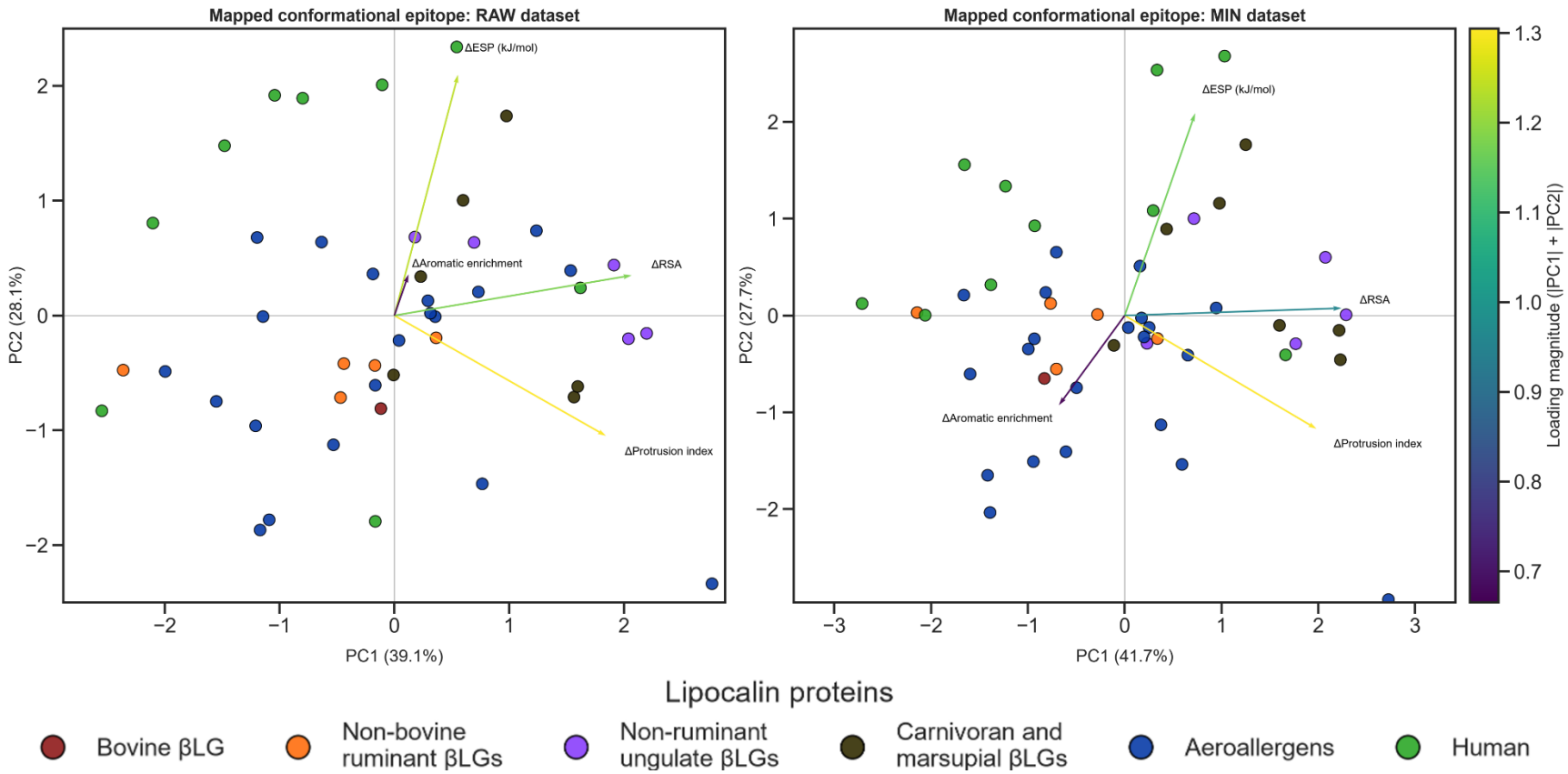

Figure 3-13. PCA biplots of retained descriptor differences for MCER.

PCA biplots are shown for the RAW and MIN mapped conformational epitope region (MCER) datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate loadings for the retained Descriptor differences relative to Bos d 5, including ESP, RSA, protrusion index, and aromatic enrichment. Axis labels show the percentage of variance explained by PC1 and PC2.

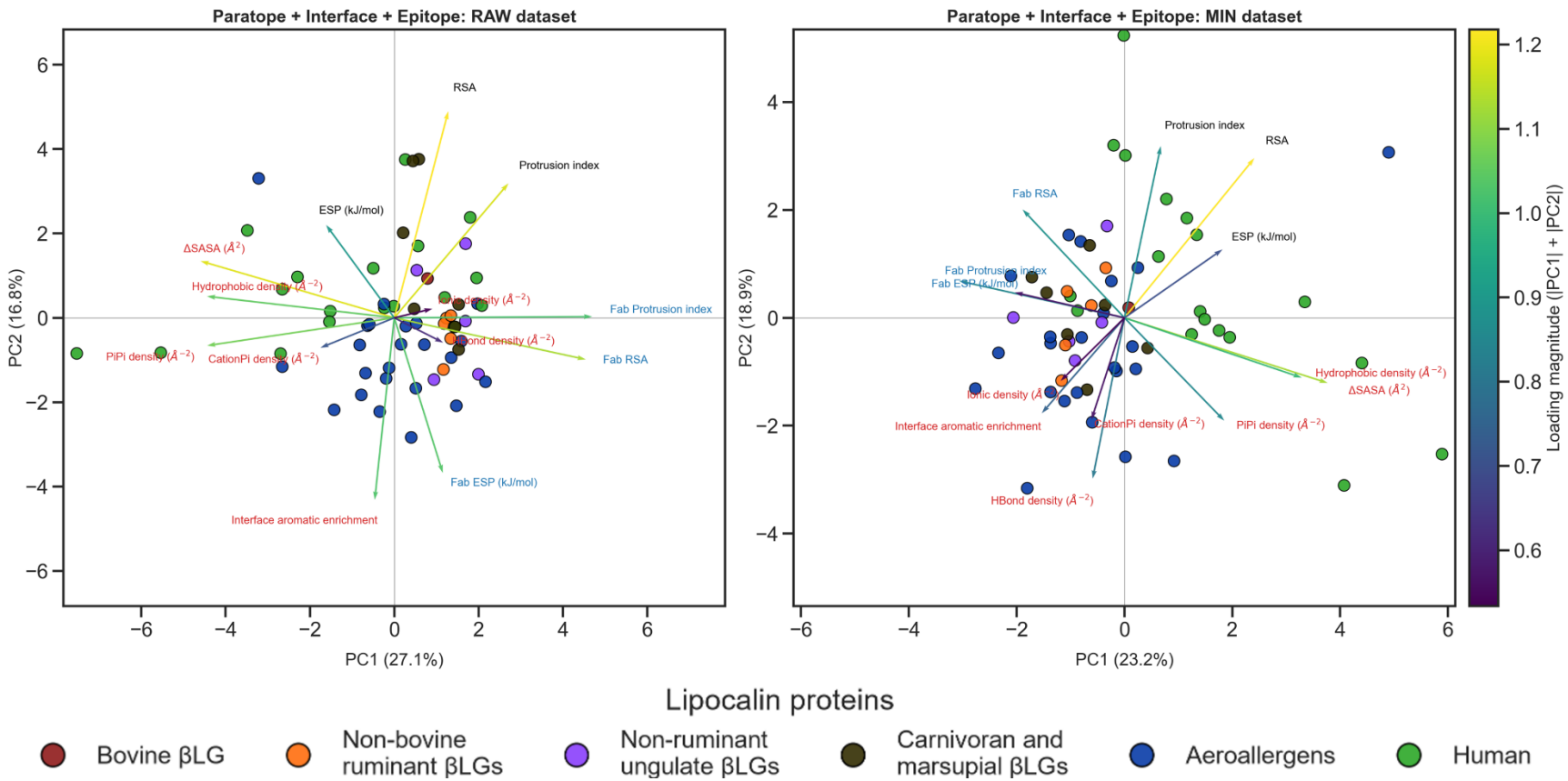

Figure 3-14. PCA biplots of retained composite descriptors.

PCA biplots are shown for the retained composite descriptor set in the RAW and MIN datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate descriptor loadings from comparative epitope, paratope, and interface descriptor classes. Axis labels show the percentage of variance explained by PC1 and PC2.

Table 3-12. PCA variance summary for retained descriptor sets.

Descriptor set	Protocol	PC1 (%)	PC2 (%)	PC1 + PC2 (%)	PCs retained for 95% CEV
Interface	RAW	30.5	22.1	52.6	PC1 - PC6
Interface	MIN	29.3	23.1	52.4	PC1 - PC6
Epitope	RAW	38.6	28.8	67.4	PC1 - PC4
Epitope	MIN	42.4	25.2	67.6	PC1 - PC4
Paratope	RAW	40.5	35.1	75.6	PC1 - PC5
Paratope	MIN	37.1	36.6	73.7	PC1 - PC6
Composite descriptors	RAW	27.1	16.8	43.9	PC1 - PC10
Composite descriptors	MIN	23.2	18.9	42.1	PC1 - PC11
Mapped conformational epitope	RAW	39.1	28.1	67.2	PC1 - PC4
Mapped conformational epitope	MIN	41.7	27.7	65.8	PC1 - PC4

PC1, PC2, cumulative PC1 + PC2 variance, and the number of PCs required to retain 95% variance are shown for each descriptor set and protocol. PCA, principal component analysis; PC, principal component; CEV, cumulative explained variance.

3.3.6 Nearest-neighbor ranking based on Euclidean distance to Bos d 5 in PCA space

3.3.6.1 Overview of nearest-neighbor ranking analyses

Rankings were generated separately for each descriptor set and protocol using ascending Euclidean distance to the matched Bos d 5 reference profile in PCA space. Full rankings are provided in Supplementary Tables S3.1–S3.6.

3.3.6.2 MLER rankings

Nearest-neighbor rankings were generated for each Bos d 5 mapped linear IgE-binding region using Euclidean distances in the retained reference-relative descriptor PCA space. Rankings were calculated separately for the RAW and MIN protocols after applying the predefined ≥60% structural coverage threshold. Full rankings are provided in Supplementary Table S3.1, while Table 3-13 summarizes the top ten ranked aeroallergens.

Across the 12 mapped linear epitope regions (MLERs), non-bovine ruminant β-lactoglobulins were the group most consistently ranked nearest to Bos d 5 (Table 3-13; Figure S3. 26-S3.30). They ranked first in 8 of 12 RAW analyses and 7 of 12 MIN analyses, and accounted for most top-five placements in both protocols. This pattern was most evident for epitopes No. 2, No. 3, No. 7, No. 9, No. 12, No. 15, No. 18, and No. 20.

Several regions showed descriptor-specific departures from this pattern. Epitope No. 4 showed mixed nearest-neighbor placement, including an aeroallergen as the closest-ranked structure in the MIN protocol. Epitope No. 8 was closest to a carnivoran or marsupial β-lactoglobulin in both protocols, while epitopes No. 10 and No. 15 showed closest placements among human lipocalins. These results indicate that local similarity to Bos d 5 varied among individual mapped linear regions.

Aeroallergens were distributed across a broad range of ranks (Table 3-13; Figure S3. 26-S3.30). They appeared within the top ten for most mapped regions, but top-five aeroallergen ranks were limited to selected epitope/protocol combinations, including epitopes No. 2, No. 3, No. 4, No. 7, No. 9, No. 12, and No. 16. No aeroallergen was ranked within the top ten for epitope No. 20 in either protocol. Epitope No. 20 corresponds to the C-terminal region of Bos d 5 and, after filtering, retained only close homologous non-bovine ruminant β-lactoglobulins.

3.3.6.3 MCER rankings

Nearest-neighbor rankings were generated for the mapped conformational epitope region (MCER) using Euclidean distances in the retained reference-relative descriptor-difference PCA space. Rankings were calculated separately for the RAW and MIN protocols after applying the ≥60% structural coverage threshold. The analysis retained 45 RAW mappings and 48 MIN mappings. Full rankings are provided in Supplementary Table S3.2, while Table 3-14 shows the ten closest query structures.

Non-bovine ruminant β-lactoglobulins occupied the closest ranks to Bos d 5 in both protocols (Figure 3-15;Table 3-14). In the RAW protocol, the three nearest structures were non-bovine ruminant β-lactoglobulins, with PCA distances of 0.54, 0.68, and 0.82. A fourth non-bovine ruminant β-lactoglobulin was ranked fifth. In the MIN protocol, non-bovine ruminant β-lactoglobulins occupied ranks 1 and 2, with PCA distances of 0.32 and 0.89, and two additional non-bovine ruminant β-lactoglobulins were ranked fifth and sixth.

Aeroallergens were also represented among the ten nearest structures, but generally after the closest non-bovine ruminant β-lactoglobulins. They occupied ranks 4, 6, 7, 9, and 10 in the RAW protocol and ranks 3, 7, 8, 9, and 10 in the MIN protocol. One carnivoran or marsupial β-lactoglobulin appeared among the top ten in each protocol.

Human lipocalins and non-ruminant ungulate β-lactoglobulins were not present among the ten nearest structures in either protocol.

3.3.6.4 Rankings based on composite descriptors

Nearest-neighbor rankings were generated from descriptor sets derived from the comparative Fab–query allergen models, including the comparative epitope descriptors, comparative paratope descriptors, comparative interface descriptors, and the composite descriptor set. Rankings were calculated separately for the RAW and MIN protocols using Euclidean distances to the Bos d 5 reference profile in the retained PCA space. Full rankings are provided in Supplementary Table S3.3-S3.6, while Table 3-15 shows the ten closest query structures for each descriptor set and protocol.

The comparative interface descriptor rankings showed the strongest protocol-dependent pattern (Table 3-15; Figure 3-16). In the RAW protocol, the closest structure was a non-bovine ruminant β-lactoglobulin, followed by three human lipocalins at rank 2 - 4. Aeroallergens first appeared at ranks 5 and 6. In the MIN protocol, the three closest structures were non-bovine ruminant β-lactoglobulins, followed by an aeroallergen at rank 4. Human lipocalins appeared at ranks 6 and 8.

The comparative epitope rankings showed the most consistent enrichment of non-bovine ruminant βLGs among the closest structures. In the RAW protocol, non-bovine ruminant βLGs occupied ranks 1 and 2 and were also present at ranks 6–8. In the MIN protocol, the four closest structures were all non-bovine ruminant β-lactoglobulins. Aeroallergens and human lipocalins were present among the top ten in both protocols but did not occupy the closest ranks.

Comparative paratope descriptor rankings also placed several non-bovine ruminant βLGs close to Bos d 5. In the RAW protocol, a non-ruminant ungulate β-lactoglobulin

ranked first, followed by two non-bovine ruminant βLGs. Non-bovine ruminant βLGs accounted for five of the ten closest RAW paratope ranks. In the MIN protocol, non-bovine ruminant β βLGs occupied ranks 1, 2, and 4, while aeroallergen, human, non-ruminant ungulate, and carnivoran/marsupial lipocalins were also represented among the top ten.

The composite descriptor space showed a clear difference between RAW and MIN rankings. In the RAW protocol, the closest structure was a non-bovine ruminant βLG, but human lipocalins occupied ranks 2 and 3, and aeroallergens were also present among the top ten. In the MIN protocol, non-bovine ruminant βLGs occupied the first five ranks, followed by human lipocalins and aeroallergens at higher distances. Overall, the MIN protocol produced stronger nearest-neighbor enrichment of non-bovine ruminant β-lactoglobulins across the interface, epitope, paratope, and combined descriptor rankings.

3.3.6.5 Ranking concordance across descriptor set

Comparison of nearest-neighbor rankings across descriptor sets showed that non-bovine ruminant β-lactoglobulins were the most consistently recovered structures closest to Bos d 5. This pattern was observed across the MLERs, MCER, comparative epitope, paratope, interface, and the composite descriptor set (Figure 3-15-Figure 3-16; Figure S3. 26-S3.35; Table 3-13-Table 3-15).

The strongest concordance was observed for epitope descriptor sets. In the MLER analysis, non-bovine ruminant βLGs ranked first in most RAW and MIN rankings. The recurrent closest homologous structures included AF-P67975-F1 (mouflon βLG), AF-P02755-F1 (bβLG), 1YUP (rβLG), 4TLJ (gβLG), and 4CK4 (sβLG). A similar pattern was observed for the MCER, where AF-P67975-F1, AF-P02755-F1, 1YUP, and 4TLJ occupied the closest ranks. Comparative epitope descriptors also ranked non-bovine

ruminant βLGs as the closest structures, with AF-P67975-F1 ranked first in both RAW and MIN protocols. Epitope No. 20 was more restricted after coverage filtering and retained only close homologous non-bovine ruminant βLGs among the top-ranked structures.

Comparative paratope and interface descriptor sets showed greater rank variation. Paratope rankings still recovered several non-bovine ruminant βLGs, including 4TLJ, 4CK4, 1YUP, AF-P02755-F1, and AF-P67975-F1, among the top ten. However, other lipocalin groups also appeared, particularly in the MIN protocol, where 4N7C ranked third. Interface rankings were more protocol-dependent. In the RAW protocol, 1YUP ranked first, but several human lipocalins and aeroallergens were also present among the closest structures. In contrast, the MIN protocol placed 1YUP, 4CK4, and 4TLJ in the first three ranks, indicating stronger recovery of close βLG homologs after filtering.

The composite descriptor rankings reflected this protocol dependence. In the RAW protocol, 1YUP ranked first, but human lipocalins and aeroallergens also appeared within the top ten. The aeroallergens included 8AMC (cat Fel d 4.0101), AF-D7PBH4-F1 (dog Can f 4.0101), 4ODD (dog Can f 4), and 1I06 (mouse Mus m 1.0101). In contrast, the MIN protocol placed non-bovine ruminant β-lactoglobulins in the first five ranks, including 1YUP, 4TLJ, 4CK4, AF-P67975-F1, and AF-P02755-F1. Aeroallergens were still present in the MIN combined ranking, but only at lower positions, represented by 1I06 (mouse Mus m 1.0101) and 8EPV (cat Fel d 7.0101).

Aeroallergens appeared repeatedly across descriptor sets, but they were less consistently ranked nearer than non-bovine ruminant βLGs. In the MLER analysis, the strongest aeroallergen enrichment was observed for epitope No. 4, where 8AEJ (dog Can f 6.0101) ranked first in the MIN protocol and AF-A0A3Q2HX90-F1 (horse Equ c

1.0102), 8AEJ (dog Can f 6.0101), 1EW3 (horse Equ c 1.0101), 8AMC (cat Fel d 4.0101), AF-S0BDX9-F1 (Cav p 6.0101), 2A2U (rat Rat n 1.0101), and AF-S5ZYD3-F1 (Siberian hamster Phod s 1.0101) appeared among the top ten. Epitopes No. 9 and No. 12 also showed repeated aeroallergen placements, including AF-D7PBH4-F1 (dog Can f 4.0101), 2NND (mouse Mus m 1.0102), 1I06 (mouse Mus m 1.0101), AF-S0BDX9-F1 (Cav p 6.0101), 2A2U (rat Rat n 1.0101), AF-A0A3Q2HX90-F1 (horse Equ c 1.0102), and 8AEJ (dog Can f 6.0101).

Across the mapped linear epitope descriptors, mapped conformational epitope descriptors, and comparative Fab–query model descriptor sets, the most recurrent aeroallergen structures were 8AMC (cat Fel d 4.0101), AF-D7PBH4-F1 (dog Can f 4.0101), 1I06 (mouse Mus m 1.0101), 8EPV (cat Fel d 7.0101), and AF-A0A484HRI4-F1 (Cav p 1.0201). These structures appeared in more than one descriptor set and were therefore the most consistent aeroallergen candidates in the cross-region comparison. Other aeroallergens, including AF-S5ZYD3-F1 (Phod s 1.0101), 1EW3 (horse Equ c 1.0102), 4ODD, 8A0D, 2A2U (rat Rat n 1.0101), 8AEJ (dog Can f 6.0101), and AF-S0BDX9-F1 (Cav p 6.0101), showed region-specific proximity but were less broadly ranked as the closest query structures.

Human lipocalins also appeared among the nearest neighbors in selected descriptor sets, particularly the RAW interface descriptor set, the RAW composite descriptor set, and several comparative paratope and comparative epitope descriptor sets. The recurrent human structures included 4O9S, AF-Q8WX39-F1, 3CBC, 2XST, 4R0B, and 2HZQ. This pattern suggests that these descriptor sets captured broader lipocalin similarity. In contrast, human lipocalins did not show the same consistency across descriptor sets as the non-bovine ruminant β-lactoglobulins.

Overall, comparison across descriptor sets showed that similarity to Bos d 5 was strongest and most reproducible for close β-lactoglobulin homologs. Aeroallergens were repeatedly recovered across several descriptor sets, particularly 8AMC (cat Fel d 4.0101), AF-D7PBH4-F1 (dog Can f 4.0101), 1I06 (mouse Mus m 1.0101), 8EPV (cat Fel d 7.0101), and AF-A0A484HRI4-F1 (Cav p 1.0201), but their proximity was more descriptor-set-dependent.

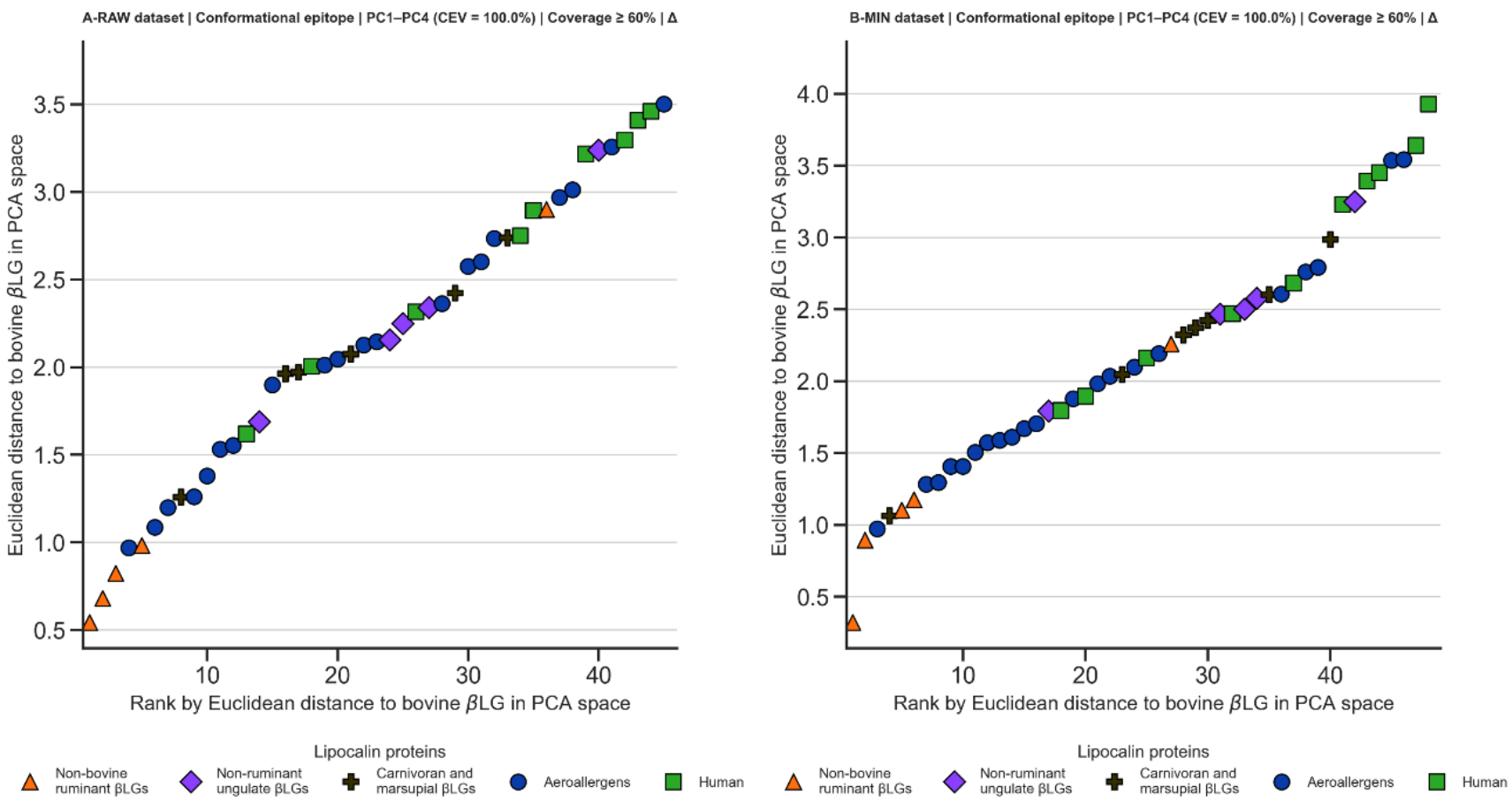

Figure 3-15. Nearest-neighbor rankings for the MCER.

Rankings for mapped conformational epitope region (MCER). Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in retained PCA space for the RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater similarity to Bos d 5.

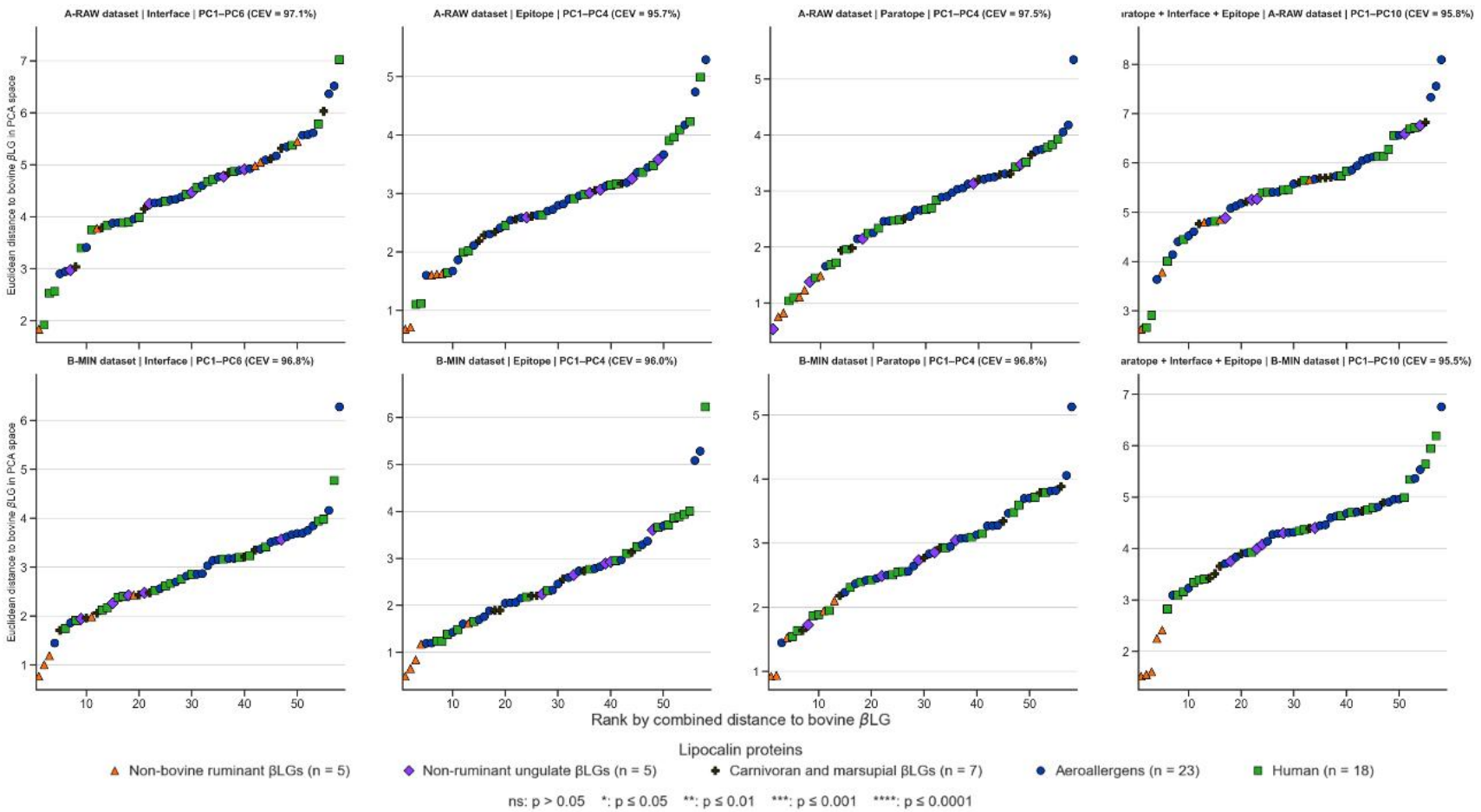

Figure 3-16. Nearest-neighbor rankings for composite descriptor sets.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in retained PCA space. Rankings are shown for interface, interface, epitope, paratope, and composite descriptor sets under RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater descriptor-space similarity to Bos d 5.

Table 3-13. Aeroallergens among the ten closest query structures for each mapped linear epitope region.

Epitope No. 2						Epitope No. 3					
RAW protocol			MIN protocol			RAW protocol			MIN protocol		
Rank	PDB/ AFDB ID	d_{PCA}	Rank	PDB/ AFDB ID	d_{PCA}	Rank	PDB/ AFDB ID	d_{PCA}	Rank	PDB/ AFDB ID	d_{PCA}
6	8AMC	1.63	5	8AMC	1.30	4	AF-P83508-F1	0.75	7	4WFV	1.29
7	AF-A0A3Q2HX90-F1	2.28	8	AF-A0A3Q2HX90-F1	2.47	6	8AEJ	0.77	8	AF-P83508-F1	1.31
			9	AF-S0BDX9-F1	2.54	7	AF-A0A3Q2HX90-F1	0.95			
						8	4WFV	1.06			
Epitope No. 4						**Epitope No. 7**					
2	AF-A0A3Q2HX90-F1	1.03	1	8AEJ	1.29	6	1I06	1.27	5	1I06	1.24
3	8AEJ	1.17	2	AF-S0BDX9-F1	1.42	9	2A2U	1.50	7	2A2U	1.34
5	1EW3	1.38	5	AF-A0A3Q2HX90-F1	1.51	10	AF-Q9QXU1-F1	1.52	8	AF-Q9QXU1-F1	1.35
6	8AMC	1.46	7	2A2U	1.60				10	8EPV	1.51
			10	AF-S5ZYD3-F1	1.65						
Epitope No. 8						**Epitope No. 9**					
	None in top 10		7	AF-U6C8D6-F1	1.18	5	AF-D7PBH4-F1	1.02	3	4N7C	0.90
						7	2NND	1.15	4	AF-D7PBH4-F1	0.99
						9	1I06	1.26	7	2NND	1.32
						10	AF-S0BDX9-F1	1.29	9	1I06	1.34
Epitope No. 10						**Epitope No. 12**					
6	8AMC	1.42	7	8AMC	1.43	4	2A2U	0.62	4	AF-S0BDX9-F1	1.04
8	AF-D7PBH4-F1	1.52	8	AF-Q9QXU1-F1	1.54	7	AF-S0BDX9-F1	1.07	6	2A2U	1.09
			10	AF-D7PBH4-F1	1.63	9	AF-A0A3Q2HX90-F1	1.33	8	AF-A0A3Q2HX90-F1	1.34
						10	8AEJ	1.39			
Epitope No. 15						**Epitope No. 16**					
10	8AEI	1.02		*None in top 10*		6	1I06	1.21	4	8EPV	1.42
						7	2NND	1.24	9	1I06	1.95
						8	AF-A0A484HRI4-F1	1.33	10	AF-A0A484HRI4-F1	1.98
						9	8EPV	1.38			
Epitope No. 18						**Epitope No. 20**					
8	4WFV	1.29	7	4ODD	1.32		*None in top 10*			*None in top 10*	
10	AF-D7PBH4-F1	1.41	8	AF-D7PBH4-F1	1.41						
			10	AF-P83508-F1	1.49						

For each mapped Bos d 5 linear epitope region (MLER), the table reports aeroallergen query structures occurring among the ten closest structures under the RAW and MIN protocols. Rank position and Euclidean distance in the retained PCA space (d_{PCA}) are shown. Lower d_{PCA} values indicate greater similarity to the Bos d 5 reference descriptor profile.

Table 3-14. Ten closest query structures to the Bos d 5 reference for the mapped conformational epitope region.

	RAW protocol			MIN Protocol		
Rank	PDB/ AFDB ID	Lipocalin group	d_{PCA}	PDB/ AFDB ID	Lipocalin group	d_{PCA}
1	AF-P67975-F1	Non-bovine ruminant βLGs	0.54	AF-P02755-F1	Non-bovine ruminant βLGs	0.32
2	AF-P02755-F1	Non-bovine ruminant βLGs	0.68	AF-P67975-F1	Non-bovine ruminant βLGs	0.89
3	1YUP	Non-bovine ruminant βLGs	0.82	AF-A0A484HRI4-F1	Aeroallergens	0.97
4	AF-D7PBH4-F1	Aeroallergens	0.97	AF-P33685-F1	Carnivoran and marsupial βLGs	1.06
5	4TLJ	Non-bovine ruminant βLGs	0.98	4TLJ	Non-bovine ruminant βLGs	1.10
6	8AMC	Aeroallergens	1.09	1YUP	Non-bovine ruminant βLGs	1.17
7	1I06	Aeroallergens	1.20	8A0D	Aeroallergens	1.28
8	AF-P33685-F1	Carnivoran and marsupial βLGs	1.26	8AMC	Aeroallergens	1.29
9	AF-A0A484HRI4-F1	Aeroallergens	1.26	AF-S5ZYD3-F1	Aeroallergens	1.41
10	8EPV	Aeroallergens	1.38	1EW3	Aeroallergens	1.41

For the mapped Bos d 5 conformational epitope region (MCER), the table reports query structures appearing among the closest ranks in the RAW and MIN protocols. Rank position and Euclidean distance in retained PCA space (d_{PCA}) are shown. Lower d_{PCA} values indicate greater similarity to the Bos d 5 reference descriptor profile.

Table 3-15. Ten closest query structures to the Bos d 5 reference for each descriptor set.

	RAW protocol			MIN Protocol		
Rank	PDB/ AFDB ID	Lipocalin group	d_{PCA}	PDB/ AFDB ID	Lipocalin group	d_{PCA}
			Comparative interface ranking			
1	1YUP	Non-bovine ruminant βLGs	1.84	1YUP	Non-bovine ruminant βLGs	0.77
2	4ORR	Human	1.92	4CK4	Non-bovine ruminant βLGs	1.00
3	AF-Q8WX39-F1	Human	2.53	4TLJ	Non-bovine ruminant βLGs	1.19
4	4O9S	Human	2.56	8EPV	Aeroallergens	1.45
5	AF-D7PBH4-F1	Aeroallergens	2.90	AF-Q29146-F1	Carnivoran and marsupial βLGs	1.71
6	AF-S5ZYD3-F1	Aeroallergens	2.94	2HZQ	Human	1.74
7	AF-P07380-F1	Non-ruminant ungulate βLGs	2.97	AF-S5ZYD3-F1	Aeroallergens	1.86
8	AF-Q29146-F1	Carnivoran and marsupial βLGs	3.04	AF-Q8WX39-F1	Human	1.91
9	AF-Q9NPH6-F1	Human	3.40	AF-P07380-F1	Non-ruminant ungulate βLGs	1.94

10	8AMC	Aeroallergens	3.41	AF-P33688-F1	Carnivoran and marsupial βLGs	1.96
			Comparative paratope ranking			
1	3KZA	Non-ruminant ungulate βLGs	0.53	4CK4	Non-bovine ruminant βLGs	0.92
2	4TLJ	Non-bovine ruminant βLGs	0.75	4TLJ	Non-bovine ruminant βLGs	0.94
3	4CK4	Non-bovine ruminant βLGs	0.82	4N7C	Aeroallergens	1.45
4	2XST	Human	1.04	1YUP	Non-bovine ruminant βLGs	1.52
5	4R0B	Human	1.10	AF-Q6JVE9-F1	Human	1.54
6	AF-P02755-F1	Non-bovine ruminant βLGs	1.11	3CBC	Human	1.63
7	AF-P67975-F1	Non-bovine ruminant βLGs	1.23	AF-P33687-F1	Carnivoran and marsupial βLGs	1.64
8	1EXS	Non-ruminant ungulate βLGs	1.38	1EXS	Non-ruminant ungulate βLGs	1.72
9	3CBC	Human	1.45	AF-Q8WX39-F1	Human	1.87
10	1YUP	Non-bovine ruminant βLGs	1.49	2HZQ	Human	1.88
			Comparative epitope ranking			
1	AF-P67975-F1	Non-bovine ruminant βLGs	0.68	AF-P67975-F1	Non-bovine ruminant βLGs	0.49
2	AF-P02755-F1	Non-bovine ruminant βLGs	0.71	4TLJ	Non-bovine ruminant βLGs	0.66
3	4O9S	Human	1.10	4CK4	Non-bovine ruminant βLGs	0.84
4	AF-Q8WX39-F1	Human	1.12	1YUP	Non-bovine ruminant βLGs	1.17
5	AF-A0A484HRI4-F1	Aeroallergens	1.60	AF-A0A484HRI4-F1	Aeroallergens	1.19
6	4TLJ	Non-bovine ruminant βLGs	1.61	1EW3	Aeroallergens	1.20
7	1YUP	Non-bovine ruminant βLGs	1.62	AF-Q8WX39-F1	Human	1.24
8	4CK4	Non-bovine ruminant βLGs	1.63	4O9S	Human	1.24
9	4ES7	Human	1.64	4R0B	Human	1.38
10	8AMC	Aeroallergens	1.67	8A0D	Aeroallergens	1.43
			Composite ranking			
1	1YUP	Non-bovine ruminant βLGs	2.63	1YUP	Non-bovine ruminant βLGs	1.51
2	4O9S	Human	2.66	4TLJ	Non-bovine ruminant βLGs	1.55
3	AF-Q8WX39-F1	Human	2.91	4CK4	Non-bovine ruminant βLGs	1.60
4	8AMC	Aeroallergens	3.64	AF-P67975-F1	Non-bovine ruminant βLGs	2.25
5	4CK4	Non-bovine ruminant βLGs	3.78	AF-P02755-F1	Non-bovine ruminant βLGs	2.41
6	3CBC	Human	4.01	AF-Q8WX39-F1	Human	2.82
7	AF-D7PBH4-F1	Aeroallergens	4.14	1I06	Aeroallergens	3.09
8	4ODD	Aeroallergens	4.40	2HZQ	Human	3.10
9	2XST	Human	4.44	4O9S	Human	3.15
10	1I06	Aeroallergens	4.52	8EPV	Aeroallergens	3.23

For each descriptor set, the ten query structures with the smallest Euclidean distances (d_{PCA}) from the Bos d 5 reference in the retained PCA space are shown separately for the RAW and MIN protocols. Lower d_{PCA} values indicate greater similarity to the Bos d 5 reference descriptor profile.

3.3.7 Statistical testing of group differences

3.3.7.1 Group differences in MLER distances

Group differences in reference-relative PCA distances to Bos d 5 were assessed for each mapped linear epitope region (MLER) using pairwise permutation tests. Results were broadly consistent with the nearest-neighbor rankings, with non-bovine ruminant βLGs generally showing the lowest distance distributions across RAW and MIN protocols (Figure 3-17 - 3-24). Complete pairwise test results are provided in Supplementary Table S3.7.

In the RAW protocol, significant group differences were most frequent for epitopes No. 2, No. 7, No. 8, No. 9, No. 15, No. 16, and No. 18. These differences mainly reflected lower distances for non-bovine ruminant βLGs compared with aeroallergens and/or human lipocalins. Epitopes No. 3, No. 4, and No. 12 showed no significant pairwise separation, while epitope No. 10 showed a more limited difference pattern. The MIN protocol showed a similar overall trend, with significant differences observed for epitopes No. 2, No. 3, No. 7, No. 8, No. 9, No. 10, No. 12, No. 15, No. 16, and No. 18. As in the RAW protocol, significant comparisons were mainly driven by lower distances among non-bovine ruminant βLGs relative to aeroallergens and/or human lipocalins. Epitope No. 4 remained weakly separated, with no significant pairwise differences.

Non-ruminant ungulate β-lactoglobulins and carnivoran/marsupial βLGs generally showed intermediate distance distributions and were less consistently separated from the other groups. Aeroallergens and human lipocalins showed broader distance distributions, with some low-distance observations but less consistent proximity to Bos d 5 across mapped linear regions.

3.3.7.2 Group differences in MCER distances

PCA distances to Bos d 5 were compared among the protein groups for the mapped conformational epitope region (MCER).The resulting group differences were consistent with the nearest-neighbor rankings (Figure 3-19; Table 3-16). In the RAW protocol, non-bovine ruminant βLGs had the lowest median distance to Bos d 5 (median = 0.82), followed by carnivoran and marsupial βLGs (median = 2.02), aeroallergens (median = 2.09), non-ruminant ungulate βLGs (median = 2.25), and human lipocalins (median = 2.89). After FDR adjustment, the difference between non-bovine ruminant β-lactoglobulins and aeroallergens remained significant (FDR-adjusted p = 0.0311).

In the MIN protocol, non-bovine ruminant βLGs again showed the lowest median distance to Bos d 5 (median = 1.10). Aeroallergens showed an intermediate median distance (median = 1.70), whereas human lipocalins showed the highest median distance among the planned comparisons (median = 2.96). After FDR adjustment, human lipocalins were significantly more distant than aeroallergens (FDR-adjusted p = 0.0500) and non-bovine ruminant βLGs (FDR-adjusted p = 0.0500). The difference between non-bovine ruminant βLGs and aeroallergens did not remain significant after FDR correction in the MIN protocol.

3.3.7.3 Group differences in composite distances

PCA distances to Bos d 5 were compared among the protein groups for the comparative interface, epitope, paratope, and comparative composite sets. Planned comparisons focused on aeroallergens, non-bovine ruminant βLGs, and human lipocalins.

For the interface descriptor set, RAW distances showed no significant group differences after FDR correction (Figure 3-20; Table 3-17). In the MIN protocol, non-bovine ruminant βLGs had a lower median distance than both aeroallergens (1.19 versus 3.18; FDR-

adjusted p = 0.0079) and human lipocalins (1.19 versus 2.71; FDR-adjusted p = 0.0089). Aeroallergens and human lipocalins were not significantly different.

For the epitope descriptor set, non-bovine ruminant βLGs had the lowest median distance in both protocols. In RAW, they were significantly closer to Bos d 5 than aeroallergens (1.61 versus 2.80; FDR-adjusted p = 0.0281). In MIN, they were significantly closer than both aeroallergens (0.84 versus 2.33; FDR-adjusted p = 0.0079) and human lipocalins (0.84 versus 3.03; FDR-adjusted p = 0.0089). Aeroallergens and human lipocalins were not significantly different in either protocol. For the paratope descriptor set, non-bovine ruminant βLGs again showed the lowest median distances. In RAW, they were significantly closer to Bos d 5 than aeroallergens (1.11 versus 3.03; FDR-adjusted p = 0.0079), while the comparison with human lipocalins did not remain significant after FDR correction. In MIN, non-bovine ruminant βLGs were significantly closer than both aeroallergens (1.52 versus 3.08; FDR-adjusted p = 0.0285) and human lipocalins (1.52 versus 2.53; FDR-adjusted p = 0.0267).

For the combined paratope, interface, and epitope descriptor space, RAW distances showed no significant group separation after FDR correction. In MIN, non-bovine ruminant βLGs were significantly closer to Bos d 5 than aeroallergens (1.60 versus 4.46; FDR-adjusted p = 0.0079) and human lipocalins (1.60 versus 4.51; FDR-adjusted p = 0.0089). Aeroallergens and human lipocalins were not significantly different.

Together, the permutation tests confirmed that close β-lactoglobulin homologs were the group most consistently closer to Bos d 5. Aeroallergens appeared among top-ranked structures in selected rankings, but their group-level distance distributions did not show the same consistent pattern like non-bovine ruminant β-lactoglobulins.

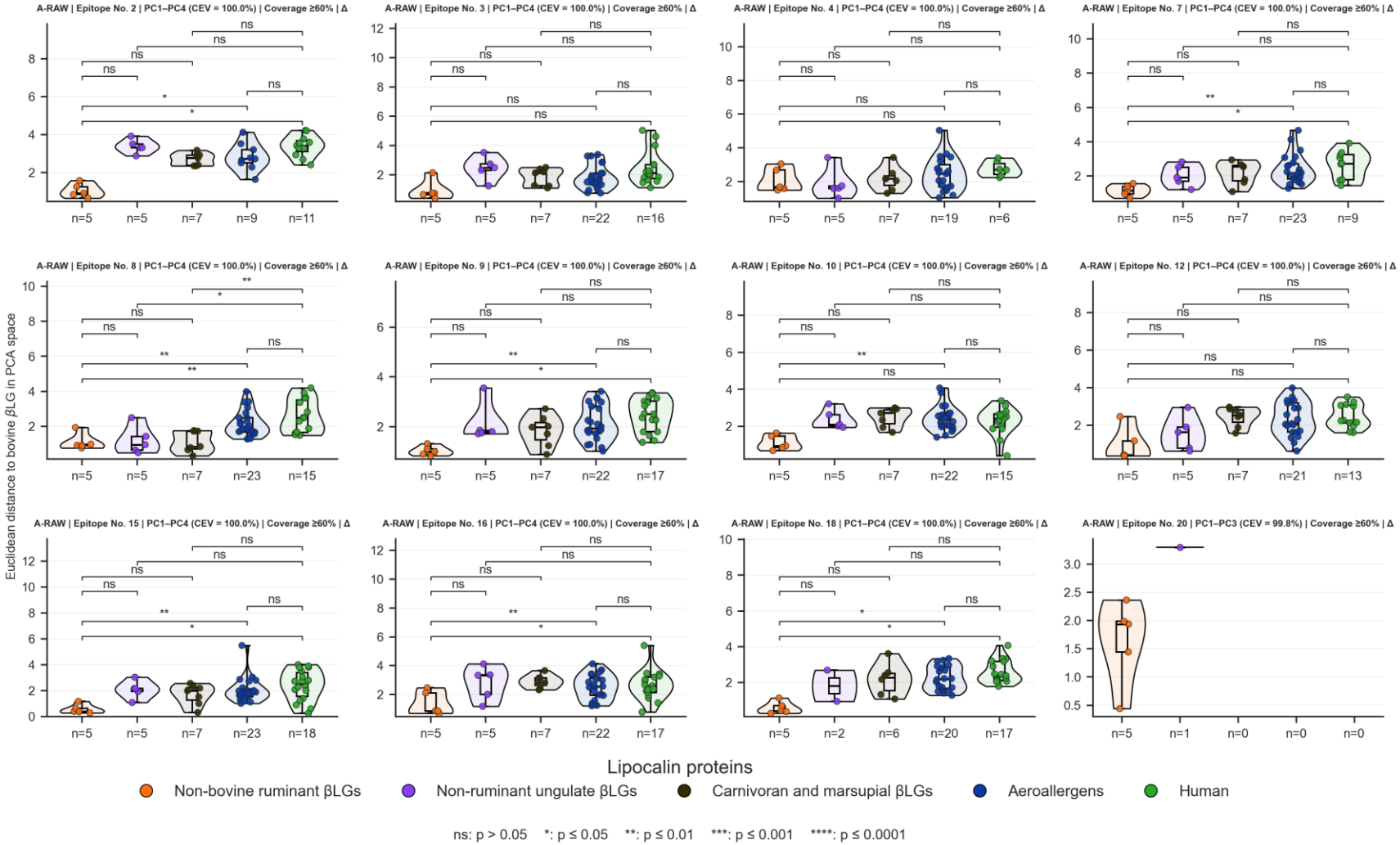

Figure 3-17. Group differences in RAW MLER Euclidean distances to Bos d 5.

Group comparisons of Bos d 5 mapped linear epitope regions (MLERs). Violin and box plots show Euclidean distances from query structures to the matched Bos d 5 reference profile in retained reference-relative descriptor-difference PCA space. Panels correspond to mapped Bos d 5 linear IgE-binding regions. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after multiple-testing correction. Lower distances indicate greater descriptor-space similarity to Bos d 5.

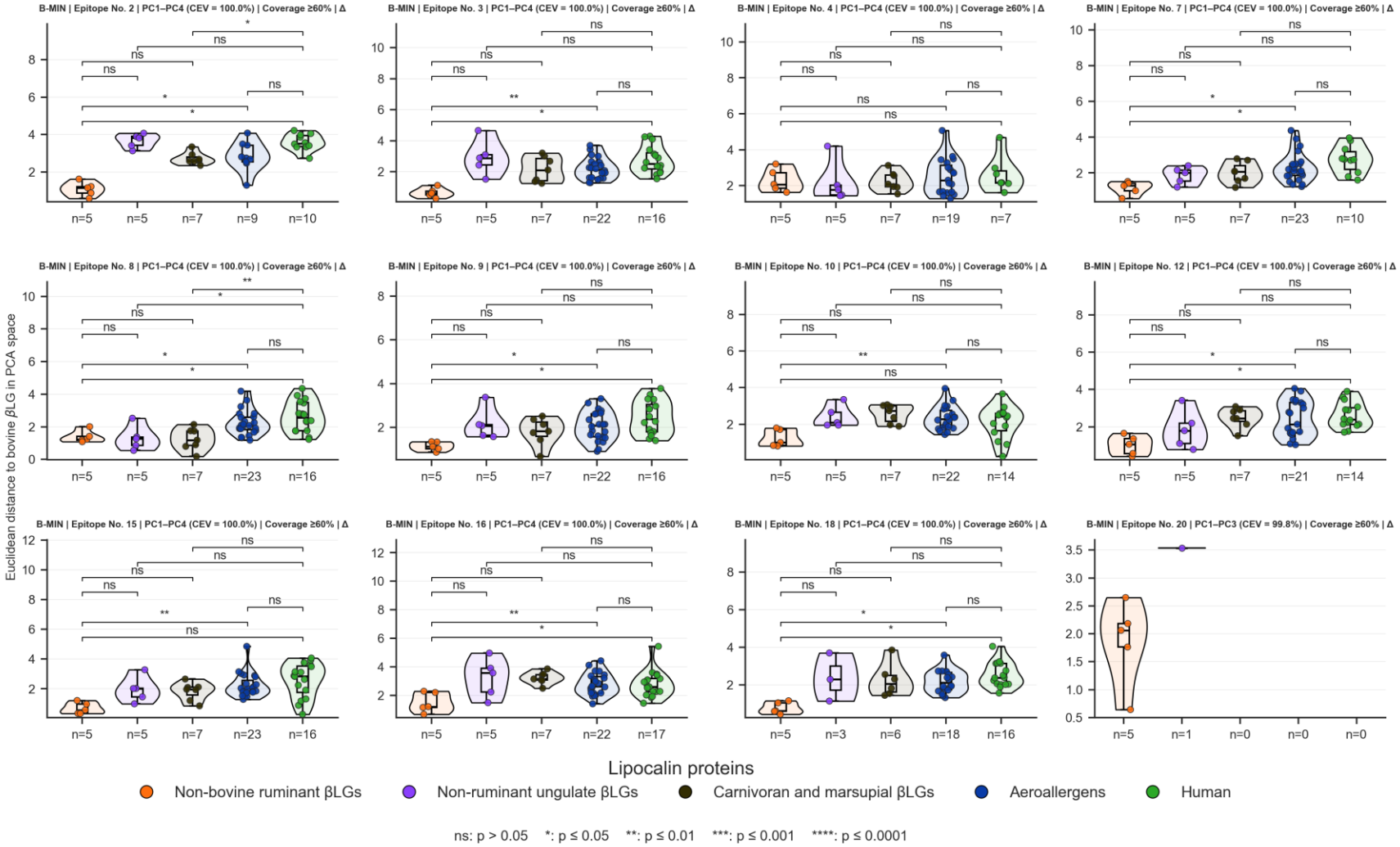

Figure 3-18. Group differences in MIN MLER Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the matched Bos d 5 reference profile in retained reference-relative descriptor-difference PCA space after structural coverage filtering. Panels correspond to mapped Bos d 5 linear IgE-binding regions. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after multiple-testing correction.

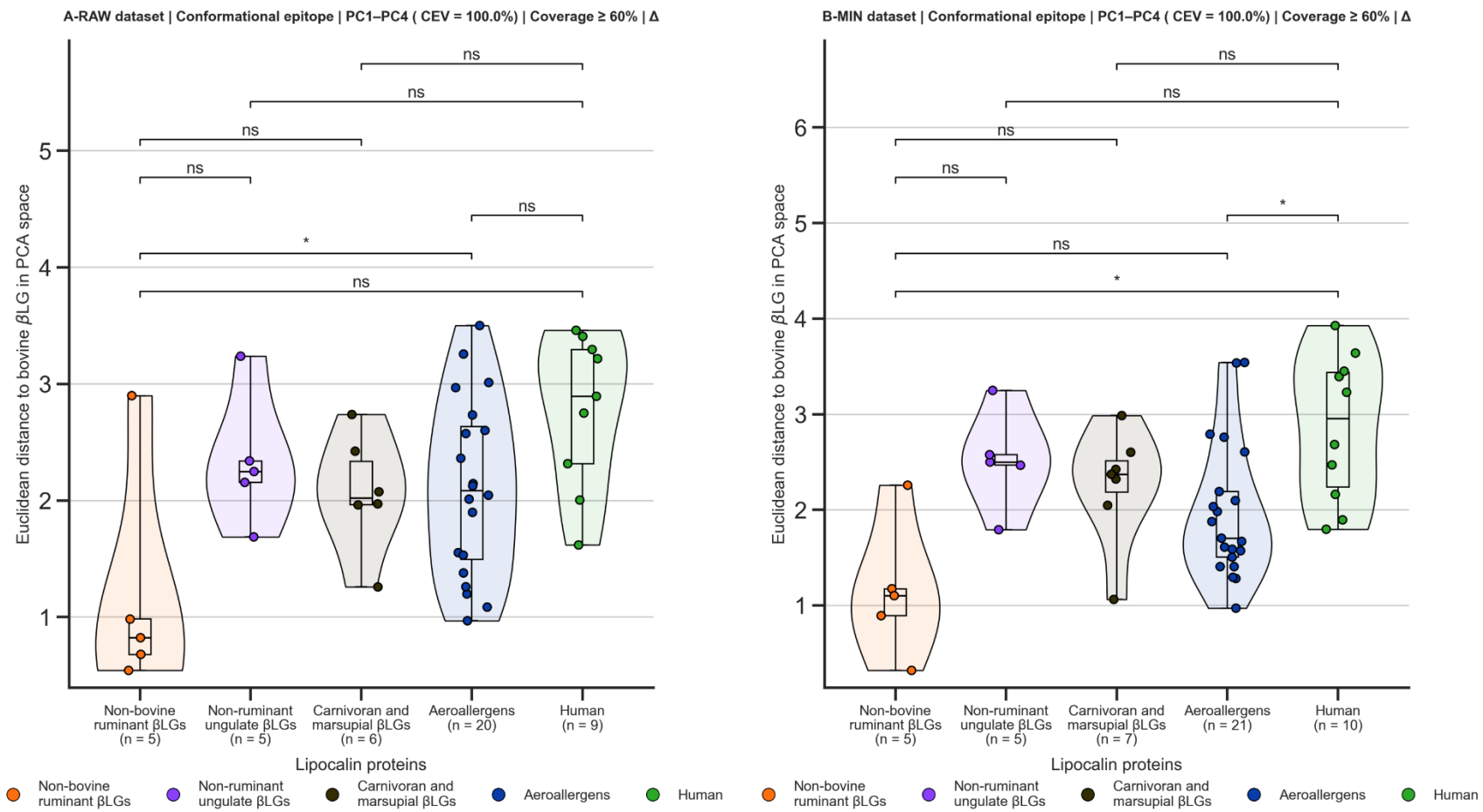

Figure 3-19. Group differences in MCER Euclidean distances to Bos d 5.

Group comparisons of Bos d 5 PCA distances for the mapped conformational epitope region (MCER). Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in retained reference-relative descriptor-difference PCA space for the RAW and MIN protocols. Points are colored by biological group. Brackets indicate planned pairwise permutation-test results after FDR correction. Lower distances indicate greater descriptor-space similarity to Bos d 5.

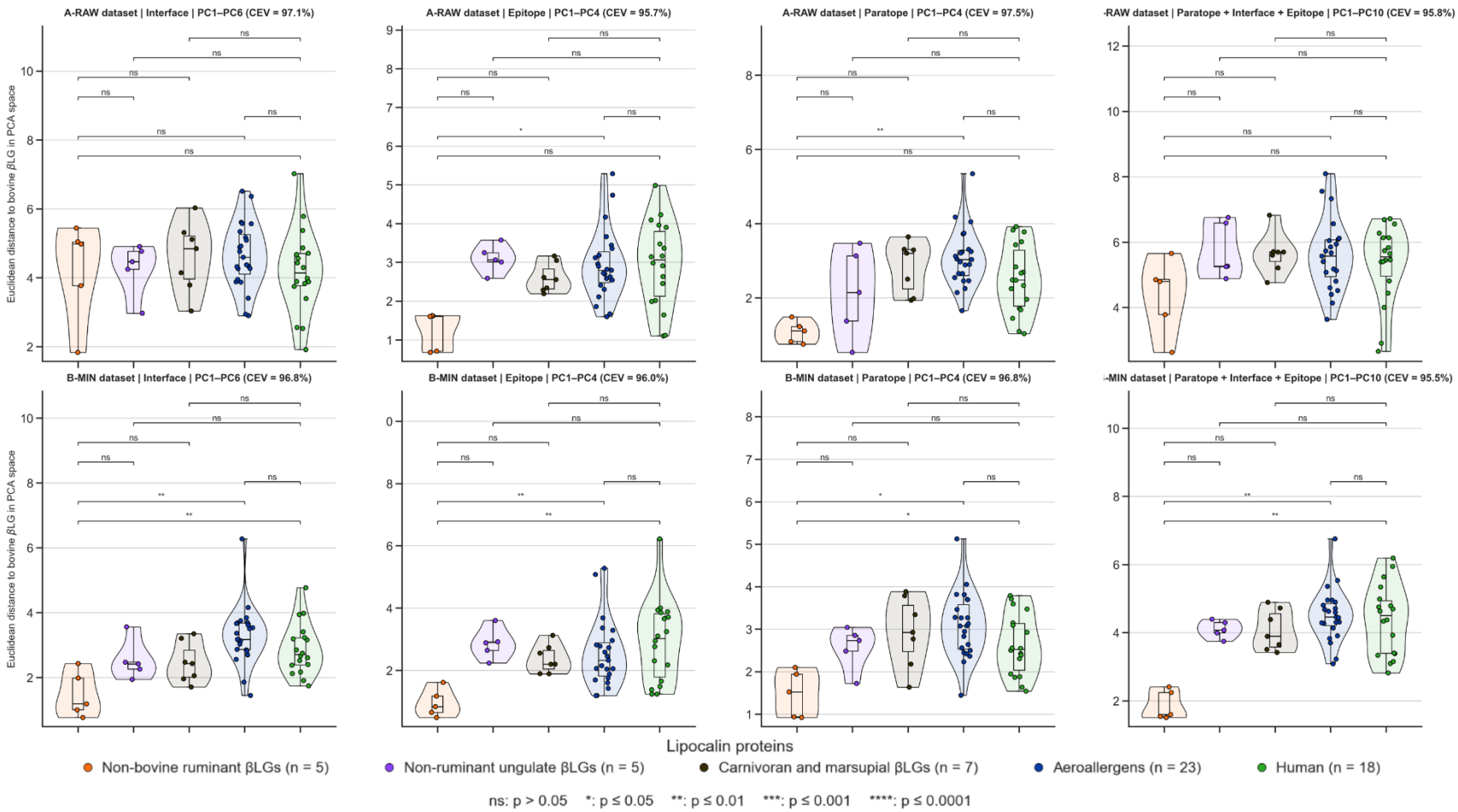

Figure 3-20. Group differences in composite descriptor Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in retained PCA space for comparative interface, comparative epitope, paratope, and composite descriptor sets. Results are shown for RAW and MIN protocols. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after FDR correction.

Table 3-16. Pairwise permutation-test results for the mapped conformational epitope median Euclidean distances.

Protocol	Lipocalin group x	Lipocalin group y	Median x d_{PCA}	Median y d_{PCA}	p	FDR-adjusted p	Results
RAW	Aeroallergens	Non-bovine ruminant βLGs	2.09	0.82	0.0062	0.0311	*
RAW	Aeroallergens	Human	2.09	2.89	0.0945	0.1890	ns
RAW	Non-bovine ruminant βLGs	Human	0.82	2.89	0.0400	0.0999	ns
MIN	Aeroallergens	Non-bovine ruminant βLGs	1.70	1.10	0.0310	0.0774	ns
MIN	Aeroallergens	Human	1.70	2.96	0.0055	0.0500	*
MIN	Non-bovine ruminant βLGs	Human	1.10	2.96	0.0100	0.0500	*

The table reports planned pairwise comparisons of PCA distances from the Bos d 5 reference for the mapped conformational epitope region (MCER) under RAW and MIN protocols. Group median distances, permutation-test p-values, FDR-adjusted p-values. Not significant (ns): p>0.05; *: p≤0.05, **: p ≤0.01, ***: p ≤0.001, ****: p ≤0.0001.

Table 3-17. Pairwise permutation-test results for comparative Fab-allergen model descriptor distances.

Protocol	Lipocalin group x	Lipocalin group y	Median x d_{PCA}	Median y d_{PCA}	p	FDR-adjusted p	Results
			Interface				
RAW	Aeroallergens	Non-bovine ruminant βLGs	4.60	4.98	0.6841	0.9773	ns
RAW	Aeroallergens	Human	4.60	4.14	0.1589	0.8219	ns
RAW	Non-bovine ruminant βLGs	Human	4.98	4.14	0.2351	0.8219	ns
MIN	Aeroallergens	Non-bovine ruminant βLGs	3.18	1.19	0.0016	0.0079	**
MIN	Aeroallergens	Human	3.18	2.71	0.0509	0.0849	ns
MIN	Non-bovine ruminant βLGs	Human	1.19	2.71	0.0027	0.0089	**
			Epitope				
RAW	Aeroallergens	Non-bovine ruminant βLGs	2.80	1.61	0.0056	0.0281	*
RAW	Aeroallergens	Human	2.80	3.06	0.3667	0.4583	ns
RAW	Non-bovine ruminant βLGs	Human	1.61	3.06	0.0518	0.1035	ns
MIN	Aeroallergens	Non-bovine ruminant βLGs	2.33	0.84	0.0016	0.0079	**
MIN	Aeroallergens	Human	2.33	3.03	0.1533	0.2555	ns
MIN	Non-bovine ruminant βLGs	Human	0.84	3.03	0.0027	0.0089	**
			Paratope				
RAW	Aeroallergens	Non-bovine ruminant βLGs	3.03	1.11	0.0016	0.0079	**
RAW	Aeroallergens	Human	3.03	2.48	0.0519	0.0897	ns
RAW	Non-bovine ruminant βLGs	Human	1.11	2.48	0.0196	0.0631	ns
MIN	Aeroallergens	Non-bovine ruminant βLGs	3.08	1.52	0.0085	0.0285	*
MIN	Aeroallergens	Human	3.08	2.53	0.1516	0.2526	ns
MIN	Non-bovine ruminant βLGs	Human	1.52	2.53	0.0027	0.0267	*
			Composite				
RAW	Aeroallergens	Non-bovine ruminant βLGs	5.57	4.80	0.2358	0.5152	ns
RAW	Aeroallergens	Human	5.57	5.56	1.0000	1.0000	ns
RAW	Non-bovine ruminant βLGs	Human	4.80	5.56	0.2576	0.5152	ns
MIN	Aeroallergens	Non-bovine ruminant βLGs	4.46	1.60	0.0016	0.0079	**
MIN	Aeroallergens	Human	4.46	4.51	0.9929	0.9929	ns
MIN	Non-bovine ruminant βLGs	Human	1.60	4.51	0.0027	0.0089	**

The table reports planned pairwise comparisons of PCA distances from the Bos d 5 reference for the comparative interface, epitope, paratope, and composite descriptor sets under the RAW and MIN protocols. Group median distances, permutation-test p-values, FDR-adjusted p-values. Not significant (ns): p>0.05; *: p≤0.05, **: p ≤0.01, ***: p ≤0.001, ****: p ≤0.0001.

4 DISCUSSION

This study introduced and applied a template-based structural workflow to examine whether lipocalin homologs preserve IgE-binding surface features observed in Bos d 5 IgE binding regions. The main finding was that the workflow reproduced the interface descriptors defined by the reference template in the validation controls and, when applied to Bos d 5, consistently identified close β-lactoglobulin homologs as the most similar proteins. Several respiratory lipocalin allergens also showed local similarity to Bos d 5 in specific descriptor sets, but this pattern was less consistent. Human lipocalins sometimes overlapped with allergens in descriptor space, showing that structural similarity alone is not sufficient to infer IgE binding or clinical cross-reactivity.

The method validation supported the internal reproducibility of the IgE Fab–allergen complex construction and descriptor extraction workflow. This workflow follows the general logic of template-based protein–protein complex modelling, in which a known complex structure provides a structural framework for constructing related complexes[187,230–233], and of protein–protein interface analysis, where geometric and interaction descriptors are commonly extracted from complex structures.[156–160] In the identity control tests, each template allergen was reprocessed as its own query. The reconstructed complexes reproduced the expected interface descriptor values under both RAW and MIN protocols, showing that alignment, template replacement, contact definition, and descriptor extraction did not introduce artificial descriptor differences when no allergen substitution was made.

The validation also showed that the workflow produced descriptor outputs for most comparisons. The RAW protocol produced descriptor outputs for all template–query comparisons, while MIN produced a small number of unavailable descriptor records because some structures did not converge during energy minimization. This supports the practical use of the workflow at the scale tested here. Energy minimization was used to relax local geometry optimize intermolecular contacts, reduce steric clashes, and examine protein complex stability. However, although coarse geometric descriptors were stable between RAW and MIN protocols, geometry-sensitive interaction counts shifted more strongly. This aligns with established practice in docking and refinement, where near-native ranking and interface scoring improves when side-chain conformations are optimized and interface clashes are relieved.[234] A practical implication is that RAW and MIN represent distinct analytical states and should not be regarded as interchangeable. RAW more directly reflects the transferred coordinates and is less confounded by energy-function biases introduced by minimization. MIN better reflects whether the transferred interface can support locally consistent polar geometry and packing after relaxation, but it is more expensive and can fail when starting geometries contain severe clashes or topology issues.

A further validation result was that structural alignment quality depended on fold compatibility. Lipocalin template complexes showed stronger alignment support across the lipocalin query set, whereas non-lipocalin templates showed weaker support and shorter aligned regions. This defines an important boundary of the method: template-based replacement is most interpretable when the template allergen and query allergen share a conserved fold. This is consistent with the logic of template-based structural

comparison, where the biological meaning of the model depends on whether the query can reasonably occupy the same structural frame as the template.[187,230–233]

The Bos d 5 application tested whether lipocalins from different biological groups preserve structural and physicochemical features of the Bos d 5 IgE-binding surface. The comparison included close non-bovine ruminant β-lactoglobulins, more distant β-lactoglobulin homologs, respiratory lipocalin allergens, and human lipocalins. This design allowed the analysis to compare expected positive controls, candidate respiratory allergens, and tolerated human comparators within the same lipocalin fold.

The closest and most consistent lipocalin queries came from non-bovine ruminant β-lactoglobulins. These proteins were closest to Bos d 5 across mapped linear epitope regions, the contact-defined conformational epitope, and the generated IgE Fab–query allergen complexes. This result is biologically expected because goat, sheep, buffalo, and related β-lactoglobulins are highly conserved homologs of bovine β-lactoglobulin. Experimental studies have shown extensive immunological cross-reactivity among ruminant β-lactoglobulins, with IgE from cow's milk-allergic patients recognizing homologous proteins from goat, sheep, and buffalo milk, and inhibition assays demonstrating substantial sharing of IgE-binding determinants.[40,42,47,111,118,210,212,235–238] The molecular basis for this cross-reactivity is the exceptional conservation of both sequence and structure within the ruminant β-lactoglobulin family. Reported sequence identities commonly exceed 94%, and crystallographic studies show that caprine and ovine β-lactoglobulins superimpose closely on Bos d 5, with RMSD values as low as 0.35–0.7 Å, indicating near-identical positioning of surface residues.[106,112,113] Conserved linear and conformational epitopes have also been identified across ruminant β-

lactoglobulins.[106] Given this established structural and immunological conservation, recovery of these proteins as the closest matches across all query types is consistent with known biology. The result therefore provides an internal positive control for the workflow, demonstrating that proteins with the greatest experimentally supported preservation of Bos d 5-relevant surface features are correctly identified as the nearest structural neighbors.

More distant β-lactoglobulin homologs showed intermediate behavior. Although these proteins retained the conserved lipocalin β-barrel fold, they were less consistently similar to Bos d 5 than the non-bovine ruminant β-lactoglobulins. This trend is consistent with experimental evidence showing that distant homologs such as porcine and equine β-lactoglobulins preserve the overall β-lactoglobulin scaffold despite substantially lower sequence identity.[112,118,239] Studies of β-lactoglobulin cross-reactivity indicate that while the structural core remains highly conserved, divergence accumulates primarily in surface-exposed residues and loop regions that contribute to antibody recognition.[66,76,155] Consequently, preservation of the global fold does not necessarily imply preservation of Bos d 5-relevant epitopes. The observed pattern therefore supports a graded interpretation in which structural features observed for Bos d 5 are most stable among close homologs and become progressively less conserved as evolutionary distance increases.

Respiratory lipocalin allergens showed a more variable pattern. Several aeroallergens appeared repeatedly among the closest structures in selected descriptor sets, especially Fel d 4, Can f 4, Mus m 1, Fel d 7, and Cav p 1. These recurrent hits suggest that some inhalant lipocalins can present local surface features that resemble parts of the Bos d 5

IgE-binding region. This is plausible because mammalian lipocalins share a conserved β-barrel fold even when their sequence identity is low.[81,82,84,101] However, the aeroallergen ranking was region- and descriptor-dependent. No aeroallergen was consistently closest to Bos d 5 across all analyses, and group-level tests did not show that aeroallergens were consistently distinct from human lipocalins. Therefore, these proteins should be considered structural candidates for further experimental testing. This pattern is consistent with previous reports in the literature. IgE cross-reactivity among lipocalins has been reported, but it is often incomplete, asymmetric, and dependent on the specific allergen pair and patient serum.[66,129,131,152] The present results add structural candidates to this problem. The aeroallergen findings are still useful because they narrow the set of proteins that could be tested experimentally. Rather than testing all respiratory lipocalins equally, follow-up studies could prioritize the recurrent candidates identified here. For example, inhibition ELISA, basophil activation testing, or monoclonal IgE binding assays could test whether these aeroallergens inhibit IgE binding to Bos d 5 or whether Bos d 5 inhibits IgE binding to these respiratory lipocalins. These experiments would determine whether the observed structural compatibility has immunological relevance.

Human lipocalins were included as tolerated comparators because they are endogenous self-proteins that are normally immunologically tolerated, whereas loss of self-tolerance can contribute to autoimmune pathology. Some human lipocalins appeared near Bos d 5 in selected descriptor spaces, especially in RAW interface, RAW combined, and some epitope or paratope rankings. This result is important because it shows that the workflow can detect local structural similarity even in proteins that are not expected to be clinically allergenic. This overlap clarifies the meaning of the results. The lipocalin scaffold can generate similar local surface patches in allergenic and non-allergenic proteins.

Therefore, preserving a structural surface similar to Bos d 5 is not, by itself, enough to infer allergenicity. IgE binding also depends on immune exposure, sensitization history, antibody repertoire, epitope accessibility, affinity, and the ability to cross-link FcεRI-bound IgE on effector cells. [3,73,87,101] This finding supports a conservative interpretation of the aeroallergen hits and explains why structural comparison should be followed by experimental testing.

The results partially support the initial hypothesis. The study hypothesized that Bos d 5-like IgE-binding surface features could be detected across lipocalins and might help identify candidate links between cow's milk allergy and asthma. The strongest support came from close β-lactoglobulin homologs, where the expected structural conservation was recovered. The respiratory lipocalin results provided partial support by identifying repeated candidate aeroallergens with local features observed in Bos d 5. However, the hypothesis was not fully proven because aeroallergens were not consistently separated from human lipocalins, and no IgE-binding assays were performed. Therefore, the results suggest that some respiratory lipocalins contain local structural features that resemble parts of the Bos d 5 IgE-binding surface. These proteins should be considered candidates for targeted experimental studies to determine whether the observed structural similarities have immunological significance.

This work has several limitations. The template-based complex generation used one IgE Fab binding orientation. Patient IgE responses are polyclonal, and different individuals may recognize different Bos d 5 epitopes.[45,66,103,104] Therefore, the results describe compatibility with one known IgE binding mode, and not the full Bos d 5 IgE repertoire.

Moreover, the analysis depends on structural model quality, alignment quality, descriptor definitions, and filtering thresholds. The validation showed that fold compatibility is important, and the application analysis showed that some mapped regions were more informative than others. Regions with poor coverage or limited retained observations are less suitable for group-level interpretation.

The workflow defines the epitope, paratope and the interface based on a threshold. Such thresholds are common in interface analysis and docking evaluation.[240] However, any hard cutoff creates boundary effects, which were observed as small changes in interface after minimization.

Finally, Energy minimization provides more physically coherent geometry for interactions but introduces a major runtime cost and occasional failures due to severe model steric clashes. This mirrors broader experience in docking and refinement: expensive refinement can improve interaction but is not always robust for all starting models.[241]

5 CONCLUSION

This study presents a reproducible structural framework for evaluating IgE-relevant surface similarity across homologous lipocalins using epitope mapping and comparative Fab–allergen modelling.

Application to Bos d 5 demonstrated that non-bovine ruminant β-lactoglobulins consistently preserve Bos d 5-like structural and physicochemical descriptors across mapped epitopes, contact-defined patches, and antibody–antigen interfaces, providing an internal positive control consistent with their high evolutionary conservation.

In contrast, respiratory and more distant lipocalins showed intermittent, region-dependent similarity signals, with several aeroallergens recurrently identified among nearer structural neighbors but without consistent recovery across descriptor spaces. Human lipocalins also overlapped in selected analyses, indicating that descriptors of the shared lipocalin fold alone are insufficient to infer allergenic relevance or IgE-mediated cross-reactivity.

The RAW and MIN preprocessing protocols provide complementary views of structural similarity, with RAW preserving template-constrained geometry for scalable screening and MIN providing relaxed interfaces for refined structural interpretation but at a high computational cost.

Overall, the workflow enables prioritization of candidate proteins and interface regions for experimental validation, while highlighting the necessity of immunological testing to establish biological relevance.

6 FUTURE WORK

Future work should focus on experimental validation of the recurrent candidate lipocalins and further testing of the comparative Fab–query allergen modelling framework. The most direct next step is to test whether the structural similarities identified here correspond to measurable IgE recognition. This could be done using IgE-binding assays with Bos d 5-sensitized sera and, where available, Bos d 5-specific monoclonal IgE. Competitive inhibition assays would be particularly useful for testing whether selected homologous lipocalins can inhibit IgE binding to Bos d 5, or whether Bos d 5 can inhibit IgE binding to those proteins. Functional assays, such as FcεRI-mediated basophil or reporter-cell activation assays, would help determine whether any observed binding is sufficient to produce effector-cell activation.

A second direction is to apply the workflow to additional IgE Fab–allergen template complexes. This would test whether the similarity patterns observed in this study are specific to the Bos d 5 IgE Fab binding orientation or are also observed under other template-defined paratope constraints. Including additional IgE complex templates would also help assess how strongly the workflow depends on epitope topology, antibody orientation, and allergen conformational state.

A third direction is to test the workflow beyond the lipocalin family to evaluate whether RAW and MIN protocol behavior remains consistent in other systems. Lipocalins are well suited for this study because they share a conserved fold despite variation in sequence and surface chemistry. Applying the same approach to other allergen families would help

determine whether the workflow is transferable to protein groups with different folds, different epitope architectures, and different levels of structural conservation.

Finally, further computational refinement could focus on descriptor robustness and protocol sensitivity. This includes testing the effect of structural alignment quality, contact-distance cutoffs, patch definitions, changes in pH, and differences between RAW and MIN outputs. Such analyses would help clarify which descriptor patterns are stable across protocols and which are sensitive to local model preparation. In summary, these future studies would determine whether the framework can improve prioritization of candidate IgE cross-reactivity tests beyond structural similarity alone.

7 REFERENCES

(1) Virtanen, T.; Kinnunen, T.; Rytkönen-Nissinen, M. *Clinical and Experimental Allergy* **2012**, *42* (4), 494.

(2) Khatri, K.; Ball, A.; Glesner, J.; Linn, C.; Vailes, L. D.; Wünschmann, S.; Gabel, S. A.; Zhang, J.; Peebles, R. S.; Borowski, T.; Mueller, G. A.; Chapman, M. D.; Smith, S. A.; Pomés, A.; Chruszcz, M. *Protein Science* **2025**, *34* (9), e70269.

(3) Seidler, C. A.; Zeindl, R.; Fernández-Quintero, M. L.; Tollinger, M.; Liedl, K. R. *Allergies* **2024**, *4* (1), 1.

(4) Cunico, D.; Giannì, G.; Scavone, S.; Buono, E. V.; Caffarelli, C. *Children* **2024**, *11* (11).

(5) Knol, E. F.; van Neerven, R. J. J. *Immunol. Rev.* **2024**, *328* (1), 314.

(6) Virtanen, T. *Frontiers in Allergy* **2021**, *2*, 824736.

(7) Aalberse, R. C.; Crameri, R. *Allergy: European Journal of Allergy and Clinical Immunology* **2011**, *66* (10), 1261.

(8) Iglesia, E. G. A.; Kwan, M.; Virkud, Y. V.; Iweala, O. I. *JAMA* **2024**, *331* (6), 510.

(9) Grandguillaume, I.; Barroso da Silva, F. L.; Etchebest, C. *J. Chem. Inf. Model.* **2025**, *65* (20), 11454.

(10) Guo, M.; Wang, C. In *Whey Protein Production, Chemistry, Functionality, and Applications*; 2019; pp 39–65.

(11) Charcosset, A.; Adel-Patient, K.; Dupont, C.; Bernard, H. *J. Agric. Food Chem.* **2016**, *64* (17), 3394.

(12) Khatri, K.; Richardson, C. M.; Glesner, J.; Kapingidza, A. B.; Mueller, G. A.; Zhang, J.; Dolamore, C.; Vailes, L. D.; Wünschmann, S.; Stokes Peebles, R.; Chapman, M. D.; Smith, S. A.; Chruszcz, M.; Pomés, A. *PNAS Nexus* **2022**, *1* (3), 1.

(13) Pomés, A.; Mueller, G. A.; Chruszcz, M. *Front. Immunol.* **2020**, *11*.

(14) Dhaliwal, B.; Yuan, D.; Pang, M. O. Y.; Henry, A. J.; Cain, K.; Oxbrow, A.; Fabiane, S. M.; Beavil, A. J.; McDonnell, J. M.; Gould, H. J.; Sutton, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (31), 12686.

(15) Sharma, E.; Vitte, J. *Journal of Allergy and Clinical Immunology: Global* **2024**, *3* (2), 100230.

(16) Padavattan, S.; Flicker, S.; Schirmer, T.; Madritsch, C.; Randow, S.; Reese, G.; Vieths, S.; Lupinek, C.; Ebner, C.; Valenta, R.; Markovic-Housley, Z. *The Journal of Immunology* **2009**, *182* (4), 2141.

(17) Niemi, M.; Jylhä, S.; Laukkanen, M. L.; Söderlund, H.; Mäkinen-Kiljunen, S.; Kallio, J. M.; Hakulinen, N.; Haahtela, T.; Takkinen, K.; Rouvinen, J. *Structure* **2007**, *15* (11), 1413.

(18) Pomés, A.; Smith, S. A.; Chruszcz, M.; Mueller, G. A.; Brackett, N. F.; Chapman, M. D. *Journal of Allergy and Clinical Immunology* **2024**, *153* (3), 560.

(19) Schroeder, A.; Kumar, R.; Pongracic, J. A.; Sullivan, C. L.; Caruso, D. M.; Costello, J.; Meyer, K. E.; Vucic, Y.; Gupta, R.; Kim, J. S.; Fuleihan, R.; Wang, X. *Clinical and Experimental Allergy* **2009**, *39* (2), 261.
(20) Singer, A. G.; Kosowan, L.; Soller, L.; Chan, E. S.; Nankissoor, N. N.; Phung, R. R.; Abrams, E. M. *Journal of Allergy and Clinical Immunology: In Practice* **2021**, *9* (1), 193.
(21) Vogel, N. M.; Katz, H. T.; Lopez, R.; Lang, D. M. *Journal of Asthma* **2008**, *45* (10), 862.
(22) Simpson, A. B.; Glutting, J.; Yousef, E. *Pediatr. Pulmonol.* **2007**, *42* (6), 489.
(23) Malmberg, L. P.; Saarinen, K. M.; Pelkonen, A. S.; Savilahti, E.; Mäkelä, M. J. *Clinical and Experimental Allergy* **2010**, *40* (10), 1491.
(24) Esteban-Gorgojo, I.; Gorgojo, M. P.; Sastrec, J.; Quirce, S. *Allergol. Immunopathol. (Madr).* **2021**, *49* (1), 68.
(25) Liu, A. H.; Jaramillo, R.; Sicherer, S. H.; Wood, R. A.; Bock, S. A.; Burks, A. W.; Massing, M.; Cohn, R. D.; Zeldin, D. C. *Journal of Allergy and Clinical Immunology* **2010**, *126* (4), 798.
(26) Vermeulen, E. M.; Koplin, J. J.; Dharmage, S. C.; Gurrin, L. C.; Peters, R. L.; McWilliam, V.; Ponsonby, A. L.; Dwyer, T.; Lowe, A. J.; Tang, M. L. K.; Allen, K. J.; Wake, M.; Matheson, M.; Robertson, C.; Hill, D. *Journal of Allergy and Clinical Immunology: In Practice* **2018**, *6* (4), 1336.
(27) Schroeder, A.; Kumar, R.; Pongracic, J. A.; Sullivan, C. L.; Caruso, D. M.; Costello, J.; Meyer, K. E.; Vucic, Y.; Gupta, R.; Kim, J. S.; Fuleihan, R.; Wang, X. *Clinical and Experimental Allergy* **2009**, *39* (2), 261.
(28) Suleyman, A.; Tamay, Z.; Güler, N. *J. Trop. Pediatr.* **2022**, *68* (2).
(29) Wang, J.; Liu, A. H. *Curr. Opin. Allergy Clin. Immunol.* **2011**, *11* (3), 249.
(30) Hansen, M. M.; Nissen, S. P.; Halken, S.; Høst, A. *Pediatric Allergy and Immunology* **2021**, *32* (4), 727.
(31) Roberts, G.; Patel, N.; Levi-Schaffer, F.; Habibi, P.; Lack, G. *Journal of Allergy and Clinical Immunology* **2003**, *112* (1), 168.
(32) Sherenian, M. G.; Singh, A. M.; Arguelles, L.; Balmert, L.; Caruso, D.; Wang, X.; Pongracic, J.; Kumar, R. *Annals of Allergy, Asthma and Immunology* **2018**, *121* (5), 588.
(33) Roberts, G.; Patel, N.; Levi-Schaffer, F.; Habibi, P.; Lack, G. *Journal of Allergy and Clinical Immunology* **2003**, *112* (1), 168.
(34) Turner, P. J.; Arasi, S.; Ballmer-Weber, B.; Baseggio Conrado, A.; Deschildre, A.; Gerdts, J.; Halken, S.; Muraro, A.; Patel, N.; Van Ree, R.; de Silva, D.; Worm, M.; Zuberbier, T.; Roberts, G. *Allergy: European Journal of Allergy and Clinical Immunology* **2022**, *77* (9), 2634.
(35) Järvinen, K. M.; Beyer, K.; Vila, L.; Chatchatee, P.; Busse, P. J.; Sampson, H. A. *Journal of Allergy and Clinical Immunology* **2002**, *110* (2), 293.

(36) Arasi, S.; Cafarotti, A.; Fiocchi, A. *Curr. Opin. Allergy Clin. Immunol.* **2022**, *22* (3), 181.
(37) Błażowski, Ł.; Podlecka, D.; Brzozowska, A.; Jerzyńska, J.; Seweryn, M.; Błażowski, M.; Majak, P. *Nutrients* **2025**, *17* (11).
(38) Hochwallner, H.; Schulmeister, U.; Swoboda, I.; Spitzauer, S.; Valenta, R. *Methods* **2014**, *66* (1), 22.
(39) Verelst, S.; Sinnesael, R.; Taïbi, F.; Tuyls, S.; Coorevits, L.; Breynaert, C.; Bullens, D.; Schrijvers, R. *Nutrients* **2025**, *17* (15).
(40) Järvinen, K. M.; Chatchatee, P. *Curr. Opin. Allergy Clin. Immunol.* **2009**, *9* (3), 251.
(41) Ma, Y.; Zhang, L.; Zhou, P. *J. Agric. Food Chem.* **2024**, *72* (43), 23969.
(42) El-Agamy, E. I. *Small Ruminant Research* **2007**, *68* (1–2), 64.
(43) Villa, C.; Costa, J.; Oliveira, M. B. P. P.; Mafra, I. *Compr. Rev. Food Sci. Food Saf.* **2018**, *17* (1), 137.
(44) Cao, Z.; Chang, Z.; Gao, Y.; Xue, B.; Yang, Y.; Zhao, L.; Ge, W. *LWT* **2025**, *219*.
(45) Sélo, I.; Clément, G.; Bernard, H.; Chatel, J. M.; Créminon, C.; Peltre, G.; Wal, J. M. *Clinical and Experimental Allergy* **1999**, *29* (8), 1055.
(46) Maryniak, N. Z.; Hansen, E. B.; Ballegaard, A. S. R.; Sancho, A. I.; Bøgh, K. L. *Nutrients* **2018**, *10* (12), 1903.
(47) Restani, P.; Gaiaschi, A.; Plebani, A.; Beretta, B.; Cavagni, G.; Fiocchi, A.; Poiesi, C.; Velonà, T.; Ugazio, A. G.; Galli, C. L. *Clinical and Experimental Allergy* **1999**, *29* (7), 997.
(48) Teodorowicz, M.; Zenker, H. E.; Ewaz, A.; Tsallis, T.; Mauser, A.; Gensberger-Reigl, S.; de Jong, N. W.; Hettinga, K. A.; Wichers, H. J.; van Neerven, R. J. J.; Savelkoul, H. F. J. *Mol. Nutr. Food Res.* **2021**, *65* (8), 2000834.
(49) Huang, H.; Luo, W.; Wei, N.; Liang, X.; Zheng, P.; Hu, H.; Sun, B. *BMC Pediatr.* **2020**, *20* (1), 1.
(50) Monaci, L.; Tregoat, V.; Van Hengel, A. J.; Anklam, E. *European Food Research and Technology* **2006**, *223* (2), 149.
(51) Lo, R.; Groetch, M.; Brooks, J.; Anderson, E.; Rodríguez del Río, P.; Anagnostou, A. *Journal of Allergy and Clinical Immunology: In Practice* **2025**, *13* (4), 754.
(52) Bognanni, A.; Firmino, R. T.; Arasi, S.; Chu, D. K.; Chu, A. W. L.; Waffenschmidt, S.; Agarwal, A.; Dziechciarz, P.; Horvath, A.; Mihara, H.; Roldan, Y.; Terracciano, L.; Martelli, A.; Starok, A.; Said, M.; Shamir, R.; Ansotegui, I. J.; Dahdah, L.; Ebisawa, M.; Galli, E.; Kamenwa, R.; Lack, G.; Li, H.; Pawankar, R.; Warner, A.; Wong, G. W. K.; Bozzola, M.; Assa'Ad, A.; Dupont, C.; Bahna, S.; Spergel, J.; Venter, C.; Szajewska, H.; Nowak-Wegrzyn, A. H.; Vandenplas, Y.; Papadopoulos, N. G.; Waserman, S.; Fiocchi, A.; Schünemann, H. J.; Brożek, J. L. *World Allergy Organization Journal* **2024**, *17* (9), 100947.

(53) Delgado Dolset, M. I.; Obeso, D.; Rodriguez-Coira, J.; Villaseñor, A.; González Cuervo, H.; Arjona, A.; Barbas, C.; Barber, D.; Carrillo, T.; Escribese, M. M. *Front. Med. (Lausanne).* **2022**, *9*, 1009324.
(54) Krogulska, A.; Dynowski, J.; Jędrzejczyk, M.; Sardecka, I.; Małachowska, B.; Wąsowska-Królikowska, K. *Pediatr. Pulmonol.* **2016**, *51* (8), 787.
(55) Cheng, F.; He, L.; Deng, D.; Zhang, J.; Liu, C. *Journal of Asthma* **2025**, *62* (1), 45.
(56) Dezfouli, S. G.; Mothes-Luksch, N.; Jensen, A. N.; Untersmayr, E.; Kundi, M.; Jensen-Jarolim, E. *World Allergy Organization Journal* **2020**, *13* (12), 100483.
(57) Abbring, S.; Verheijden, K. A. T.; Diks, M. A. P.; Leusink-Muis, A.; Hols, G.; Baars, T.; Garssen, J.; van Esch, B. C. A. M. *Front. Immunol.* **2017**, *8* (AUG), 291826.
(58) Sierra-Heredia, C.; North, M.; Brook, J.; Daly, C.; Ellis, A. K.; Henderson, D.; Henderson, S. B.; Lavigne, É.; Takaro, T. K. *Int. J. Environ. Res. Public Health* **2018**, *15* (8), 1577.
(59) Murrison, L. B.; Brandt, E. B.; Myers, J. B.; Khurana Hershey, G. K. *Journal of Clinical Investigation* **2019**, *129* (4), 1504.
(60) Pang, S. L.; Ho, K. L.; Waterman, J.; Rambo, R. P.; Teh, A. H.; Mathavan, I.; Harris, G.; Beis, K.; Say, Y. H.; Anusha, M. S.; Sio, Y. Y.; Chew, F. T.; Ng, C. L. *Sci. Rep.* **2019**, *9* (1), 4933.
(61) Gromek, W.; Kołdej, N.; Kurowski, M.; Majsiak, E. *Clin. Rev. Allergy Immunol.* **2025**, *68* (1), 75.
(62) Saarinen, K. M.; Pelkonen, A. S.; Mäkelä, M. J.; Savilahti, E. *Journal of Allergy and Clinical Immunology* **2005**, *116* (4), 869.
(63) Emons, J. A. M.; Gerth van Wijk, R. *Curr. Treat. Options Allergy* **2018**, *5* (4), 436.
(64) Busse, W. W. *Annals of Allergy, Asthma and Immunology* **2018**, *120* (2), 116.
(65) Bublin, M.; Breiteneder, H. *Allergy: European Journal of Allergy and Clinical Immunology* **2020**, *75* (5), 1019.
(66) Aalberse, R. C.; Akkerdaas, J.; Van Ree, R. *Allergy: European Journal of Allergy and Clinical Immunology* **2001**, *56* (6), 478.
(67) Kringelum, J. V.; Nielsen, M.; Padkjær, S. B.; Lund, O. *Mol. Immunol.* **2013**, *53* (1–2), 24.
(68) Hu, W.; Xie, F.; Wu, Y.; Meng, X.; Yang, A.; Wu, Z.; Gao, J.; Li, X.; Chen, H. *J. Agric. Food Chem.* **2025**, *73* (9), 5537.
(69) Dall'Antonia, F.; Pavkov-Keller, T.; Zangger, K.; Keller, W. *Methods* **2014**, *66* (1), 3.
(70) Dall'Antonia, F.; Gieras, A.; Devanaboyina, S. C.; Valenta, R.; Keller, W. *Journal of Allergy and Clinical Immunology* **2011**, *128* (4), 872.
(71) Sela-Culang, I.; Kunik, V.; Ofran, Y. *Front. Immunol.* **2013**, *4* (OCT), 64858.

(72) Akbar, R.; Robert, P. A.; Pavlović, M.; Jeliazkov, J. R.; Snapkov, I.; Slabodkin, A.; Weber, C. R.; Scheffer, L.; Miho, E.; Haff, I. H.; Haug, D. T. T.; Lund-Johansen, F.; Safonova, Y.; Sandve, G. K.; Greiff, V. *Cell Rep.* **2021**, *34* (11), 108856.
(73) Ball, A.; Khatri, K.; Glesner, J.; Vailes, L. D.; Wünschmann, S.; Gabel, S. A.; Mueller, G. A.; Zhang, J.; Peebles, R. S.; Chapman, M. D.; Smith, S. A.; Chruszcz, M.; Pomés, A. *Journal of Allergy and Clinical Immunology* **2024**, *154* (2), 447.
(74) Ponomarenko, J. V.; Bourne, P. E. *BMC Struct. Biol.* **2007**, *7* (1), 64.
(75) Peng, H. P.; Lee, K. H.; Jian, J. W.; Yang, A. S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (26), E2656.
(76) Rubinstein, N. D.; Mayrose, I.; Halperin, D.; Yekutieli, D.; Gershoni, J. M.; Pupko, T. *Mol. Immunol.* **2008**, *45* (12), 3477.
(77) Dalkas, G. A.; Teheux, F.; Kwasigroch, J. M.; Rooman, M. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82* (9), 1734.
(78) McCoy, A. J.; Chandana Epa, V.; Colman, P. M. *J. Mol. Biol.* **1997**, *268* (2), 570.
(79) Sinha, N.; Mohan, S.; Lipschultz, C. A.; Smith-Gill, S. J. *Biophys. J.* **2002**, *83* (6), 2946.
(80) Chandrasekaran, P.; Weiskirchen, S.; Weiskirchen, R. *Int. J. Mol. Sci.* **2024**, *25* (8), 4290.
(81) Yang, Y.; He, X.; Li, F.; He, S.; Liu, M.; Li, M.; Xia, F.; Su, W.; Liu, G. *Compr. Rev. Food Sci. Food Saf.* **2024**, *23* (3), e13340.
(82) Hilger, C.; Kuehn, A.; Hentges, F. *Curr. Allergy Asthma Rep.* **2012**, *12* (5), 438.
(83) Rouvinen, J.; Rautiainen, J.; Virtanen, T.; Zeiler, T.; Kauppinen, J.; Taivainen, A.; Mäntyjärvi, R. *Journal of Biological Chemistry* **1999**, *274* (4), 2337.
(84) Clayton, G. M.; White, J.; Lee, S.; Kappler, J. W.; Chan, S. K. *PLoS One* **2019**, *14* (9), e0213052.
(85) Sawyer, L. *Front. Physiol.* **2021**, *12*, 678080.
(86) Ganfornina, M. D.; Åkerström, B.; Sanchez, D. *Front. Physiol.* **2022**, *13*.
(87) Chruszcz, M.; Kapingidza, A. B.; Dolamore, C.; Kowal, K. *PLoS One* **2018**, *13* (11), e0208276.
(88) Nakatsuji, M.; Sugiura, K.; Suda, K.; Sakurai, M.; Ubatani, M.; Muroya, H.; Okubo, R.; Noguchi, R.; Kamata, Y.; Fukutomi, Y.; Ishibashi, O.; Nishimura, S.; Inui, T. *FEBS Journal* **2022**, *289* (6), 1668.
(89) Madhurantakam, C.; Nilsson, O. B.; Uchtenhagen, H.; Konradsen, J.; Saarne, T.; Högbom, E.; Sandalova, T.; Grönlund, H.; Achour, A. *J. Mol. Biol.* **2010**, *401* (1), 68.
(90) Tiller, T.; Meffre, E.; Yurasov, S.; Tsuiji, M.; Nussenzweig, M. C.; Wardemann, H. *J. Immunol. Methods* **2008**, *329* (1–2), 112.
(91) Horst, A.; Hunzelmann, N.; Arce, S.; Herber, M.; Manz, R. A.; Radbruch, A.; Nischt, R.; Schmitz, J.; Assenmacher, M. *Clin. Exp. Immunol.* **2002**, *130* (3), 370.
(92) Cia, G.; Pucci, F.; Rooman, M. *Brief. Bioinform.* **2023**, *24* (1).

(93) Liu, F.; Yuan, C. C.; Chen, H.; Yang, F. *Sci. Rep.* **2024**, *14* (1), 1.
(94) Choi, S.; Kim, D. *Sci. Rep.* **2024**, *14* (1), 27496.
(95) Lo, Y. T.; Shih, T. C.; Pai, T. W.; Ho, L. P.; Wu, J. L.; Chou, H. Y. *BMC Genomics* **2021**, *22* (2), 1.
(96) Marini-Rapoport, O.; Fernández-Quintero, M. L.; Keswani, T.; Zong, G.; Shim, J.; Pedersen, L. C.; Mueller, G. A.; Patil, S. U. *Clin. Exp. Immunol.* **2024**, *216* (1), 25.
(97) Ren, J.; Liu, Q.; Ellis, J.; Li, J. *Bioinformatics* **2014**, *30* (12), i264.
(98) Sweredoski, M. J.; Baldi, P. *Bioinformatics* **2008**, *24* (12), 1459.
(99) Ponomarenko, J.; Bui, H. H.; Li, W.; Fusseder, N.; Bourne, P. E.; Sette, A.; Peters, B. *BMC Bioinformatics* **2008**, *9* (1), 1.
(100) Qiao, X.; Qu, L.; Guo, Y.; Hoshino, T. *Journal of Physical Chemistry B* **2021**, *125* (41), 11374.
(101) Niemi, M. H.; Rytkönen-Nissinen, M.; Miettinen, I.; Jänis, J.; Virtanen, T.; Rouvinen, J. *Sci. Rep.* **2015**, *5* (1), 1.
(102) Thomas, K.; Aalbers, M.; Bannon, G. A.; Bartels, M.; Dearman, R. J.; Esdaile, D. J.; Fu, T. J.; Glatt, C. M.; Hadfield, N.; Hatzos, C.; Hefle, S. L.; Heylings, J. R.; Goodman, R. E.; Henry, B.; Herouet, C.; Holsapple, M.; Ladics, G. S.; Landry, T. D.; MacIntosh, S. C.; Rice, E. A.; Privalle, L. S.; Steiner, H. Y.; Teshima, R.; Van Ree, R.; Woolhiser, M.; Zawodny, J. *Regulatory Toxicology and Pharmacology* **2004**, *39* (2), 87.
(103) Ball, G.; Shelton, M. J.; Walsh, B. J.; Hill, D. J.; Hosking, C. S.; Howden, M. E. H. *Clinical and Experimental Allergy* **1994**, *24* (8), 758.
(104) Cocco, R. R.; Järvinen, K. M.; Han, N.; Beyer, K.; Sampson, H. A. *Clinical and Experimental Allergy* **2007**, *37* (6), 831.
(105) Cerecedo, I.; Zamora, J.; Shreffler, W. G.; Lin, J.; Bardina, L.; Dieguez, M. C.; Wang, J.; Muriel, A.; de la Hoz, B.; Sampson, H. A. *Journal of Allergy and Clinical Immunology* **2008**, *122* (3), 589.
(106) Cong, Y. J.; Li, L. F. *J. Dairy Sci.* **2012**, *95* (11), 6307.
(107) Heinzmann, A.; Blattmann, S.; Spuergin, P.; Forster, J.; Deichmann, K. A. *Int. Arch. Allergy Immunol.* **1999**, *120* (4), 280.
(108) Järvinen, K. M.; Chatchatee, P.; Bardina, L.; Beyer, K.; Sampson, H. A. *Int. Arch. Allergy Immunol.* **2001**, *126* (2), 111.
(109) Wang, X.; Hu, Y.; Tan, H.; Dong, X.; Zhang, S.; Fu, S.; Gao, J.; Chen, H.; Liu, G.; Li, X. *J. Dairy Sci.* **2023**, *106* (11), 7382.
(110) Clement, G.; Boquet, D.; Frobert, Y.; Bernard, H.; Negroni, L.; Chatel, J. M.; Adel-Patient, K.; Creminon, C.; Wal, J. M.; Grassi, J. *J. Immunol. Methods* **2002**, *266* (1–2), 67.
(111) Ah-Leung, S.; Bernard, H.; Bidat, E.; Paty, E.; Rancé, F.; Scheinmann, P.; Wal, J. M. *Allergy: European Journal of Allergy and Clinical Immunology* **2006**, *61* (11), 1358.

(112) Crowther, J. M.; Lassé, M.; Suzuki, H.; Kessans, S. A.; Loo, T. S.; Norris, G. E.; Hodgkinson, A. J.; Jameson, G. B.; Dobson, R. C. J. *FEBS Lett.* **2014**, *588* (21), 3816.
(113) Loch, J. I.; Molenda, M.; Kopeč, M.; Świątek, S.; Lewiński, K. *Biopolymers* **2014**, *101* (8), 886.
(114) Picariello, G.; Ferranti, P.; Fierro, O.; Mamone, G.; Caira, S.; Di Luccia, A.; Monica, S.; Addeo, F. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2010**, *878* (3–4), 295.
(115) Lisson, M.; Novak, N.; Erhardt, G. *J. Dairy Sci.* **2014**, *97* (4), 1939.
(116) Oksanen, E.; Jaakola, V. P.; Tolonen, T.; Valkonen, K.; Virtanen, V.; Åkerström, B.; Kalkkinen, N.; Goldman, A. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62* (11), 1369.
(117) Lindholm Bøgh, K.; Barkholt, V.; Bernhard Madsen, C. *Int. Arch. Allergy Immunol.* **2013**, *161* (1), 21.
(118) Restani, P.; Ballabio, C.; Di Lorenzo, C.; Tripodi, S.; Fiocchi, A. *Analytical and Bioanalytical Chemistry*. September 2009, pp 47–56.
(119) Reis, P. B. P. S.; Barletta, G. P.; Gagliardi, L.; Fortuna, S.; Soler, M. A.; Rocchia, W. *Front. Mol. Biosci.* **2022**, *9*, 945808.
(120) Ferrè, F.; Ausiello, G.; Zanzoni, A.; Helmer-Citterich, M. *Nucleic Acids Res.* **2004**, *32* (suppl_1), D240.
(121) Xue, L. C.; Dobbs, D.; Bonvin, A. M. J. J.; Honavar, V. *FEBS Lett.* **2015**, *589* (23), 3516.
(122) Thornton, J. M.; Edwards, M. S.; Taylor, W. R.; Barlow, D. J. *EMBO J.* **1986**, *5* (2), 409.
(123) Rubinstein, N. D.; Mayrose, I.; Pupko, T. *Mol. Immunol.* **2009**, *46* (5), 840.
(124) Tien, M. Z.; Meyer, A. G.; Sydykova, D. K.; Spielman, S. J.; Wilke, C. O. *PLoS One* **2013**, *8* (11), e80635.
(125) Ramaraj, T.; Angel, T.; Dratz, E. A.; Jesaitis, A. J.; Mumey, B. *Biochim. Biophys. Acta Proteins Proteom.* **2012**, *1824* (3), 520.
(126) Wu, D.; Sun, J.; Xu, T.; Wang, S.; Li, G.; Li, Y.; Cao, Z. *Immunome Res.* **2010**, *6* (SUPPL. 1), S1.
(127) Janssen-Weets, B.; Kerff, F.; Swiontek, K.; Kler, S.; Czolk, R.; Revets, D.; Kuehn, A.; Bindslev-Jensen, C.; Ollert, M.; Hilger, C. *Frontiers in Allergy* **2022**, *3*, 958711.
(128) Roth-Walter, F.; Gomez-Casado, C.; Pacios, L. F.; Mothes-Luksch, N.; Roth, G. A.; Singer, J.; Diaz-Perales, A.; Jensen-Jarolim, E. *Journal of Biological Chemistry* **2014**, *289* (25), 17416.
(129) Todorović, N.; Trifonova, D.; Liu, Z.; Curin, M.; Schooltink, L.; Sagmeister, T.; Grininger, C.; Kiss, R.; Gottstein, N.; Gesslbauer, B.; Winkler, A.; Pavkov-Keller, T.; Karaulov, A.; Valenta, R.; Keller, W. *Allergy* **2025**, *0*, 1.

(130) Pomés, A.; Chruszcz, M.; Gustchina, A.; Wlodawer, A. *Curr. Allergy Asthma Rep.* **2015**, *15* (4), 8.
(131) Saarelainen, S.; Rytkönen-Nissinen, M.; Rouvinen, J.; Taivainen, A.; Auriola, S.; Kauppinen, A.; Kinnunen, T.; Virtanen, T. *Clinical and Experimental Allergy* **2008**, *38* (2), 374.
(132) Nilsson, O. B.; Binnmyr, J.; Zoltowska, A.; Saarne, T.; Van Hage, M.; Grönlund, H. *Allergy* **2012**, *67* (6), 751.
(133) Yamamoto, K.; Ishibashi, O.; Sugiura, K.; Ubatani, M.; Sakaguchi, M.; Nakatsuji, M.; Shimamoto, S.; Noda, M.; Uchiyama, S.; Fukutomi, Y.; Nishimura, S.; Inui, T. *Scientific Reports 2019 9:1* **2019**, *9* (1), 1503.
(134) Jakob, T.; Hilger, C.; Hentges, F. *Allergy* **2013**, *68* (5), 690.
(135) Lu, J.; Zhu, H.; Yang, Q.; Xu, Y.; Huang, Z.; Sun, B. *J. Asthma Allergy* **2025**, *18*, 363.
(136) Apostolovic, D.; Sánchez-Vidaurre, S.; Waden, K.; Curin, M.; Grundström, J.; Gafvelin, G.; Cirkovic Velickovic, T.; Grönlund, H.; Thomas, W. R.; Valenta, R.; Hamsten, C.; van Hage, M. *Allergy* **2016**, *71* (10), 1490.
(137) Mattsson, L.; Lundgren, T.; Olsson, P.; Sundberg, M.; Lidholm, J. *Clinical & Experimental Allergy* **2010**, *40* (8), 1276.
(138) Swiontek, K.; Kler, S.; Lehners, C.; Ollert, M.; Hentges, F.; Hilger, C. *Clinical & Experimental Allergy* **2021**, *51* (6), 829.
(139) Yih, W. T.; Siew, L. C.; Tan, C. O.; Le, Y. Y.; Yuen, S. T.; Fook, T. C.; Sivaraman, J.; Yu, K. M. *Journal of Biological Chemistry* **2009**, *284* (5), 3148.
(140) Jeal, H.; Harris, J.; Draper, A.; Newman Taylor, A.; Cullinan, P.; Jones, M. G. *Allergy* **2009**, *64* (6), 855.
(141) Apostolovic, D.; Stanic-Vucinic, D.; De Jongh, H. H. J.; De Jong, G. A. H.; Mihailovic, J.; Radosavljevic, J.; Radibratovic, M.; Nordlee, J. A.; Baumert, J. L.; Milcic, M.; Taylor, S. L.; Garrido Clua, N.; Cirkovic Velickovic, T.; Koppelman, S. J. *Sci. Rep.* **2016**, *6* (1), 1.
(142) Suutari, T. J.; Valkonen, K. H.; Karttunen, T. J.; Ehn, B. M.; Ekstrand, B.; Bengtsson, U.; Virtanen, V.; Nieminen, M.; Kokkonen, J. *J. Investig. Allergol. Clin. Immunol.* **2006**, *16* (5), 296.
(143) Pomés, A.; Davies, J. M.; Gadermaier, G.; Hilger, C.; Holzhauser, T.; Lidholm, J.; Lopata, A. L.; Mueller, G. A.; Nandy, A.; Radauer, C.; Chan, S. K.; Jappe, U.; Kleine-Tebbe, J.; Thomas, W. R.; Chapman, M. D.; van Hage, M.; van Ree, R.; Vieths, S.; Raulf, M.; Goodman, R. E. *Mol. Immunol.* **2018**, *100*, 3.
(144) Charkoftaki, G.; Wang, Y.; McAndrews, M.; Bruford, E. A.; Thompson, D. C.; Vasiliou, V.; Nebert, D. W. *Hum. Genomics* **2019**, *13* (1), 11.
(145) Yang, H. H.; Wang, X.; Li, S.; Liu, Y.; Akbar, R.; Fan, G. C. *Pharmacol. Ther.* **2023**, *244*, 108385.
(146) Liukko, A. L. K.; Kinnunen, T. T.; Rytkönen-Nissinen, M. A.; Kailaanmäki, A. H. T.; Randell, J. T.; Maillère, B.; Virtanen, T. I. *PLoS One* **2014**, *9* (5), e98461.

(147) Liu, B.; Yang, Z.; Liu, Q.; Zhang, Y.; Ding, H.; Lai, H.; Li, Q. *Front. Genet.* **2023**, *14*, 1294159.
(148) Múnera, M.; Martínez, D.; Labrada, A.; Caraballo, L.; Puerta, L. *Int. J. Mol. Sci.* **2019**, *20* (24), 6107.
(149) Kortekaas Krohn, I.; Badloe, F. M. S.; Herrmann, N.; Maintz, L.; De Vriese, S.; Ring, J.; Schmid-Grendelmeier, P.; Traidl-Hoffmann, C.; Akdis, C.; Lauener, R.; Brüggen, M. C.; Rhyner, C.; Bersuch, E.; Dreher, A.; Hammel, G.; Luschkova, D.; Lang, C.; Reiger, M.; Bieber, T.; Gutermuth, J. *Allergy* **2023**, *78* (12), 3178.
(150) Altrichter, S.; Kriehuber, E.; Moser, J.; Valenta, R.; Kopp, T.; Stingl, G. *Journal of Investigative Dermatology* **2008**, *128* (9), 2232.
(151) Laganaro, M.; Pieger, A. M.; Muñoz-González, A.; Noriega Fernández, E.; Ververis, E.; Germini, A.; Fernández Dumont, A. *Frontiers in Toxicology* **2025**, *7*, 1701391.
(152) Aalberse, R. C. *Clinical and Molecular Allergy* **2007**, *5* (1), 2.
(153) Høie, M. H.; Gade, F. S.; Johansen, J. M.; Würtzen, C.; Winther, O.; Nielsen, M.; Marcatili, P. *Front. Immunol.* **2024**, *15*, 1322712.
(154) Zhou, C.; Chen, Z.; Zhang, L.; Yan, D.; Mao, T.; Tang, K.; Qiu, T.; Cao, Z. *Nucleic Acids Res.* **2019**, *47* (W1), W388.
(155) Negi, S. S.; Braun, W. *Bioinformatics* **2017**, *33* (7), 1014.
(156) Krawczyk, K.; Liu, X.; Baker, T.; Shi, J.; Deane, C. M. *Bioinformatics* **2014**, *30* (16), 2288.
(157) Qiu, T.; Zhang, L.; Chen, Z.; Wang, Y.; Mao, T.; Wang, C.; Cun, Y.; Zheng, G.; Yan, D.; Zhou, M.; Tang, K.; Cao, Z. *Nucleic Acids Res.* **2023**, *51* (W1), W528.
(158) Nguyen, M. N.; Verma, C. S.; Zhong, P. *Nucleic Acids Res.* **2019**, *47* (W1), W482.
(159) Jubb, H. C.; Higueruelo, A. P.; Ochoa-Montaño, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L. *J. Mol. Biol.* **2017**, *429* (3), 365.
(160) Krissinel, E.; Henrick, K. *J. Mol. Biol.* **2007**, *372* (3), 774.
(161) Rosell, M.; Fernández-Recio, J. *Curr. Opin. Struct. Biol.* **2020**, *64*, 59.
(162) Lyskov, S.; Gray, J. J. *Nucleic Acids Res.* **2008**, *36* (suppl_2), W233.
(163) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. *J. Am. Chem. Soc.* **2003**, *125* (7), 1731.
(164) Sircar, A.; Gray, J. J. *PLoS Comput. Biol.* **2010**, *6* (1), e1000644.
(165) Cronin, C.; Muñoz Archidona, C.; Fernández Prudencio, B.; Gallagher, A.; Velasco Zuniga, R.; Trujillo Wurttele, J. *Antibodies 2023, Vol. 12, Page 76* **2023**, *12* (4), 76.
(166) Sorensen, K.; Meyer, R.; Grimshaw, K. E.; Cawood, A. L.; Acosta-Mena, D.; Stratton, R. J. *Immun. Inflamm. Dis.* **2022**, *10* (3), e572.
(167) Tosca, M. A.; Olcese, R.; Marinelli, G.; Schiavetti, I.; Ciprandi, G. *Children 2022, Vol. 9, Page 1872* **2022**, *9* (12), 1872.

(168) Jensen-Jarolim, E.; Pacios, L. F.; Bianchini, R.; Hofstetter, G.; Roth-Walter, F. *Allergy* **2016**, *71* (3), 286.
(169) Virtanen, T.; Zeiler, T.; Rautiainen, J.; Mäntyjärvi, R. *Immunol. Today* **1999**, *20* (9), 398.
(170) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Adesina, A.; Ahmad, S.; Bowler-Barnett, E. H.; Bye-A-Jee, H.; Carpentier, D.; Denny, P.; Fan, J.; Garmiri, P.; da Costa Gonzales, L. J.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Joshi, V.; Jyothi, D.; Kandasaamy, S.; Lock, A.; Luciani, A.; Luo, J.; Lussi, Y.; Marin, J. S. M.; Raposo, P.; Rice, D. L.; Santos, R.; Speretta, E.; Stephenson, J.; Totoo, P.; Tyagi, N.; Urakova, N.; Vasudev, P.; Warner, K.; Wijerathne, S.; Yu, C. W. H.; Zaru, R.; Bridge, A. J.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A. H.; Axelsen, K. B.; Bansal, P.; Baratin, D.; Batista Neto, T. M.; Blatter, M. C.; Bolleman, J. T.; Boutet, E.; Breuza, L.; Gil, B. C.; Casals-Casas, C.; Echioukh, K. C.; Coudert, E.; Cuche, B.; de Castro, E.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gaudet, P.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz, N.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Kerhornou, A.; Le Mercier, P.; Lieberherr, D.; Masson, P.; Morgat, A.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Poux, S.; Pozzato, M.; Pruess, M.; Redaschi, N.; Rivoire, C.; Sigrist, C. J. A.; Sonesson, K.; Sundaram, S.; Sveshnikova, A.; Wu, C. H.; Arighi, C. N.; Chen, C.; Chen, Y.; Huang, H.; Laiho, K.; Lehvaslaiho, M.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Y.; Zhang, J. *Nucleic Acids Res.* **2025**, *53* (D1), D609.
(171) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. *Nucleic Acids Res.* **2007**, *35* (SUPPL. 1), D301.
(172) Fleming, J.; Magana, P.; Nair, S.; Tsenkov, M.; Bertoni, D.; Pidruchna, I.; Lima Afonso, M. Q.; Midlik, A.; Paramval, U.; Žídek, A.; Laydon, A.; Kovalevskiy, O.; Pan, J.; Cheng, J.; Avsec, Ž.; Bycroft, C.; Wong, L. H.; Last, M.; Mirdita, M.; Steinegger, M.; Kohli, P.; Váradi, M.; Velankar, S. *J. Mol. Biol.* **2025**, *437* (15), 168967.
(173) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature 2021 596:7873* **2021**, *596* (7873), 583.
(174) Burley, S. K.; Bhatt, R.; Bhikadiya, C.; Bi, C.; Biester, A.; Biswas, P.; Bittrich, S.; Blaumann, S.; Brown, R.; Chao, H.; Chithari, V. R.; Craig, P. A.; Crichlow, G. V.; Duarte, J. M.; Dutta, S.; Feng, Z.; Flatt, J. W.; Ghosh, S.; Goodsell, D. S.; Green, R. K.; Guranovic, V.; Henry, J.; Hudson, B. P.; Joy, M.; Kaelber, J. T.; Khokhriakov, I.; Lai, J. S.; Lawson, C. L.; Liang, Y.; Myers-Turnbull, D.; Peisach,

E.; Persikova, I.; Piehl, D. W.; Pingale, A.; Rose, Y.; Sagendorf, J.; Sali, A.; Segura, J.; Sekharan, M.; Shao, C.; Smith, J.; Trumbull, M.; Vallat, B.; Voigt, M.; Webb, B.; Whetstone, S.; Wu-Wu, A.; Xing, T.; Young, J. Y.; Zalevsky, A.; Zardecki, C. *Nucleic Acids Res.* **2025**, *53* (D1), D564.

(175) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; De Hoon, M. J. L. *Bioinformatics* **2009**, *25* (11), 1422.

(176) Rossum, G. van; Boer, J. de. *CWI Quarterly*. December 1, 1991, pp 283–303.

(177) Krieger, E.; Vriend, G. *Bioinformatics* **2014**, *30* (20), 2981.

(178) Lee, J. H.; Yin, R.; Ofek, G.; Pierce, B. G. *Front. Immunol.* **2022**, *13*.

(179) Kajander, T.; Kahn, P. C.; Passila, S. H.; Cohen, D. C.; Lehtiö, L.; Adolfsen, W.; Warwicker, J.; Schell, U.; Goldman, A. *Structure* **2000**, *8* (11), 1203.

(180) Yu, X. X.; Liang, W. Y.; Yin, J. Y.; Zhou, Q.; Chen, D. M.; Zhang, Y. H. *Food Chem.* **2021**, *350*, 129139.

(181) Adhav, V. A.; Saikrishnan, K. *ACS Omega* **2023**, *8* (25), 22268.

(182) Davis, F. P.; Sali, A. *Bioinformatics* **2005**, *21* (9), 1901.

(183) Ferdous, S.; Martin, A. C. R. *Database* **2018**, *2018* (2018).

(184) Fernández-Quintero, M. L.; Pomarici, N. D.; Fischer, A. L. M.; Hoerschinger, V. J.; Kroell, K. B.; Riccabona, J. R.; Kamenik, A. S.; Loeffler, J. R.; Ferguson, J. A.; Perrett, H. R.; Liedl, K. R.; Han, J.; Ward, A. B. *Antibodies* **2023**, *12* (4), 67.

(185) Stave, J. W.; Lindpaintner, K. *The Journal of Immunology* **2013**, *191* (3), 1428.

(186) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. *Proteins: Structure, Function and Genetics* **2006**, *64* (3), 559.

(187) Kilambi, K. P.; Gray, J. J. *Sci. Rep.* **2017**, *7* (1), 8145.

(188) Breiteneder, H. *Allergy: European Journal of Allergy and Clinical Immunology* **2018**, *73* (11), 2107.

(189) Mueller, G. A. *Int. Arch. Allergy Immunol.* **2017**, *174* (2), 57.

(190) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696.

(191) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103* (19), 8577.

(192) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13* (8), 952.

(193) Benjamin Stranges, P.; Kuhlman, B. *Protein Science* **2013**, *22* (1), 74.

(194) Da Silva, B. M.; Myung, Y.; Ascher, D. B.; Pires, D. E. V. *Brief. Bioinform.* **2022**, *23* (1), 1.

(195) Shashkova, T. I.; Umerenkov, D.; Salnikov, M.; Strashnov, P. V.; Konstantinova, A. V.; Lebed, I.; Shcherbinin, D. N.; Asatryan, M. N.; Kardymon, O. L.; Ivanisenko, N. V. *Front. Immunol.* **2022**, *13*, 960985.

(196) Wang, M.; Zhu, D.; Zhu, J.; Nussinov, R.; Ma, B. *Journal of Molecular Recognition* **2018**, *31* (5), e2693.

(197) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55* (3), 379.

(198) Chen, J.; Sawyer, N.; Regan, L. *Protein Science* **2013**, *22* (4), 510.
(199) Chakravarty, D.; Guharoy, M.; Robert, C. H.; Chakrabarti, P.; Janin, J. *Protein Science* **2013**, *22* (10), 1453.
(200) Madsen, A. V.; Mejias-Gomez, O.; Pedersen, L. E.; Preben Morth, J.; Kristensen, P.; Jenkins, T. P.; Goletz, S. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 199.
(201) Yan, C.; Wu, F.; Jernigan, R. L.; Dobbs, D.; Honavar, V. *Protein Journal* **2008**, *27* (1), 59.
(202) Kozlova, E. E. G.; Viart, B. T.; de Avila, R. A. M.; Felicori, L. F.; Chavez-Olortegui, C. *BMC Bioinformatics* **2015**, *16* (19), 1.
(203) Andrews, B.; Zhang, S.; Schweitzer-Stenner, R.; Urbanc, B. *Biomolecules 2020, Vol. 10, Page 1121* **2020**, *10* (8), 1121.
(204) Hamelryck, T. *Proteins: Structure, Function and Genetics* **2005**, *59* (1), 38.
(205) Lollier, V.; Denery-Papini, S.; Larré, C.; Tessier, D. *Mol. Immunol.* **2011**, *48* (4), 577.
(206) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*; Humana Press, 2005; pp 571–607.
(207) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. *Nature*. Nature Publishing Group September 16, 2020, pp 357–362.
(208) McKinney, W. In *Proceedings of the 9th Python in Science Conference*; SciPy, 2010; pp 56–61.
(209) Schober, P.; Schwarte, L. A. *Anesth. Analg.* **2018**, *126* (5), 1763.
(210) Bellioni-Businco, B.; Paganelli, R.; Lucenti, P.; Giampietro, P. G.; Perborn, H.; Businco, L. *Journal of Allergy and Clinical Immunology* **1999**, *103* (6), 1191.
(211) Dean, T. P.; Adler, B. R.; Ruge, F.; Warner, J. O. *Clinical & Experimental Allergy* **1993**, *23* (3), 205.
(212) Restani, P.; Beretta, B.; Fiocchi, A.; Ballabio, C.; Galli, C. L. *Annals of Allergy, Asthma and Immunology* **2002**, *89* (6 SUPPL. 1), 11.
(213) Hilger, C.; van Hage, M.; Kuehn, A. *Curr. Allergy Asthma Rep.* **2017**, *17* (9), 64.
(214) Käck, U.; van Hage, M.; Grönlund, H.; Lilja, G.; Asarnoj, A.; Konradsen, J. R. *Clin. Transl. Allergy* **2022**, *12* (5), e12149.
(215) Posch, B.; Irsara, C.; Gamper, F. S.; Herrmann, M.; Bindreither, D.; Fuchs, D.; Reider, N.; Redl, B.; Heufler, C. *J. Cell. Mol. Med.* **2015**, *19* (10), 2375.
(216) Rickert, C. A.; Henkel, M.; Lieleg, O. *APL Machine Learning* **2023**, *1* (1), 16105.
(217) Hagberg, A. A.; Schult, D. A.; Swart, P. J. In *Proceedings of the 7th Python in Science Conference*; SciPy, 2008; pp 11–15.
(218) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9* (3), 90.

(219) Rácz, A.; Bajusz, D.; Héberger, K. *Mol. Inform.* **2019**, *38* (8–9), 1800154.
(220) Trinh, C.; Tbatou, Y.; Lasala, S.; Herbinet, O.; Meimaroglou, D. *Processes* **2023**, *11* (12), 3325.
(221) Pedregosa FABIANPEDREGOSA, F.; Michel, V.; Grisel OLIVIERGRISEL, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; Cournapeau, D.; Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Thirion, B.; Grisel, O.; Dubourg, V.; Passos, A.; Brucher, M.; Perrot andÉdouardand, M.; Duchesnay, andÉdouard; Duchesnay EDOUARDDUCHESNAY, Fré. *Journal of Machine Learning Research* **2011**, *12*, 2825.
(222) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. Pietro; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G. L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. *Nat. Methods* **2020**, *17* (3), 261.
(223) Seabold, S.; Perktold, J. In *Proceedings of the 9th Python in Science Conference*; SciPy, 2010; pp 92–96.
(224) Mari, A.; Scala, E.; Palazzo, P.; Ridolfi, S.; Zennaro, D.; Carabella, G. *Cell. Immunol.* **2006**, *244* (2), 97.
(225) Binbay, F. A.; Rathod, D. C.; George, A. A. P.; Imhof, D. *Pharmaceuticals* **2023**, *16* (12), 1662.
(226) Rodrigues, C. H. M.; Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. *Brief. Bioinform.* **2022**, *23* (4), 1.
(227) Feig, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2017**, *7* (3), e1307.
(228) May, A.; Zacharias, M. *Proteins: Structure, Function, and Bioinformatics* **2008**, *70* (3), 794.

(229) ADAMS, S. L.; BARNETT, D.; WALSH, B. J.; PEARCE, R. J.; HILL, D. J.; HOWDEN, M. E. H. *Immunol. Cell Biol.* **1991**, *69* (3), 191.
(230) Dourado, D. F. A. R.; Flores, S. C. *Sci. Rep.* **2016**, *6* (1), 25406.
(231) Kastritis, P. L.; Bonvin, A. M. J. J. *J. R. Soc. Interface* **2013**, *10* (79).
(232) Szilagyi, A.; Zhang, Y. *Curr. Opin. Struct. Biol.* **2014**, *24* (1), 10.
(233) Xue, L. C.; Rodrigues, J. P. G. L. M.; Dobbs, D.; Honavar, V.; Bonvin, A. M. J. J. *Brief. Bioinform.* **2017**, *18* (3), 458.
(234) Wang, C.; Schueler-Furman, O.; Baker, D. *Protein Science* **2005**, *14* (5), 1328.
(235) Hazebrouck, S.; Ah-Leung, S.; Bidat, E.; Paty, E.; Drumare, M. F.; Tilleul, S.; Adel-Patient, K.; Wal, J. M.; Bernard, H. *Clinical and Experimental Allergy* **2014**, *44* (4), 602.
(236) Cox, A. L.; Eigenmann, P. A.; Sicherer, S. H. *Journal of Allergy and Clinical Immunology: In Practice* **2021**, *9* (1), 82.
(237) Ehlayel, M.; Bener, A.; Abu Hazeima, K.; Al-Mesaifri, F. *ISRN Allergy* **2011**, *2011*, 1.
(238) Mansor, M.; Al-Obaidi, J. R.; Ismail, I. H.; Abidin, M. A. Z.; Zakaria, A. F.; Lau, B. Y. C.; Mohsin, A. Z.; Sukor, R.; Selamat, J.; Mahmud, N. K.; Jambari, N. N. *Mol. Immunol.* **2023**, *155*, 44.
(239) Inglingstad, R. A.; Devold, T. G.; Eriksen, E. K.; Holm, H.; Jacobsen, M.; Liland, K. H.; Rukke, E. O.; Vegarud, G. E. *Dairy Sci. Technol.* **2010**, *90* (5), 549.
(240) Jandova, Z.; Vargiu, A. V.; Bonvin, A. M. J. J. *J. Chem. Theory Comput.* **2021**, *17* (9), 5944.
(241) Heo, L.; Lee, H.; Seok, C. *Scientific Reports 2016 6:1* **2016**, *6* (1), 32153.

APPENDICES

APPENDIX I

METHOD DEVELOPMENT AND VALIDATION

Supplementary data

Table S2. 1. Complete RAW protocol descriptor output and metadata for MLERs.

Access this table from the google drive.

Table S2. 2. Complete MIN protocol descriptor output and metadata for MLERs.

Access this table from the google drive.

Table S2. 3. Complete RAW protocol descriptor output and metadata for MCER region

Access this table from the google drive.

Table S2. 4. Complete MIN protocol descriptor output and metadata for MCER region

Access this table from the google drive.

Table S2. 5. Complete RAW protocol descriptor output for experimental IgE Fab–allergen template complexes and comparative Fab–query allergen models.

Access this table from the google drive.

Table S2. 6. Complete MIN protocol descriptor output for experimental IgE Fab–allergen template complexes and comparative Fab–query allergen models.

Access this table from the google drive.

APPENDIX II

METHOD APPLICATION

Supplementary tables

Table S3. 1. Lipocalin structures ranked using Euclidean distance in PCA spaced based on retained linear epitope descriptors.

Access this table from the google drive.

Table S3. 2. Lipocalin structures ranked using Euclidean distance in PCA spaced based on retained conformational epitope descriptors.

Access this table from the google drive.

Table S3. 3. Lipocalin structures ranked using Euclidean distance in PCA spaced based on retained interface descriptors.

Access this table from the google drive.

Table S3. 4. Lipocalin structures ranked using Euclidean distance in PCA spaced based on retained epitope descriptors.

Access this table from the google drive.

Table S3. 5. Lipocalin structures ranked using Euclidean distance in PCA spaced based on retained paratope descriptors.

Access this table from the google drive.

Table S3. 6. Lipocalin structures ranked using Euclidean distance in PCA spaced based on retained combined descriptors.

Access this table from the google drive.

Table S3. 7. Pairwise permutation-test results for mapped linear epitope distances.

Access this table from the google drive.

Table S3. 8. . Curated set of literature-reported Bos d 5 linear IgE-binding regions.

Published linear IgE-binding regions of Bos d 5 were curated from epitope-mapping studies. Residue positions are reported relative to the UniProt Bos d 5 sequence P02754.

No.	Epitope position /P02754	Epitope sequence	Reference
1	1-15/17-31	LIVTQTMKGLDIQKV	(Cong & Li, 2012)
2	1-16/17-32	LIVTQTMKGLDIQKVA	(Järvinen et al., 2001)
3	9–29/25-45	GLDIQKVAGTWYSLAMAASDI	(X. Wang et al., 2023)
4	31-60/47-76	LLDAQSAPLRVYVEELKPTPEGDLEILLQK	(Järvinen et al., 2001)
5	41-60/57-76	VYVEELKPTPEGDLEILLQK	(Sélo et al., 1999)
6	45–57//61-73	ELKPTPEGDLEIL	(X. Wang et al., 2023)
7	56-70/72-86	ILLQKWENGECAQKK	(Cong & Li, 2012)
8	58-77/74-93	LQKWENGECAQKKIIAEKTK	(Cerecedo et al., 2008)
9	67-86/83-102	AQKKIIAEKTKIPAVFKIDA	(Järvinen et al., 2001)
10	76-90/92-106	TKIPAVFKIDALNEN	(Cong & Li, 2012)
11	77–80/93-96	KIPA	(X. Wang et al., 2023)
12	95-113/111-129	LDTDYKKYLLFCMENSAEP	(Heinzmann et al., 1999)
13	97-108/113-124	TDYKKYLLFCME	(Ball et al., 1994)
14	98–101/114-117	DYKK	(X. Wang et al., 2023)
15	102-124/118-140	YLLFCMENSAEPEQSLVCQCLVR	(Sélo et al., 1999)
16	121–135/137-151	CLVRTPEVDDEALEK	(X. Wang et al., 2023)
17	124-134/140-150	RTPEVDDEALE	(ADAMS et al., 1991)
18	127-156/143-172	EVDDEALEKFDKALKALPMHIRLSFNPTQL	(Järvinen et al., 2001)
19	136-150/152-166	FDKALKALPMHIRLS	(Cong & Li, 2012)
20	149-162/165-178	LSFNPTQLEEQCHI	(Sélo et al., 1999)

Supplementary figures

Descriptor correlation structure and redundancy

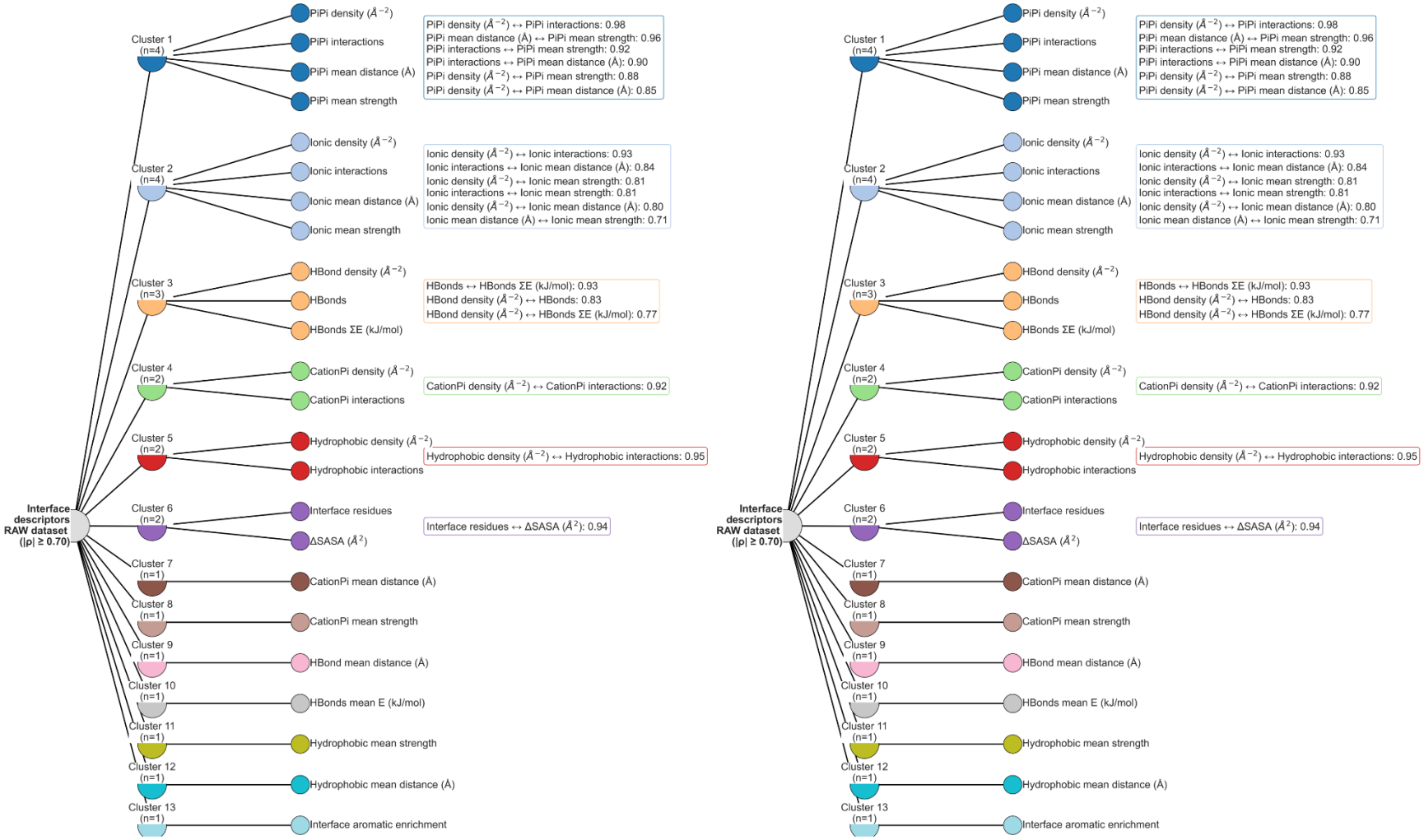

Figure S3. 1. Correlation networks of RAW and MIN Fab–allergen interface descriptors.

Nodes represent interface descriptors, and edges connect descriptor pairs with absolute Spearman correlation $|\rho| \geq 0.70$. Networks are shown separately for RAW and MIN interface descriptor datasets. Edge labels indicate pairwise $|\rho_s|$ values for clustered descriptors.

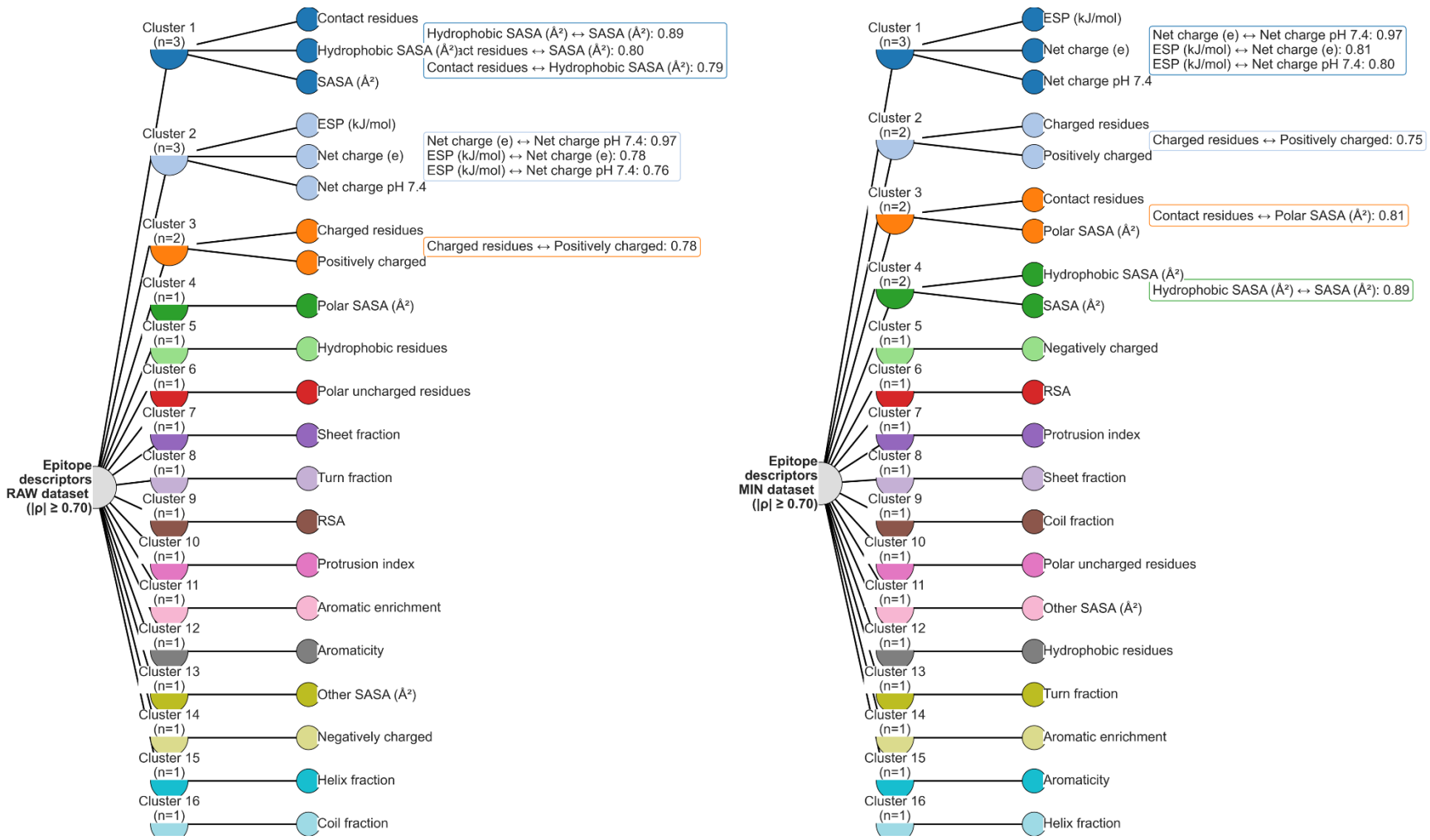

Figure S3. 1. Correlation networks of epitope descriptors.

Nodes represent interface descriptors, and edges connect descriptor pairs with absolute Spearman correlation |ρ|≥0.70. Networks are shown separately for RAW and MIN interface descriptor datasets. Edge labels indicate pairwise |ρs| values for clustered descriptors.

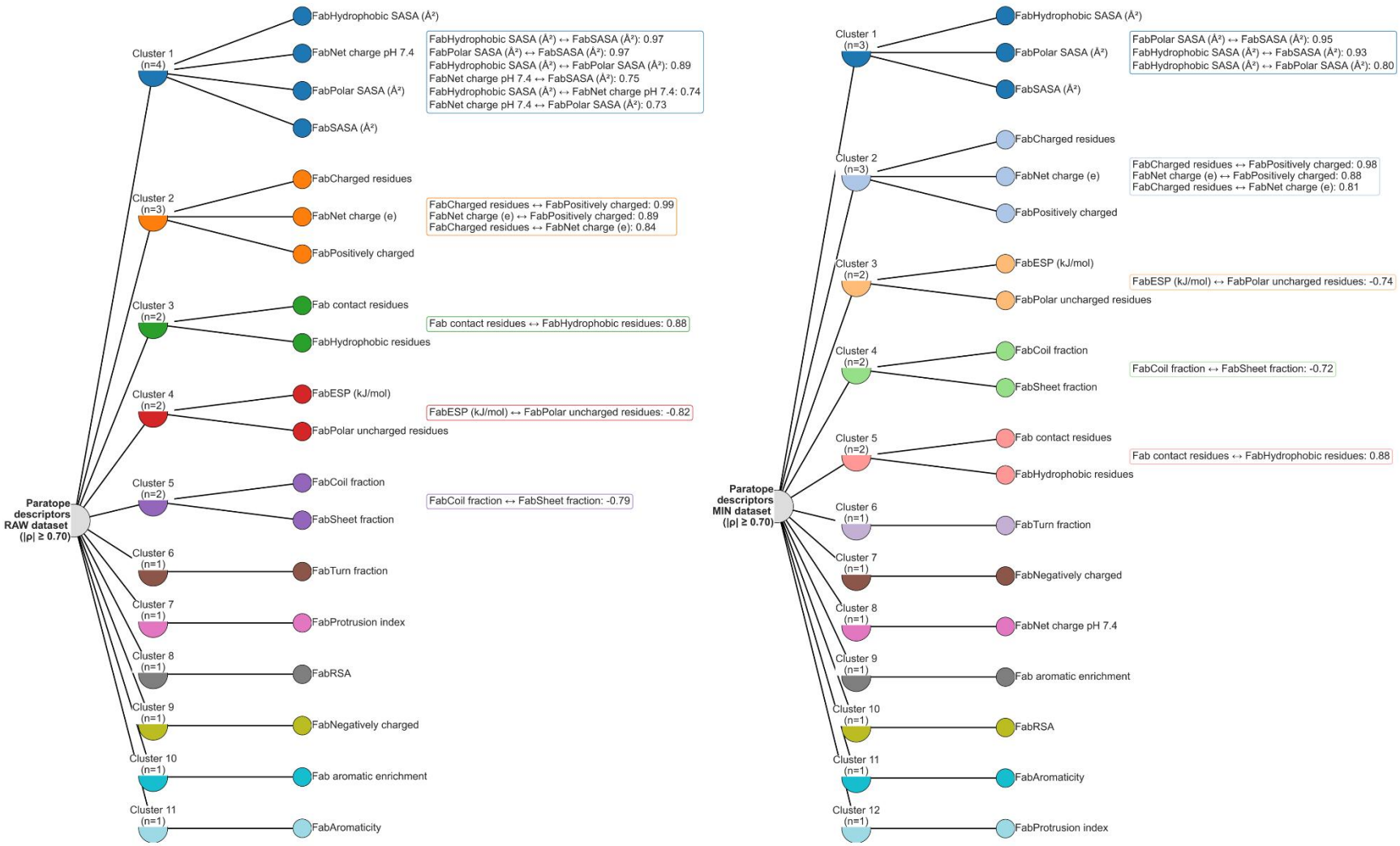

Figure S3. 2. Correlation networks of paratope descriptors.

Nodes represent interface descriptors, and edges connect descriptor pairs with absolute Spearman correlation |ρ|≥0.70. Networks are shown separately for RAW and MIN interface descriptor datasets. Edge labels indicate pairwise |ρs| values for clustered descriptors.

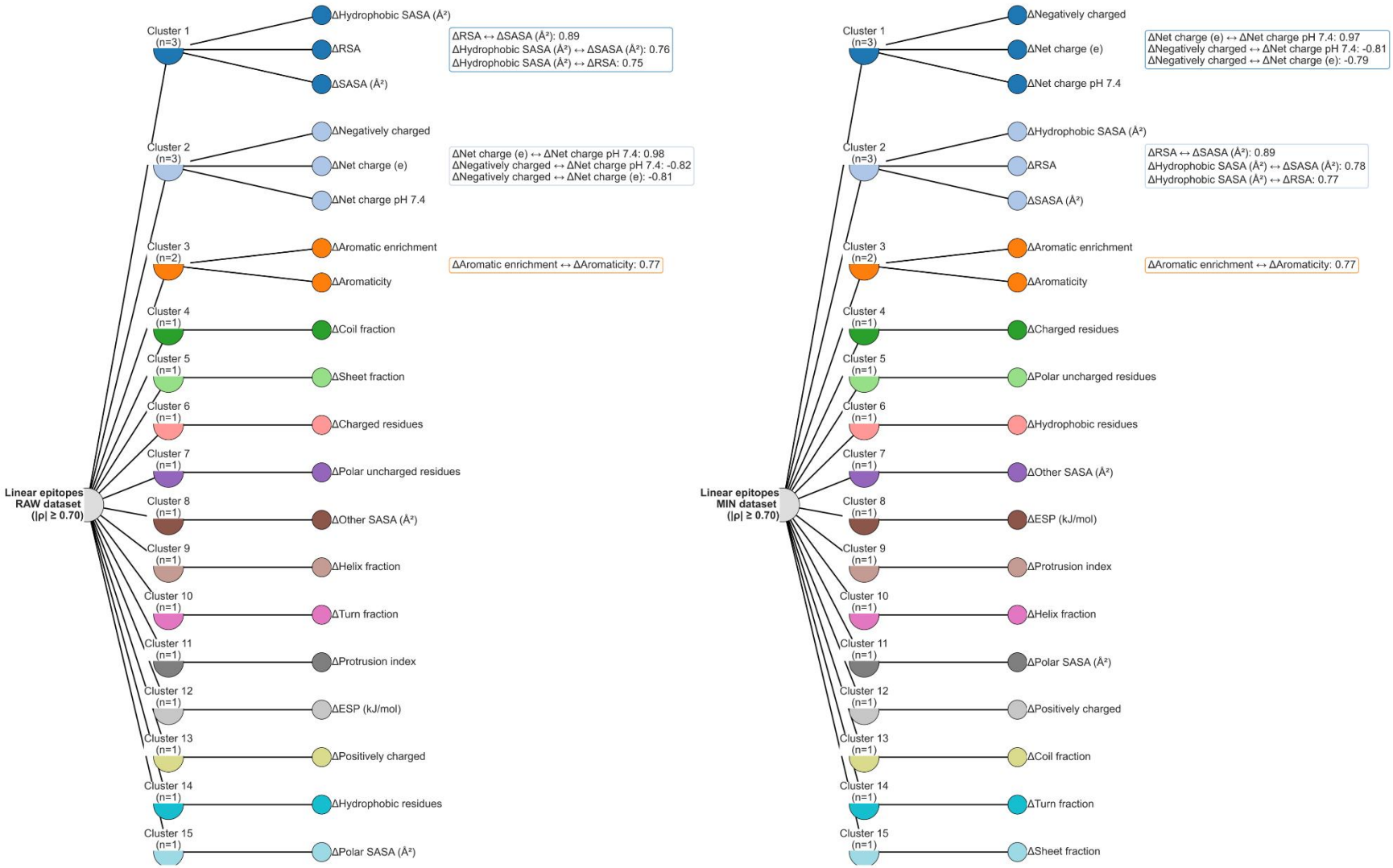

Figure S3. 3. Correlation networks of MLER descriptors.

Nodes represent interface descriptors, and edges connect descriptor pairs with absolute Spearman correlation |ρ|≥0.70. Networks are shown separately for RAW and MIN interface descriptor datasets. Edge labels indicate pairwise |ρs| values for clustered descriptors.

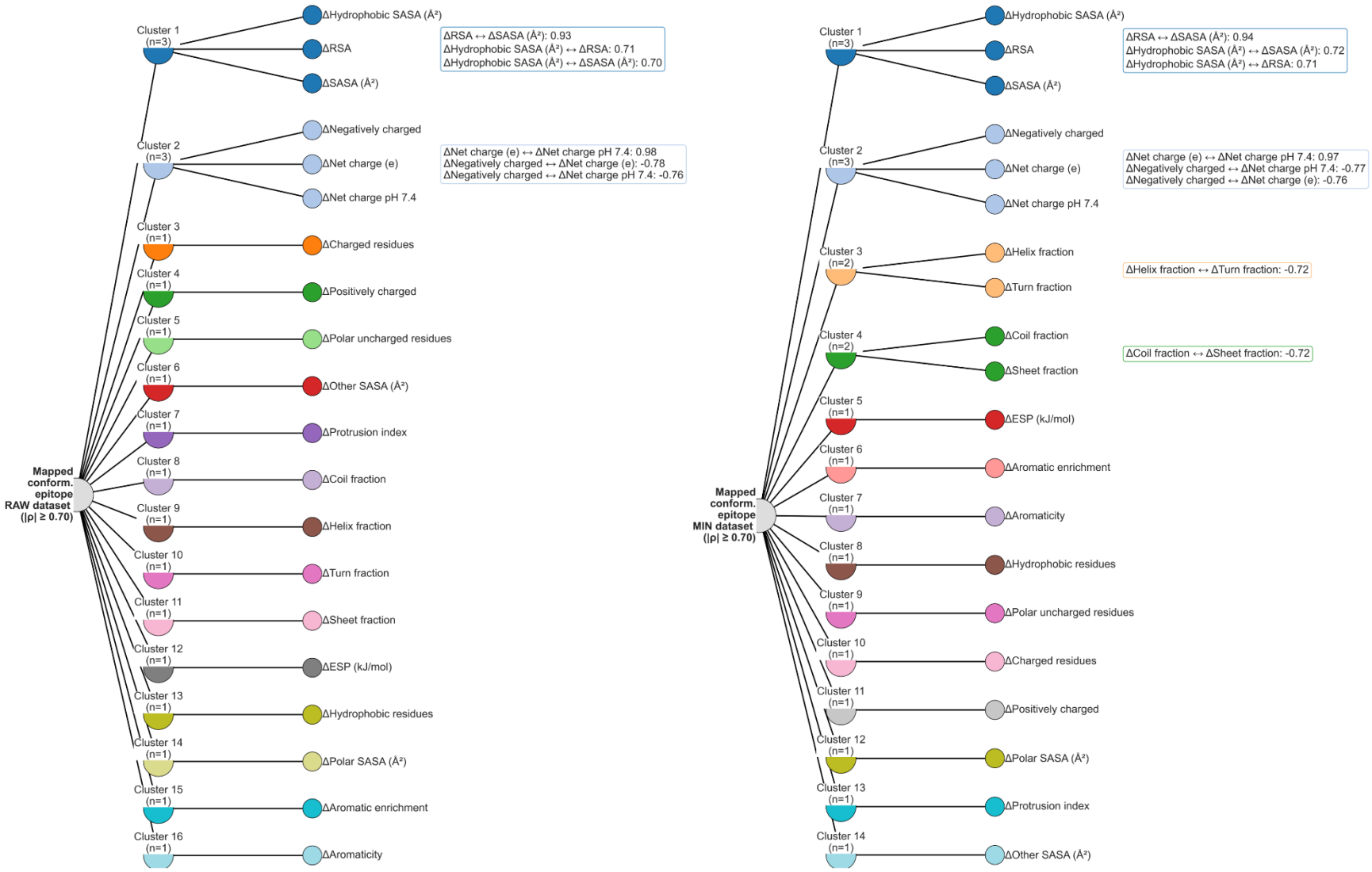

Figure S3. 4. Correlation networks of MCER descriptors.

Correlation networks of the mapped conformational epitope region (MCER) descriptors. Nodes represent interface descriptors, and edges connect descriptor pairs with absolute Spearman correlation |ρ|≥0.70. Networks are shown separately for RAW and MIN interface descriptor datasets. Edge labels indicate pairwise |ρs| values for clustered descriptors.

Exploratory PCA

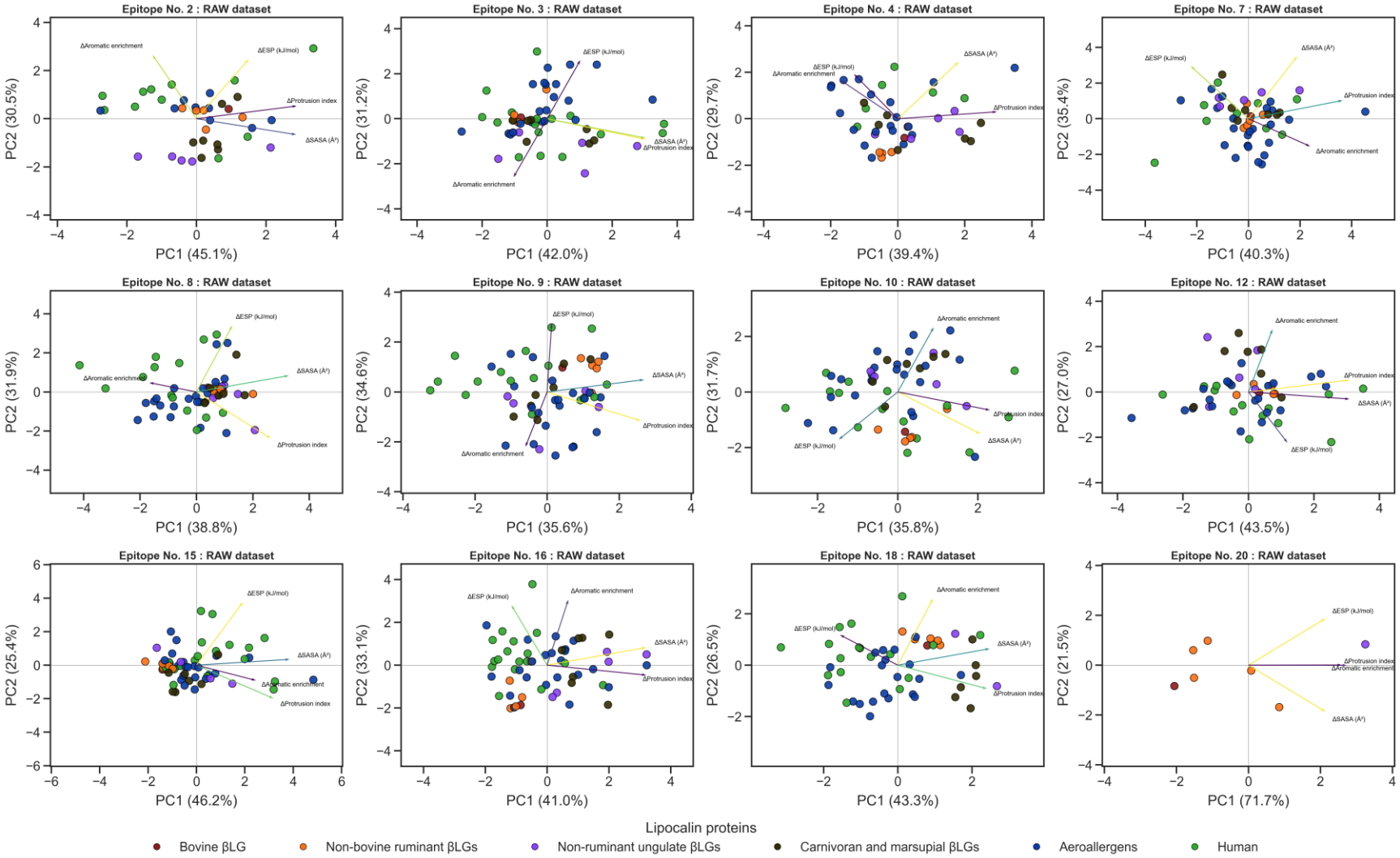

Figure S3. 5. RAW dataset PCA biplots of retained descriptor differences for MLERs.

PC1–PC2 score distribution after applying the structural coverage filter are shown for each mapped linear epitope region (MLER). Points represent query structures and the Bos d 5 reference, colored by lipocalin protein group. Arrows indicate loadings for the retained descriptor differences. Epitope 20 mostly contains retained close homologous non-bovine ruminant βLG mappings after filtering.

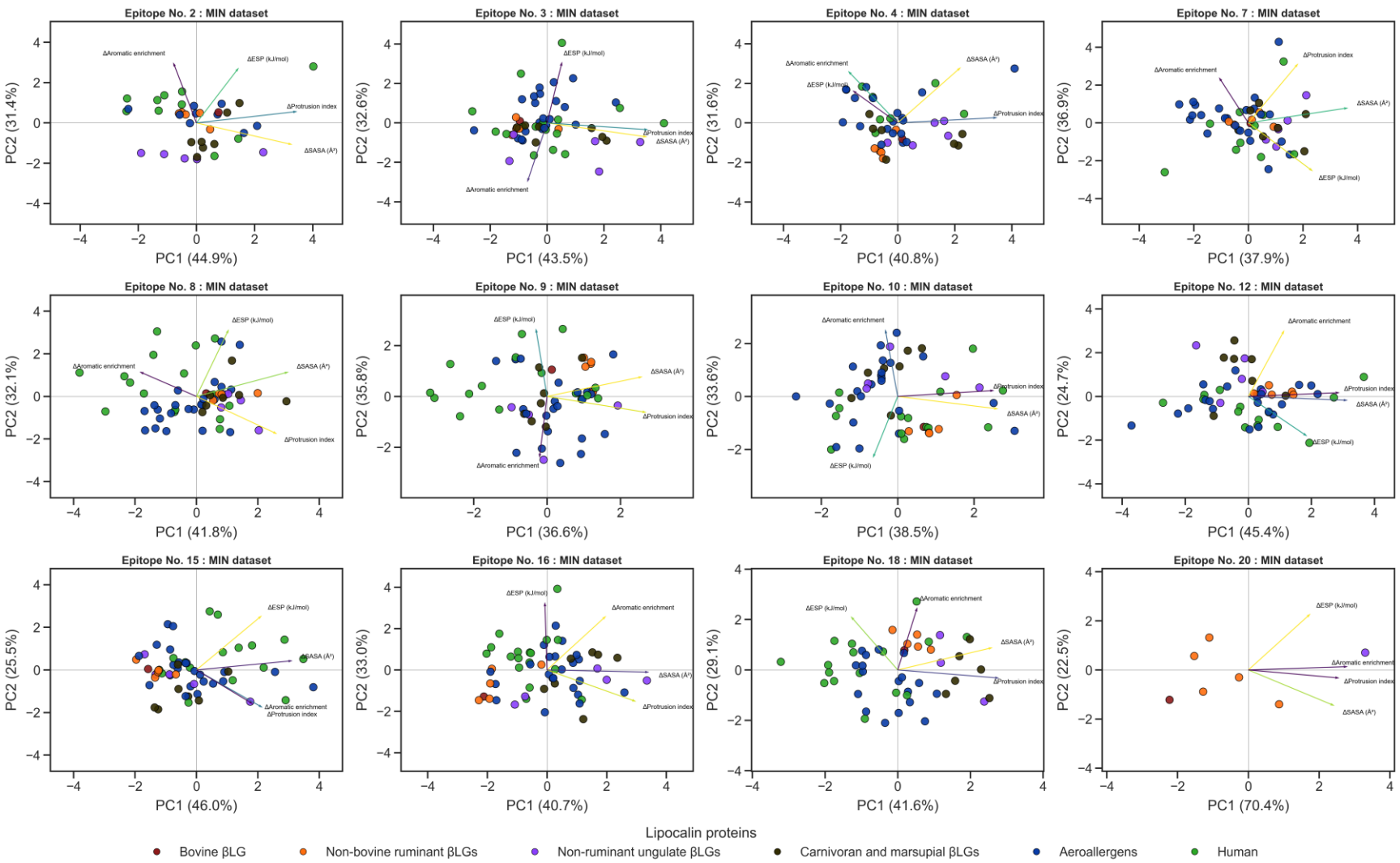

Figure S3. 6. MIN dataset PCA biplots of retained descriptor differences for MLERs.

PC1–PC2 score distribution after applying the structural coverage filter are shown for each mapped linear epitope region (MLER). Points represent query structures and the Bos d 5 reference, colored by lipocalin protein group. Arrows indicate loadings for the retained descriptor differences. Epitope 20 mostly contains retained close homologous non-bovine ruminant βLG mappings after filtering.

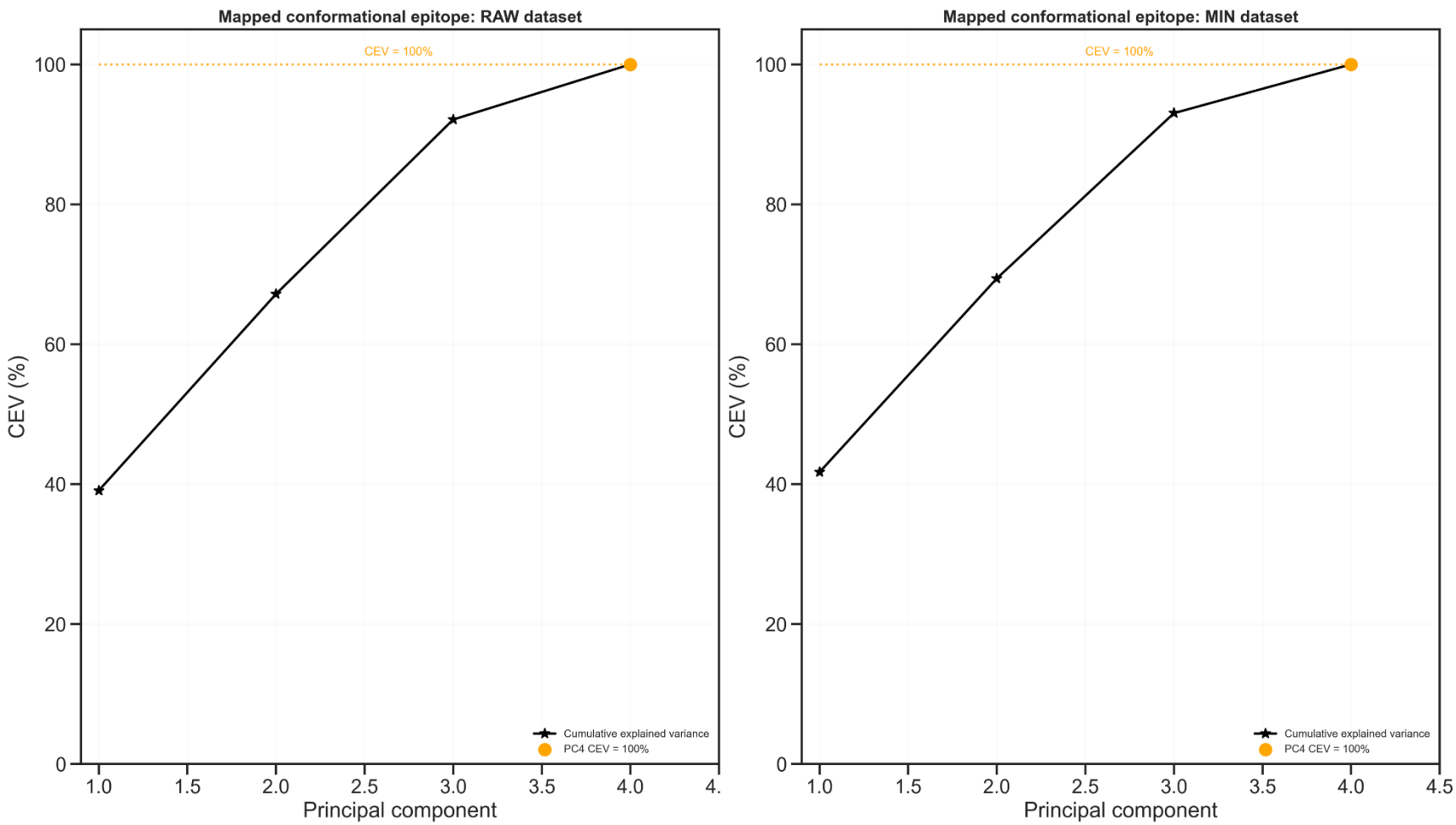

Figure S3. 7. CEV of PCA models fitted to MCER descriptors.

Cumulative explained variance (CEV) is shown for PCA models fitted to the retained mapped conformational epitope region (MCER) descriptor set in the RAW and MIN datasets. In both datasets, PC1–PC3 captured more than 90% of the variance, while PC1–PC4 captured 100% of the retained descriptor variance.

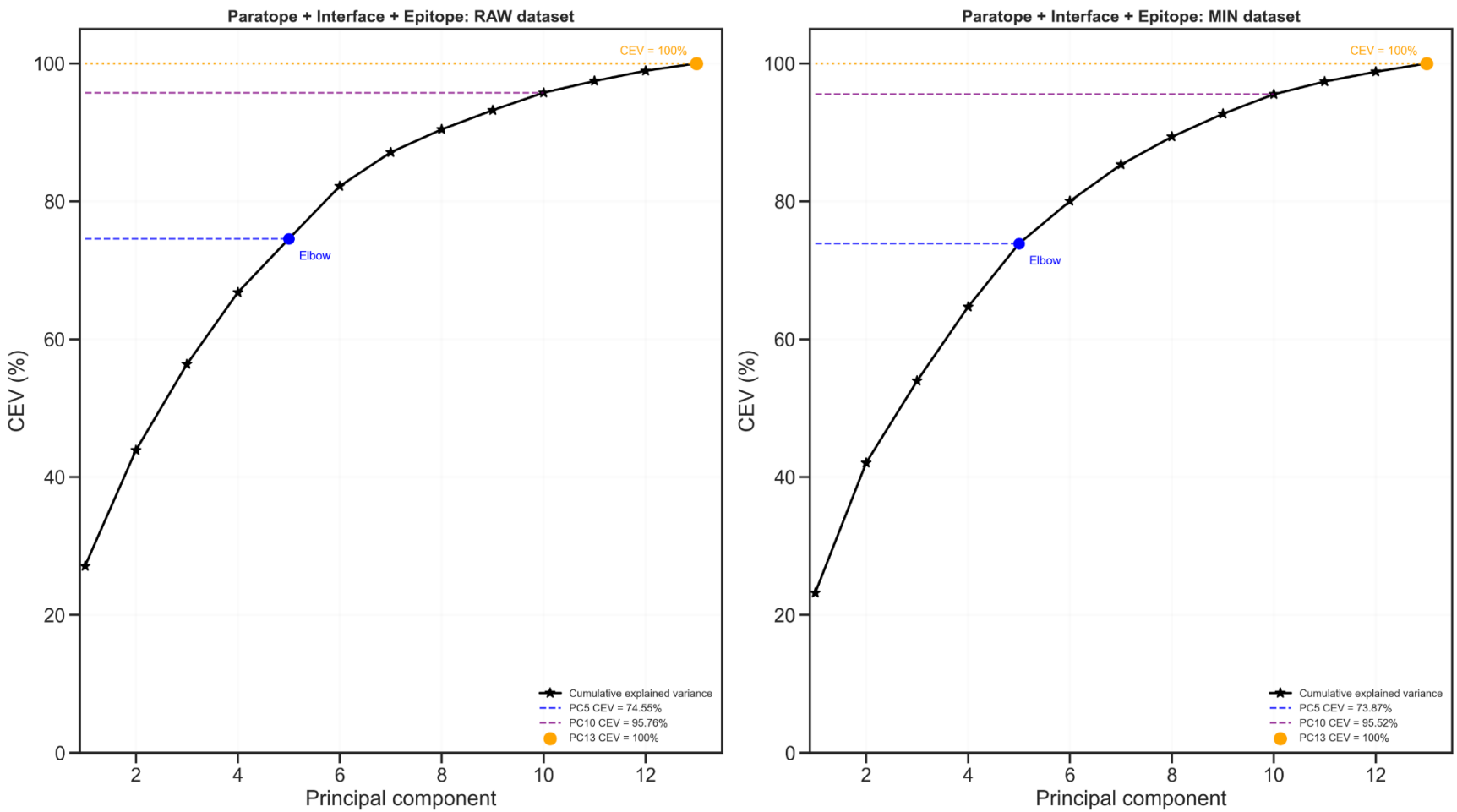

Figure S3. 8. CEV of PCA models fitted to composite descriptors.

Cumulative explained variance (CEV) is shown for PCA models fitted to the retained composite descriptor set in the RAW and MIN datasets. The combined descriptor set included comparative paratope, interface, and epitope descriptors. In the RAW dataset, PC1–PC5 captured 74.55% of variance and PC1–PC10 captured 95.76%. In the MIN dataset, PC1–PC5 captured 73.87% of variance and PC1–PC10 captured 95.52%. All 13 components captured 100% of the retained descriptor variance.

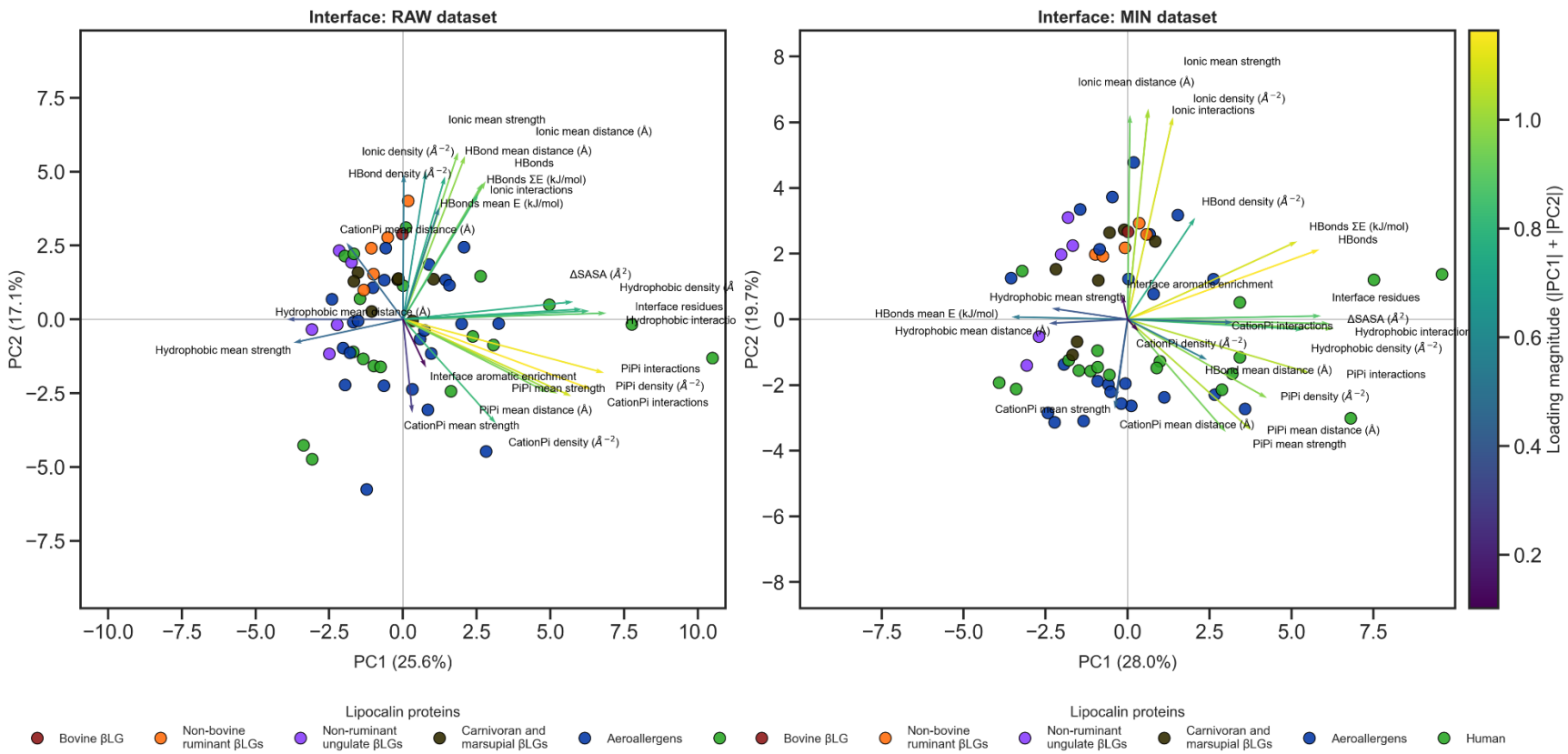

Figure S3. 9. PCA biplots of full interface descriptors.

PCA biplots are shown for the full interface descriptor set in the RAW and MIN datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate descriptor loadings from interface descriptor classes. Axis labels show the percentage of variance explained by PC1 and PC2.

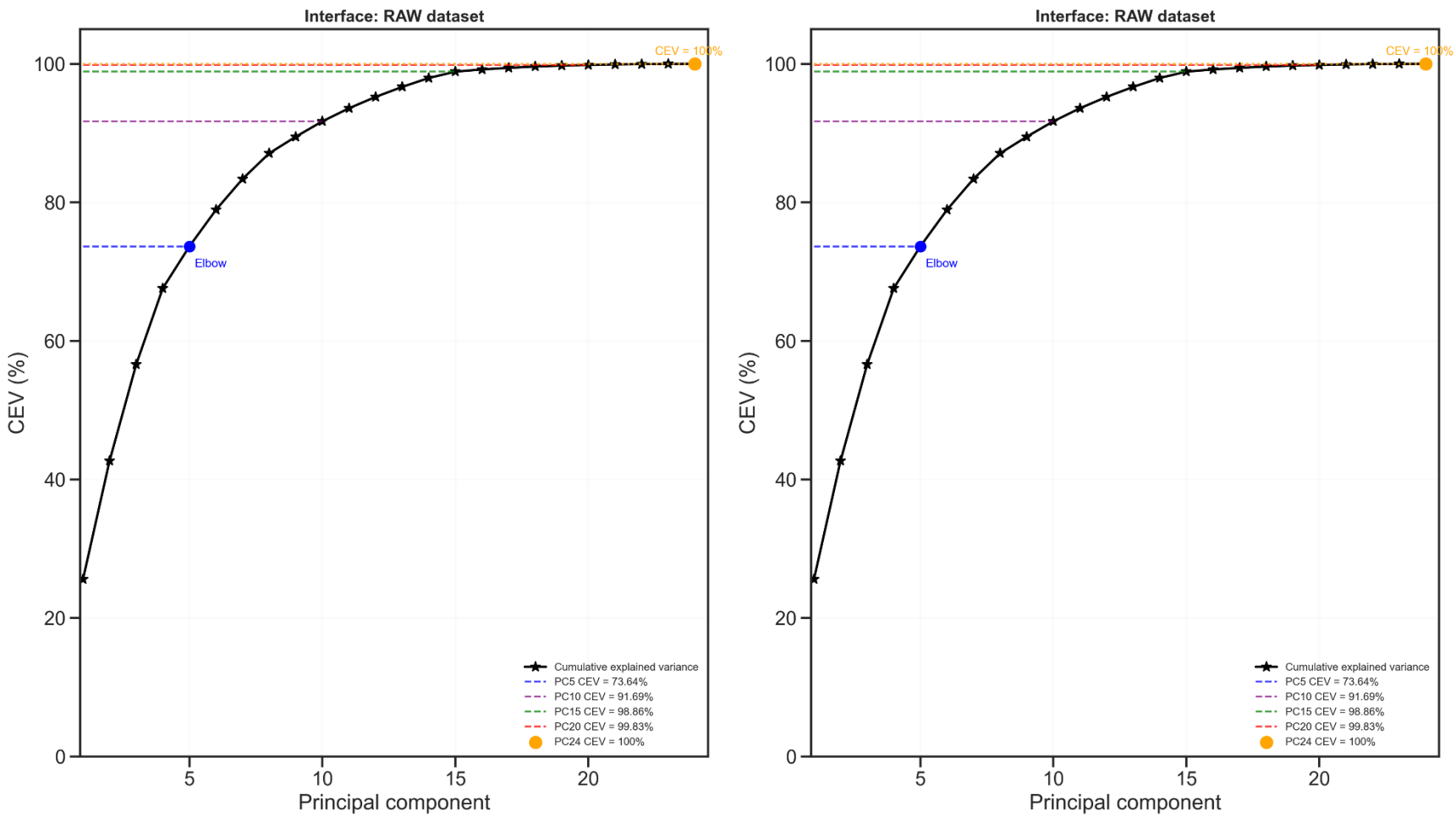

Figure S3. 10. Cumulative explained variance of PCA models fitted to interface descriptors.

Cumulative explained variance is shown for PCA models fitted to the full interface descriptor set in the RAW and MIN datasets. The combined descriptor set included comparative paratope, interface, and epitope descriptors.

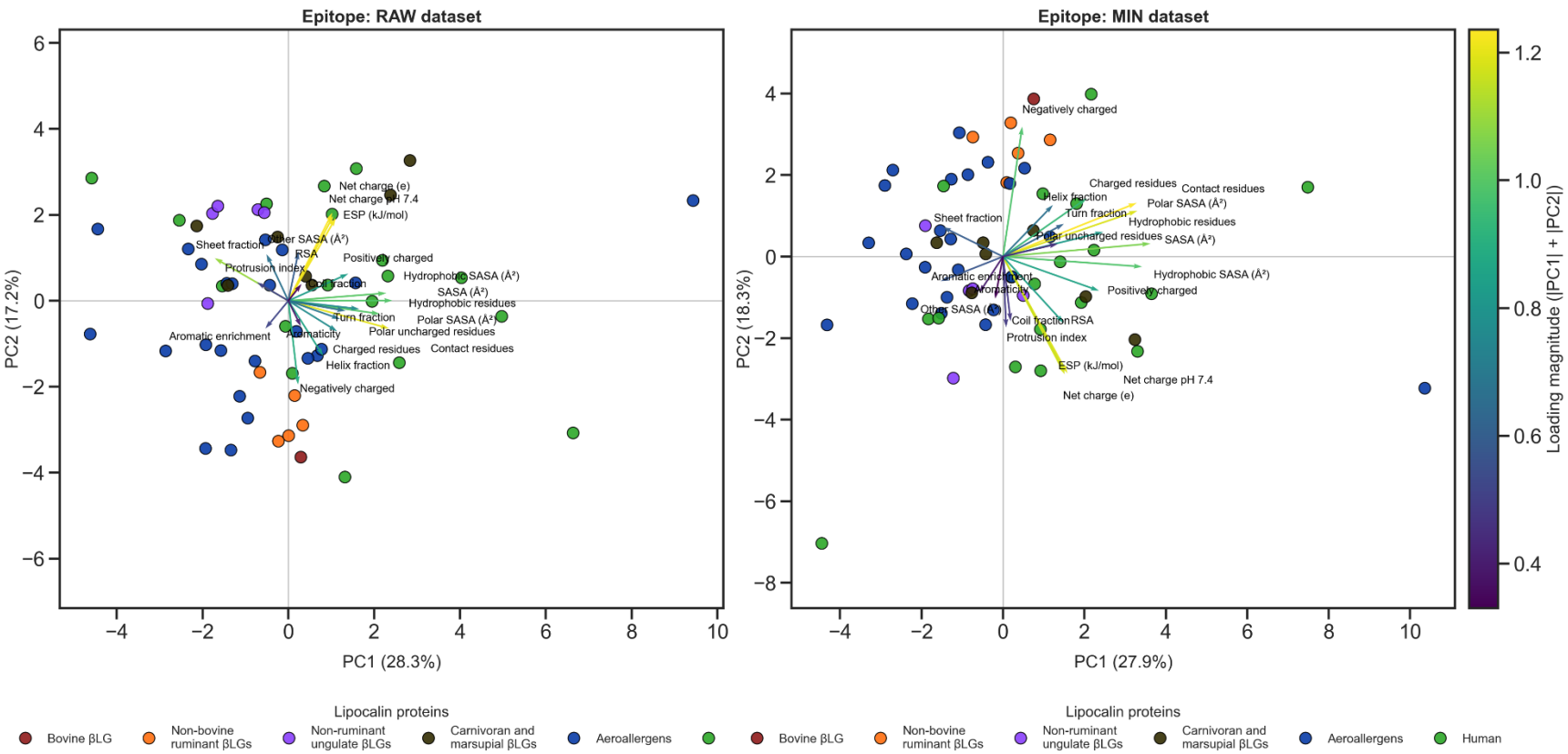

Figure S3. 11. PCA biplots of full epitope descriptors.

PCA biplots are shown for the full epitope descriptor set in the RAW and MIN datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate descriptor loadings from epitope descriptor classes. Axis labels show the percentage of variance explained by PC1 and PC2.

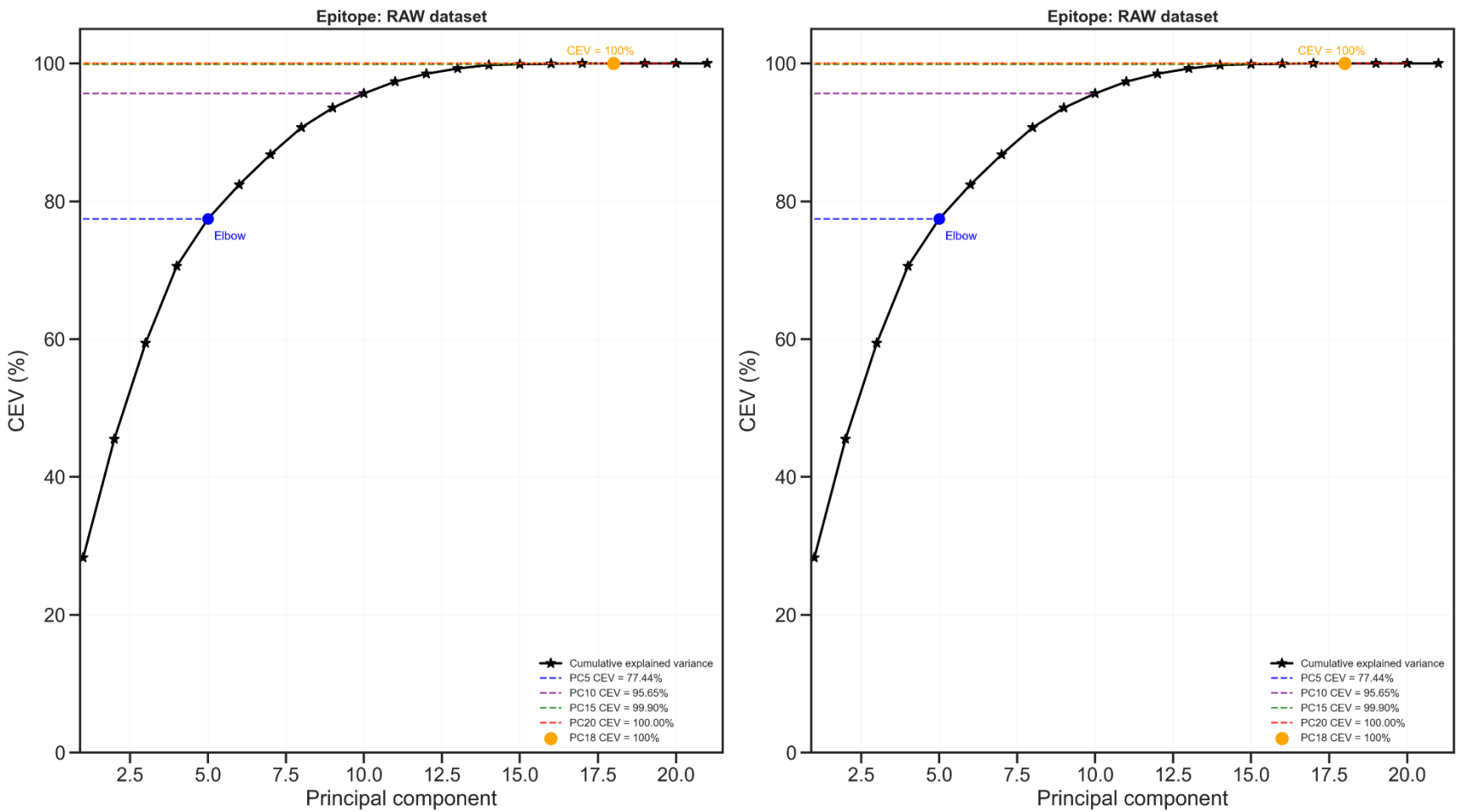

Figure S3. 12. Cumulative explained variance of PCA models fitted to epitope descriptors.

Cumulative explained variance is shown for PCA models fitted to the full epitope descriptor set in the RAW and MIN datasets. The combined descriptor set included comparative paratope, interface, and epitope descriptors.

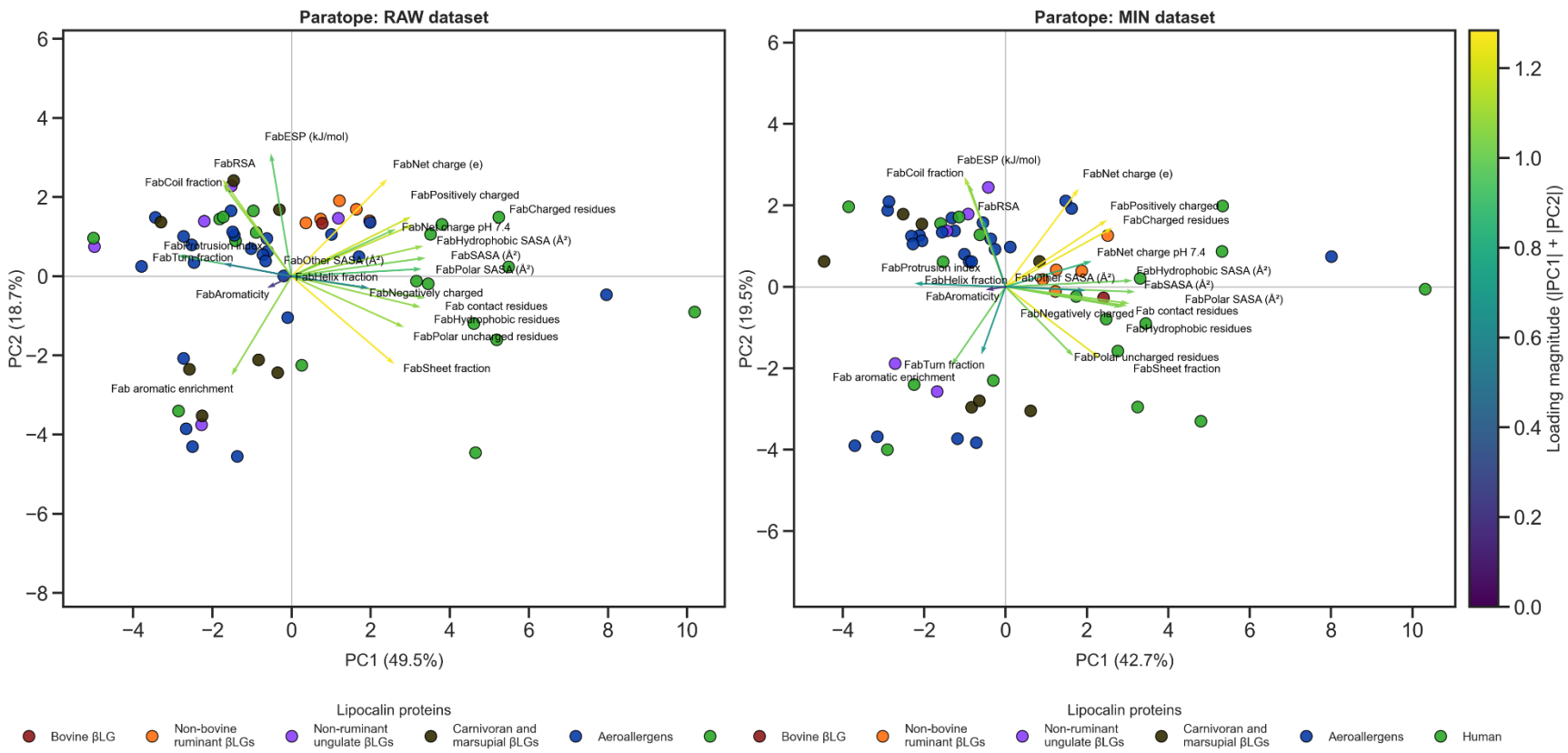

Figure S3. 13. PCA biplots of full paratope descriptors.

PCA biplots are shown for the full paratope descriptor set in the RAW and MIN datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate descriptor loadings from epitope descriptor classes. Axis labels show the percentage of variance explained by PC1 and PC2.

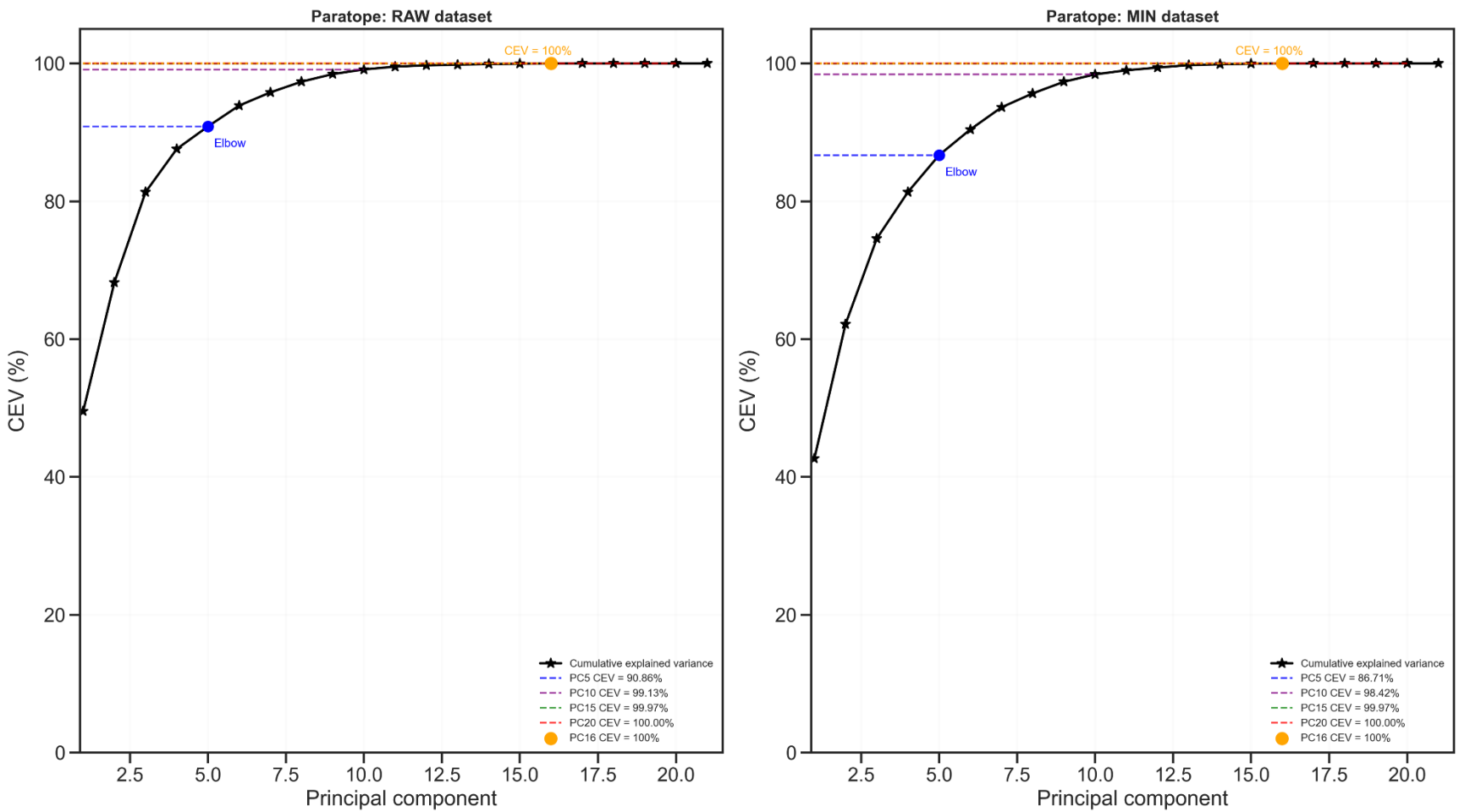

Figure S3. 14. Cumulative explained variance of PCA models fitted to paratope descriptors.

Cumulative explained variance is shown for PCA models fitted to the full paratope descriptor set in the RAW and MIN datasets. The combined descriptor set included comparative paratope, interface, and epitope descriptors.

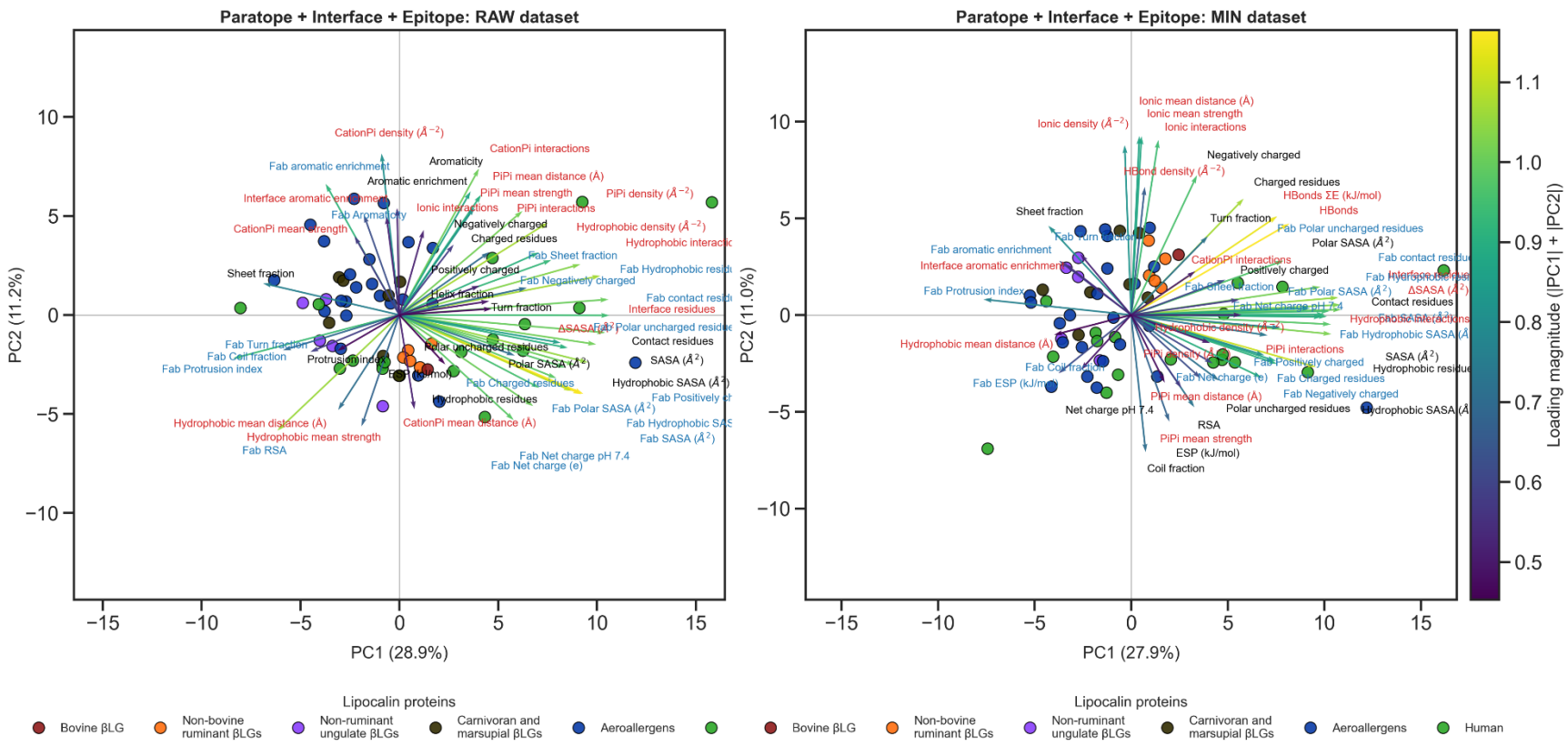

Figure S3. 15. PCA biplots of full composite descriptors set.

PCA biplots are shown for the full composite descriptor set in the RAW and MIN datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate descriptor loadings from comparative epitope, paratope, and interface descriptor classes. Axis labels show the percentage of variance explained by PC1 and PC2.

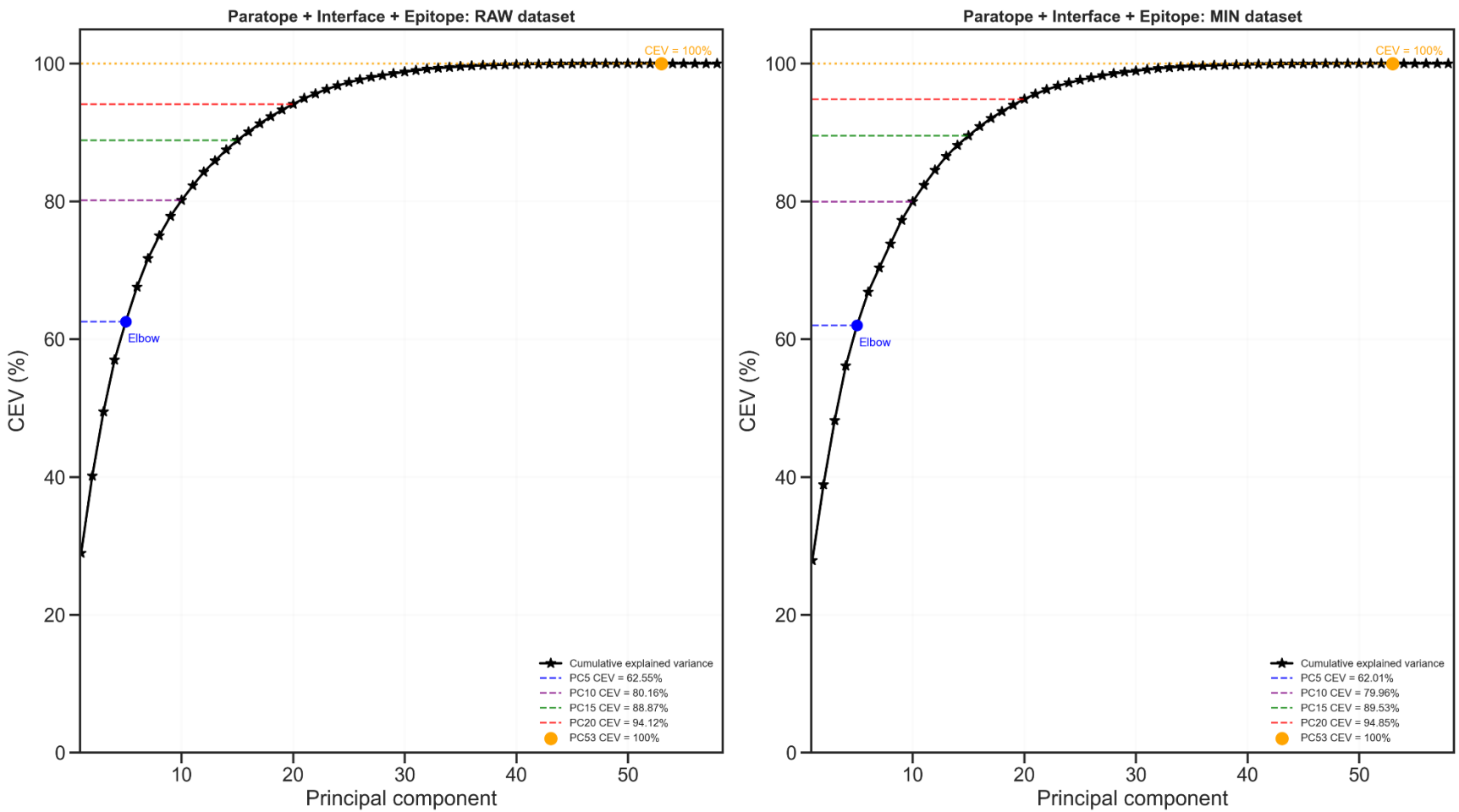

Figure S3. 16. Cumulative explained variance of PCA models fitted to composite descriptor set.

Cumulative explained variance is shown for PCA models fitted to the full composite descriptor set in the RAW and MIN datasets. The combined descriptor set included comparative paratope, interface, and epitope descriptors.

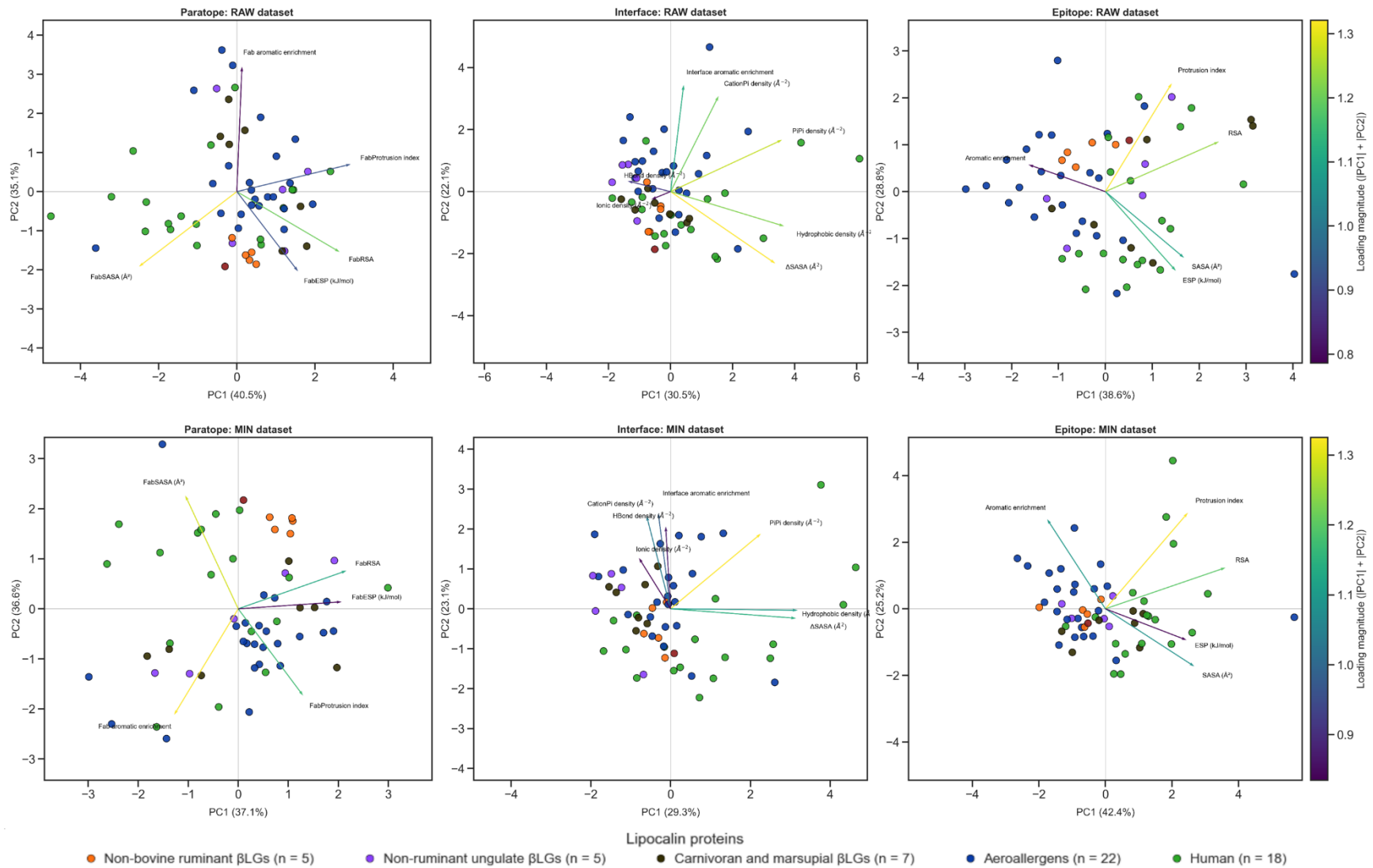

Figure S3. 17. PCA biplots of individual comparative Fab-allergen model descriptor targets.

PCA biplots are shown separately for comparative paratope, interface, and epitope descriptor sets in the RAW and MIN datasets. Points represent query lipocalin structures, coloured by biological group. Arrows indicate descriptor loadings for each retained descriptor set. Axis labels show the percentage of variance explained by PC1 and PC2.

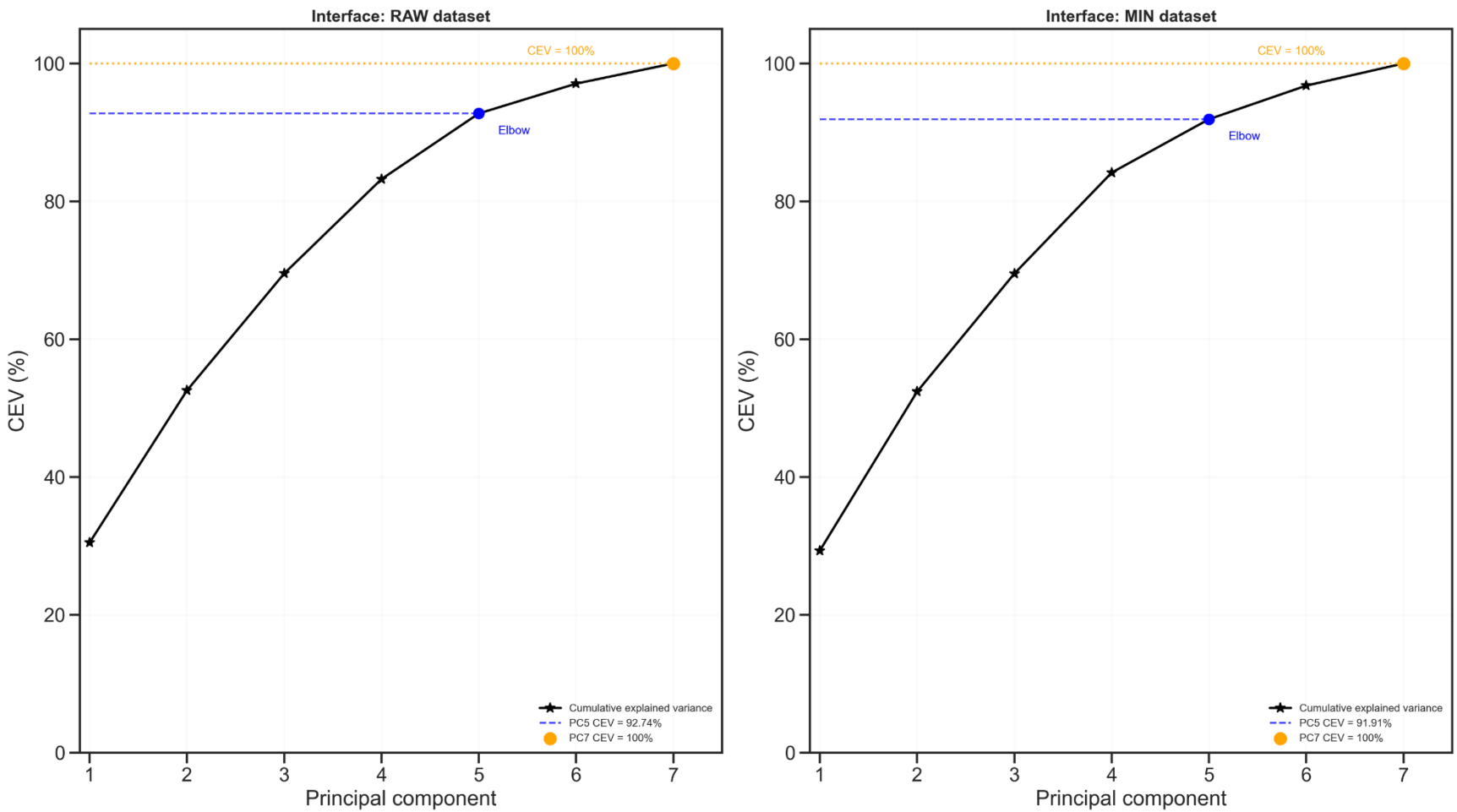

Figure S3. 18. Cumulative explained variance of PCA models fitted to retained interface descriptors.

Cumulative explained variance is shown for PCA models fitted to the retained interface descriptor set in the RAW and MIN datasets.

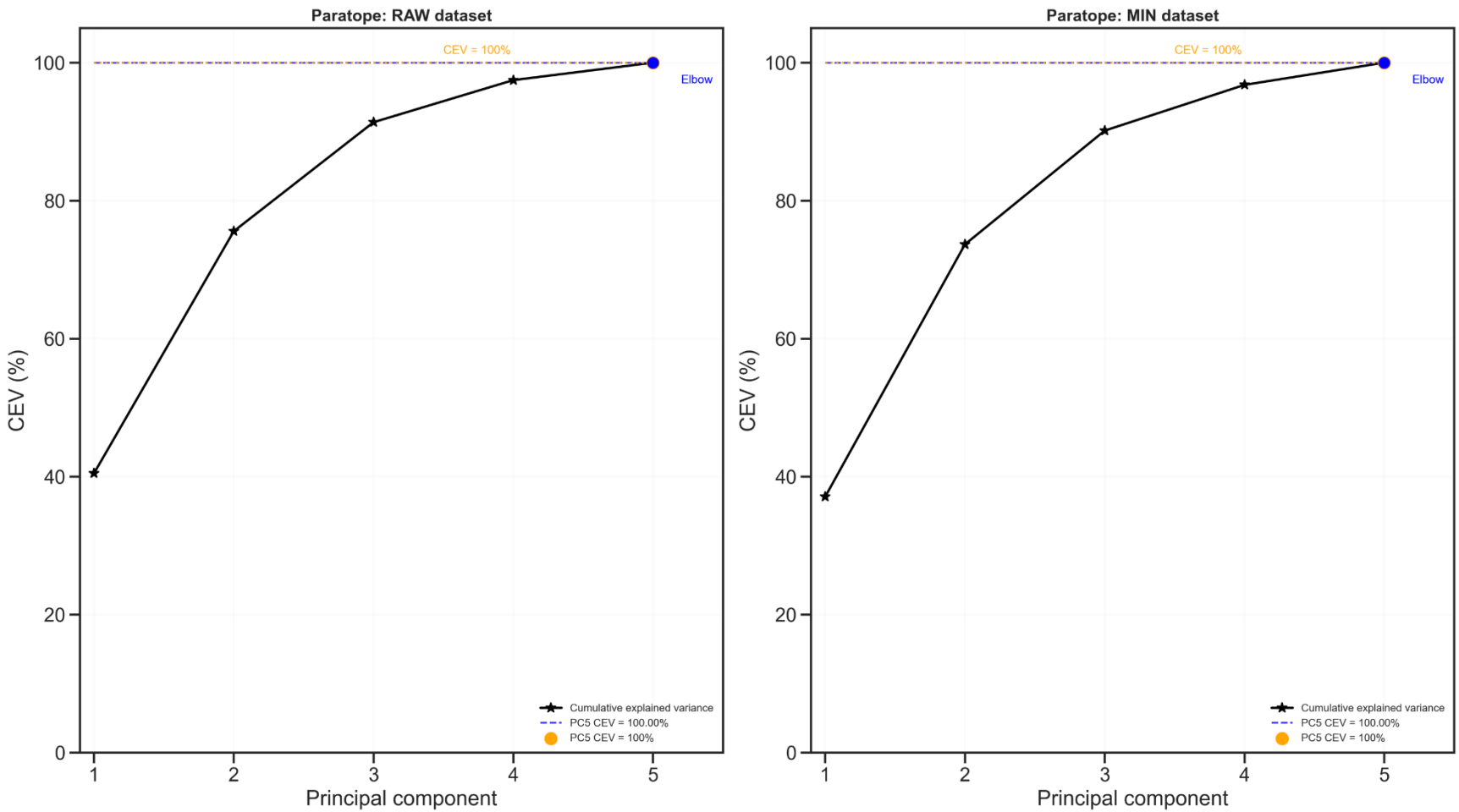

Figure S3. 19. Cumulative explained variance of PCA models fitted to retained paratope descriptors.

Cumulative explained variance is shown for PCA models fitted to the retained paratope descriptor set in the RAW and MIN datasets.

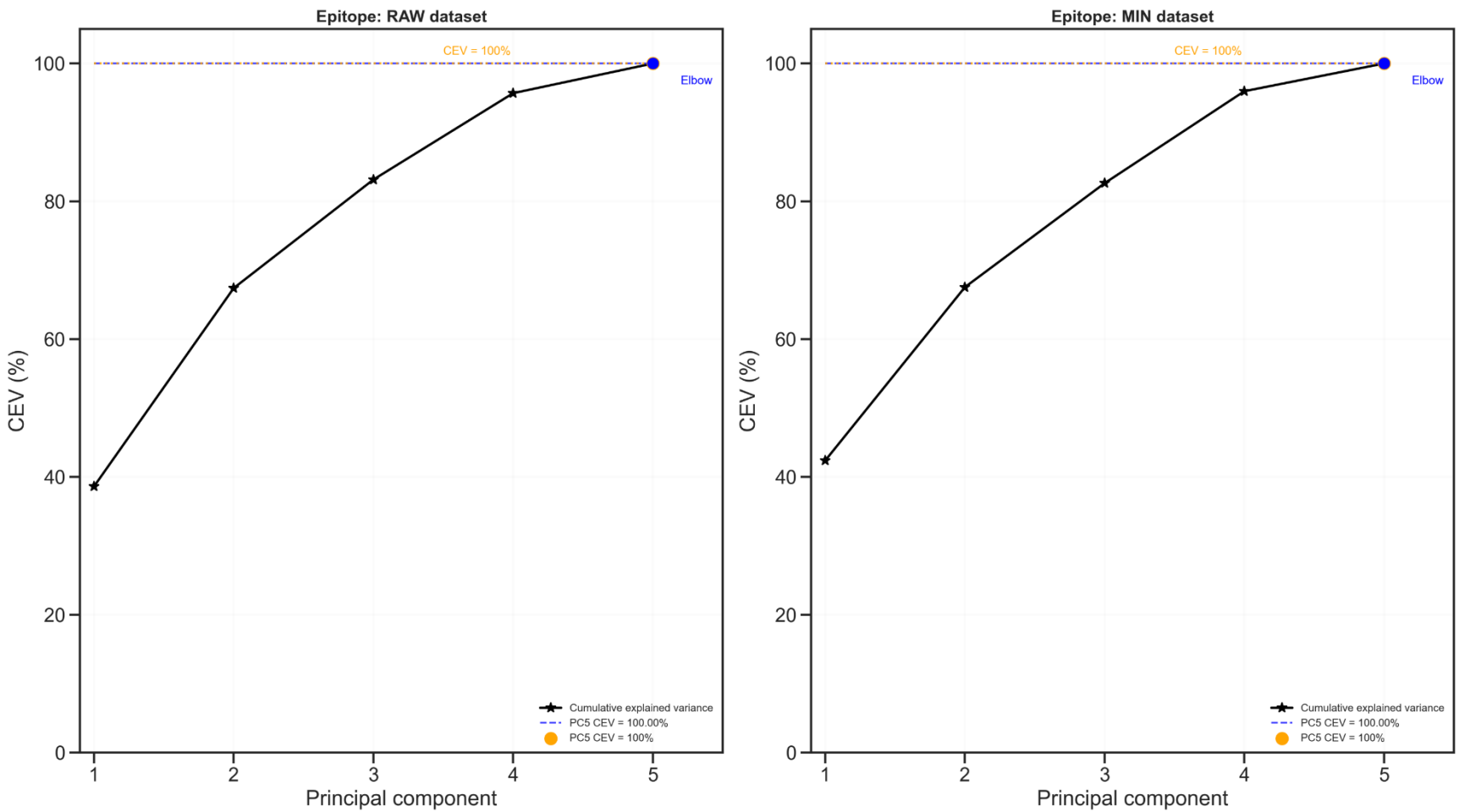

Figure S3. 20. Cumulative explained variance of PCA models fitted to retained epitope descriptors.

Cumulative explained variance is shown for PCA models fitted to the retained epitope descriptor set in the RAW and MIN datasets.

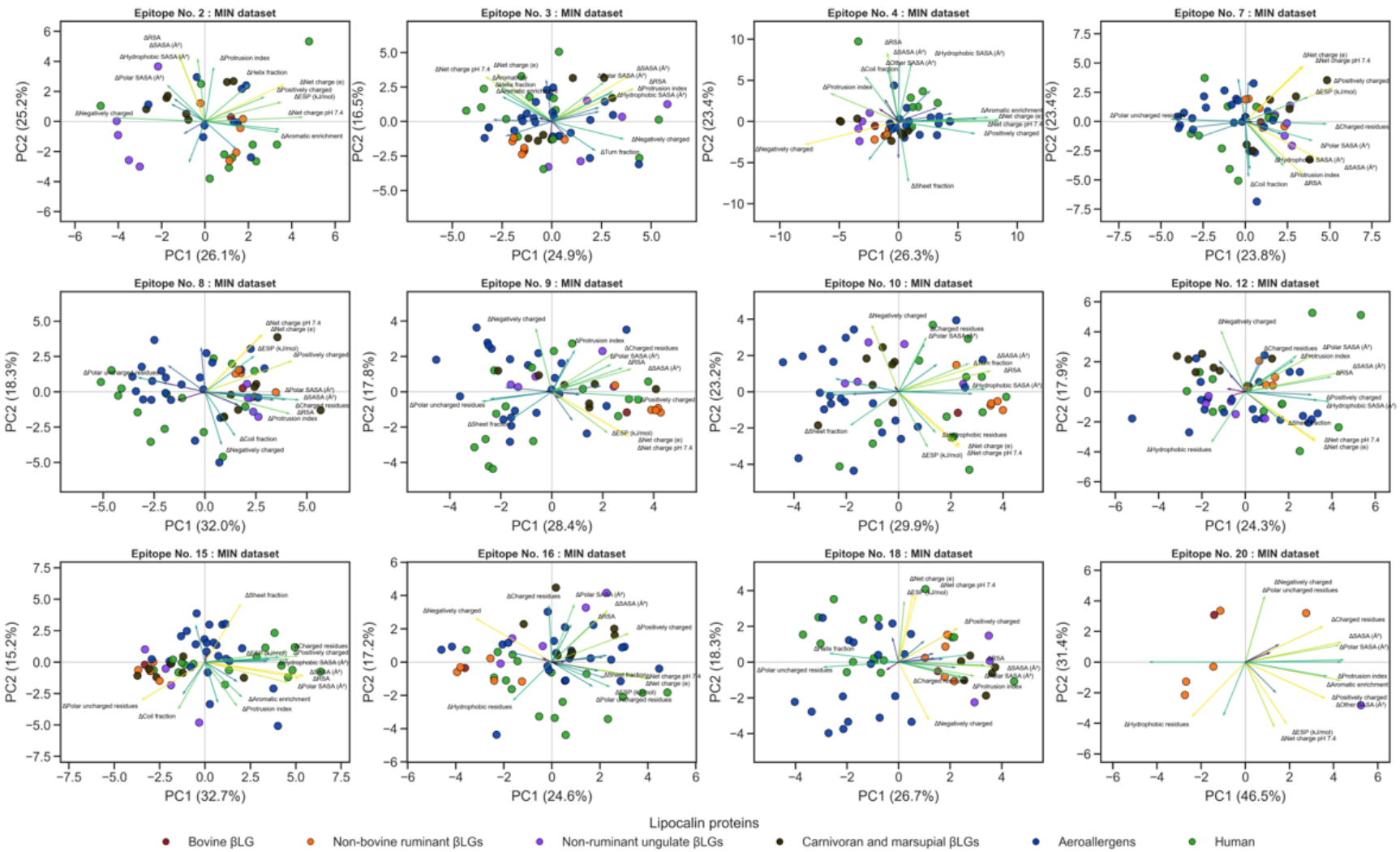

Figure S3. 21. PCA biplots of full Descriptor differences relative to Bos d 5 for MLERs in the MIN dataset.

PC1–PC2 score distribution after applying the structural coverage filter are shown for each mapped linear epitope region (MLER). Points represent query structures and the Bos d 5 reference, colored by lipocalin protein group. Arrows indicate loadings for the full descriptor differences. Epitope 20 mostly contains retained close homologous non-bovine ruminant βLG mappings after filtering.

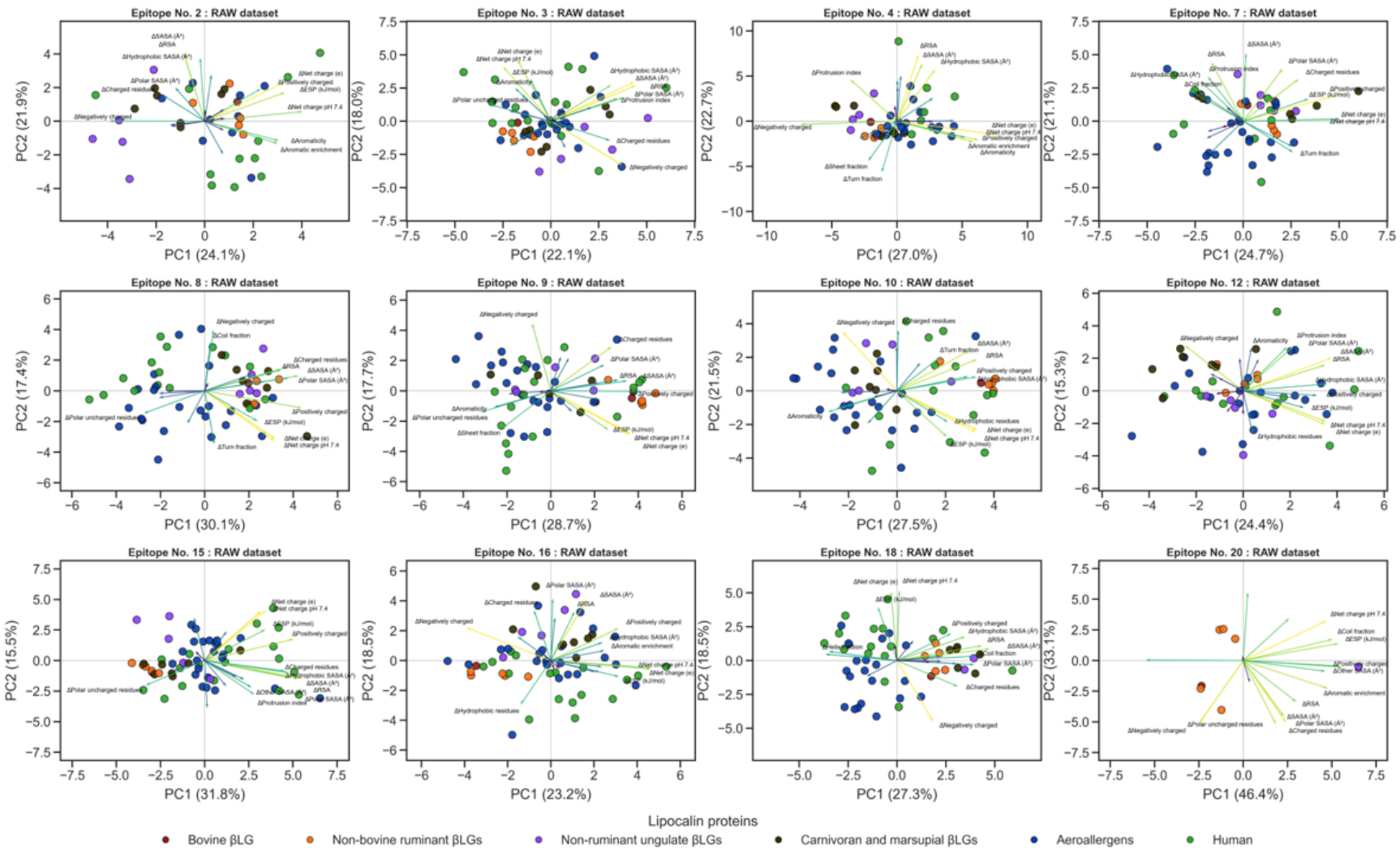

Figure S3. 22. PCA biplots of full Descriptor differences relative to Bos d 5 for MLERs in the RAW dataset.

PC1–PC2 score distribution after applying the structural coverage filter are shown for each mapped linear epitope region (MLER). Points represent query structures and the Bos d 5 reference, colored by lipocalin protein group. Arrows indicate loadings for the full descriptor differences. Epitope 20 mostly contains retained close homologous non-bovine ruminant βLG mappings after filtering.

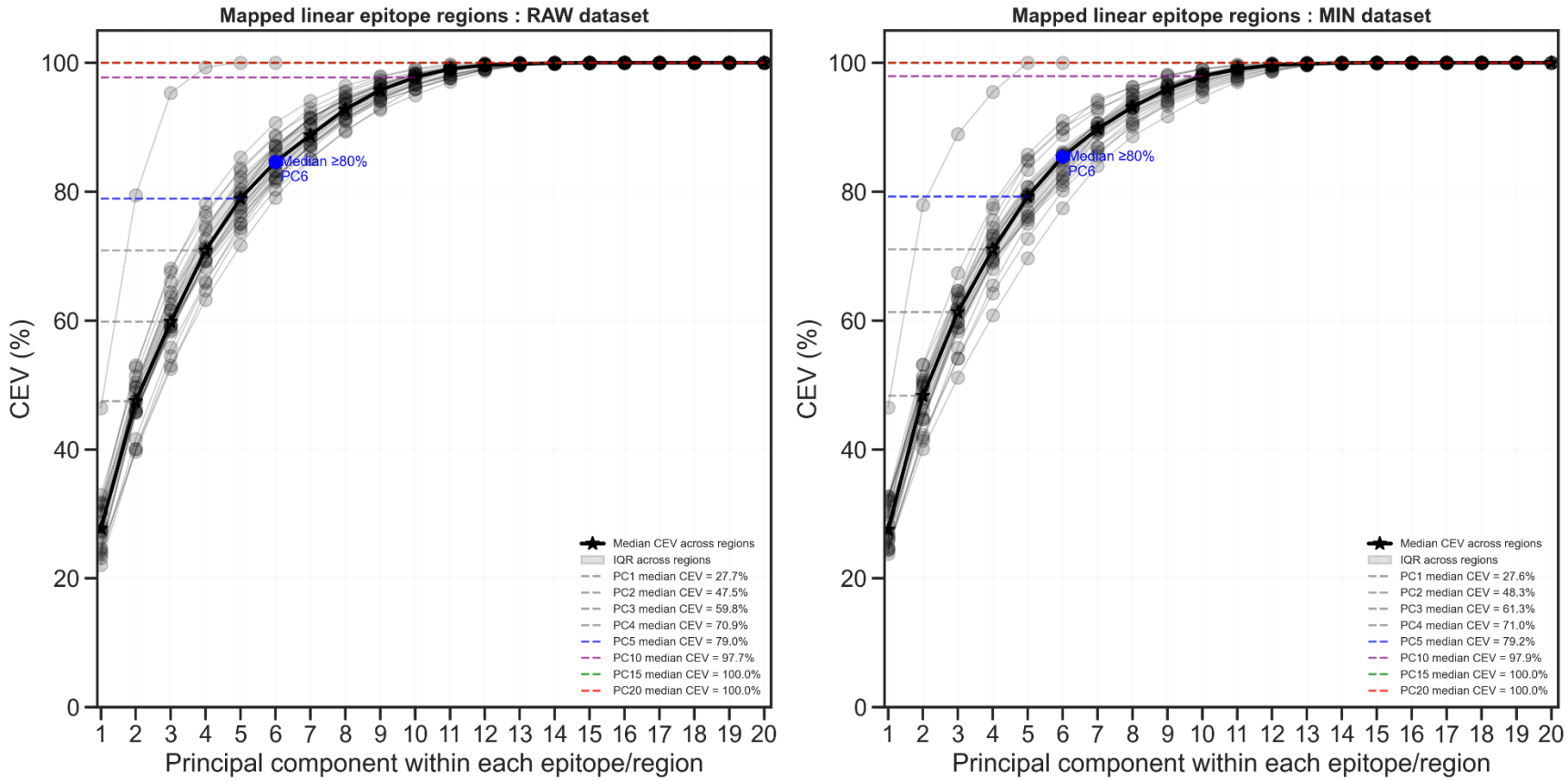

Figure S3. 23. CEV of PCA models fitted to full MLER descriptor set.

Cumulative explained variance (CEV) is shown for PCA models fitted separately to each mapped linear epitope region (MLER) in the RAW and MIN datasets. Grey lines represent individual mapped regions, and the black line indicates the median CEV across regions. In both datasets, PC1–PC3 captured more than 90% of the median variance, while PC1–PC4 captured 100% of the retained descriptor variance.

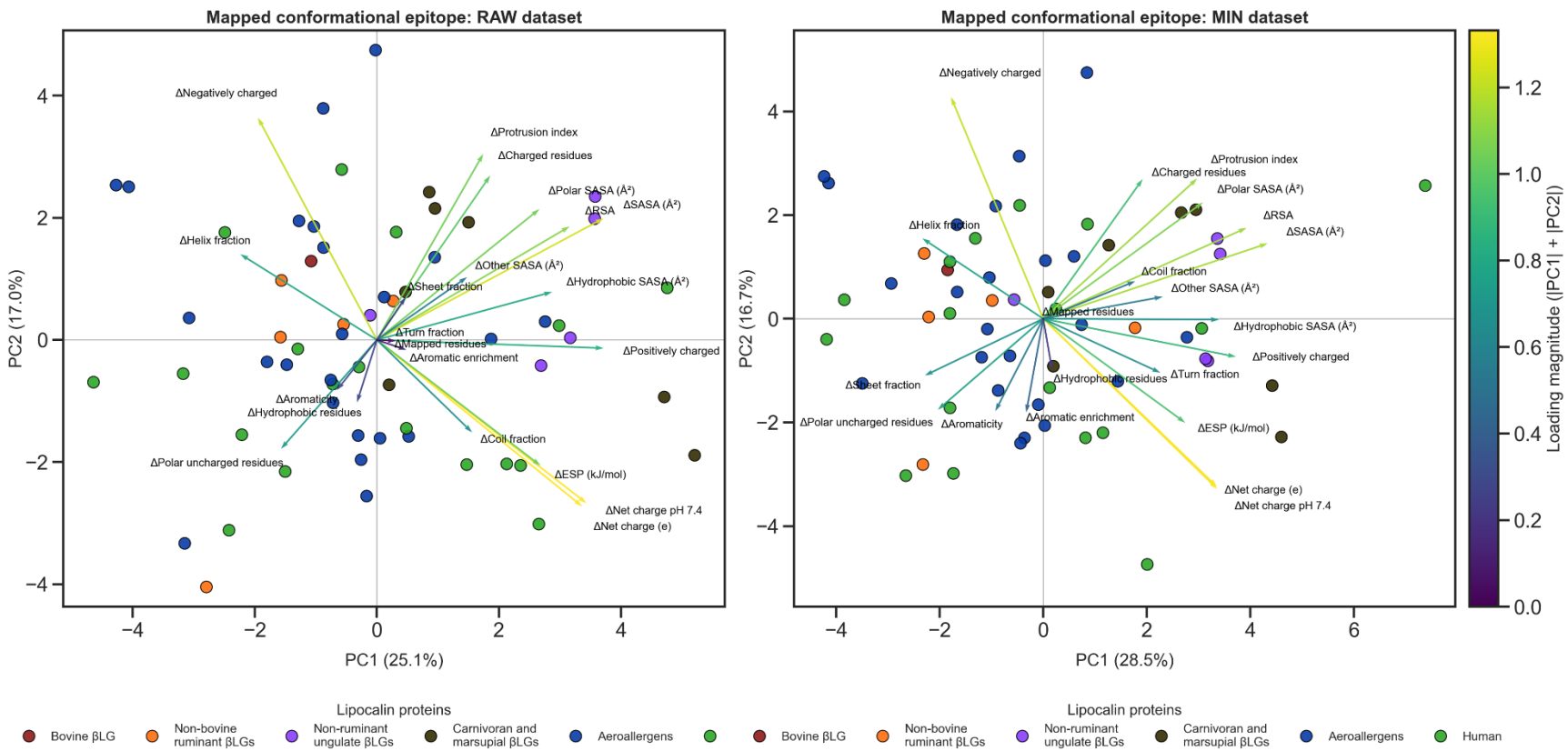

Figure S3. 24. PCA biplots of full Descriptor differences relative to Bos d 5 for the MCER.

PCA biplots are shown for the RAW and MIN mapped conformational epitope region (MCER) datasets. Points represent Bos d 5 and query lipocalin structures, colored by biological group. Arrows indicate loadings for the full Descriptor differences relative to Bos d 5. Axis labels show the percentage of variance explained by PC1 and PC2.

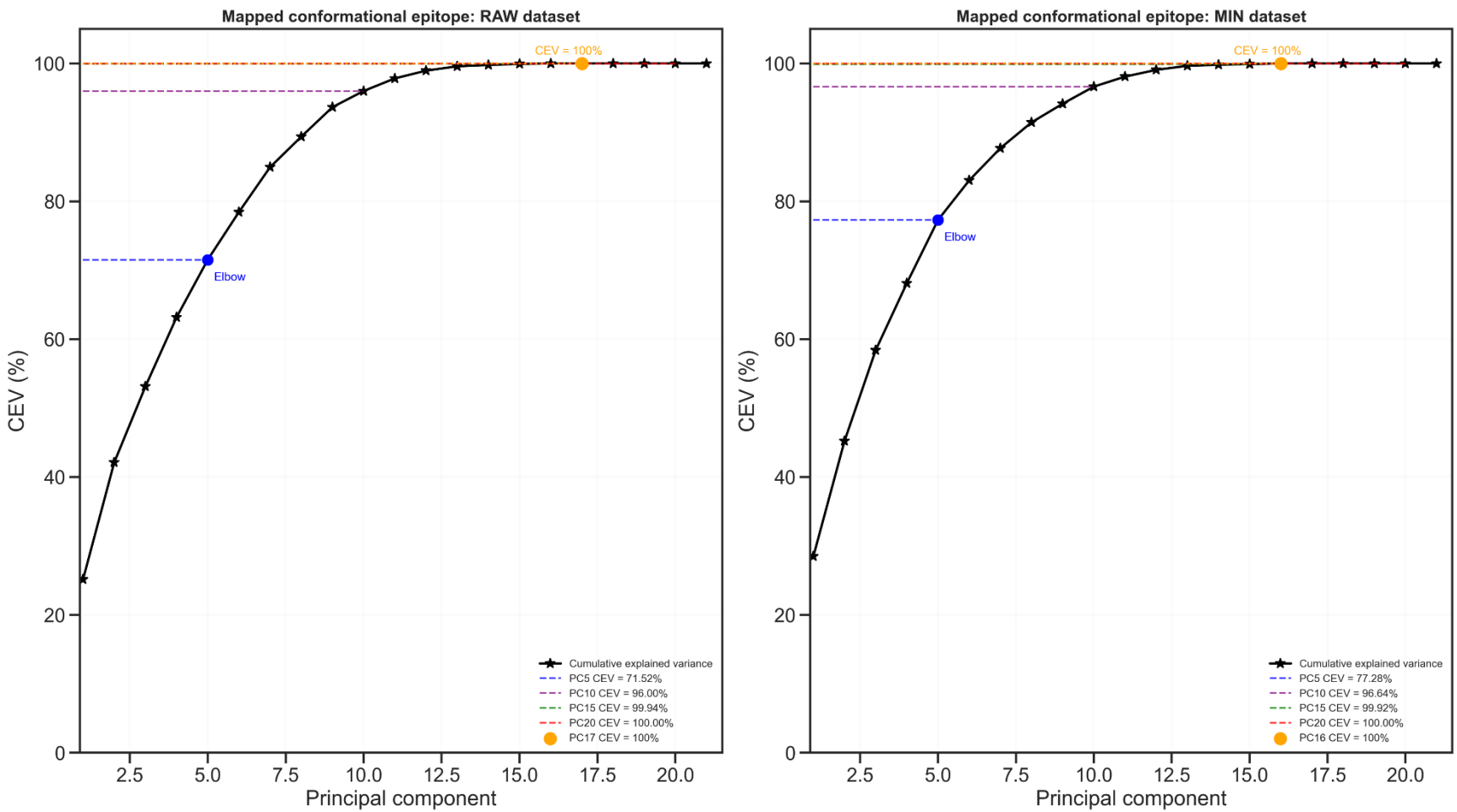

Figure S3. 25. Cumulative explained variance of PCA models fitted to MCER descriptors.

Cumulative explained variance is shown for PCA models fitted to the full mapped conformational epitope region (MCER) descriptor set in the RAW and MIN datasets.

Nearest-neighbor ranking based on Euclidean distance to Bos d 5 in PCA space

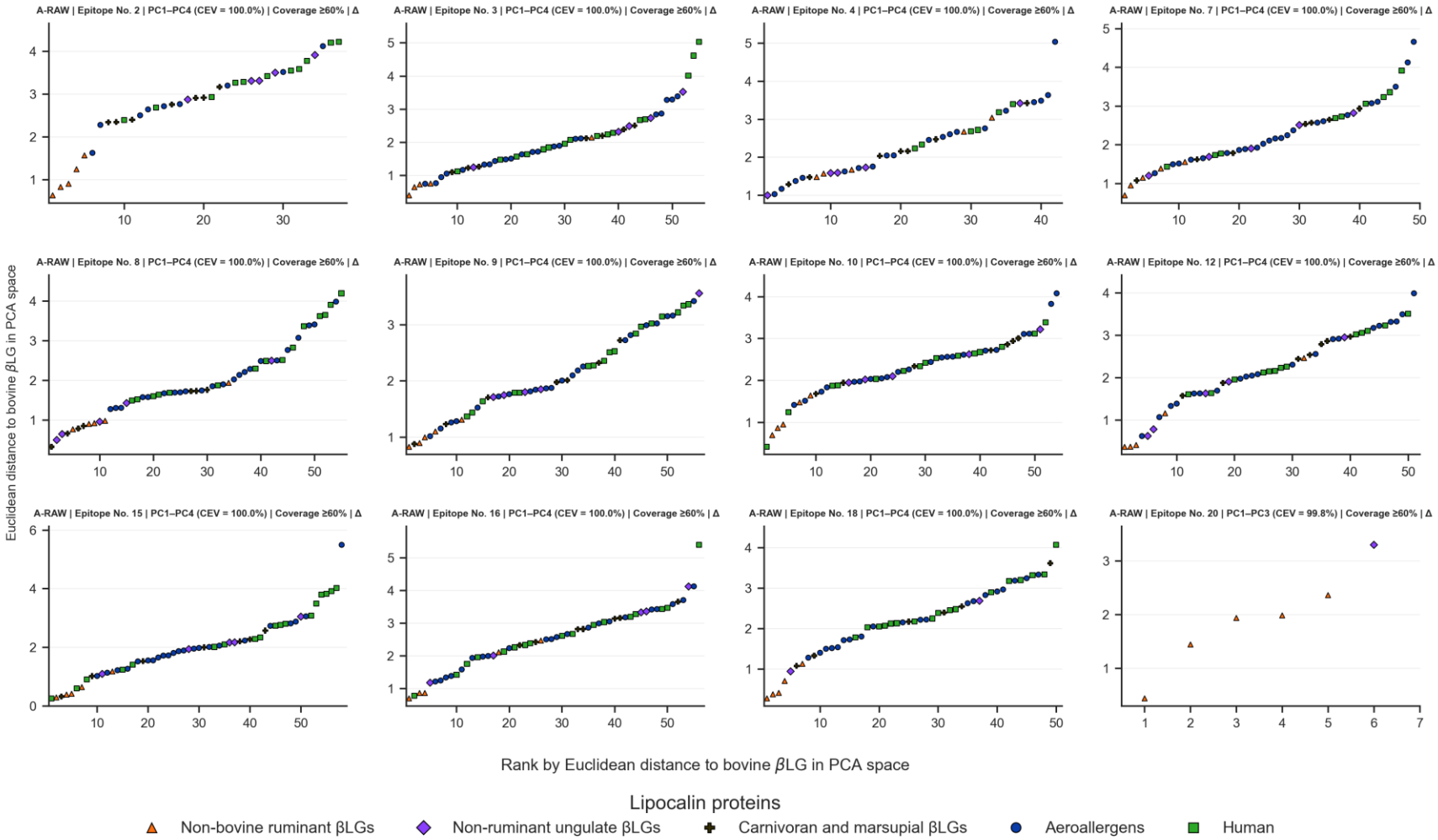

Figure S3. 26. RAW nearest-neighbor rankings for MLERs.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the matched Bos d 5 reference in retained PCA space for each mapped linear epitope region (MLER) under the RAW protocol. Points are colored by biological group. Lower ranks indicate greater similarity to the Bos d 5 reference region.

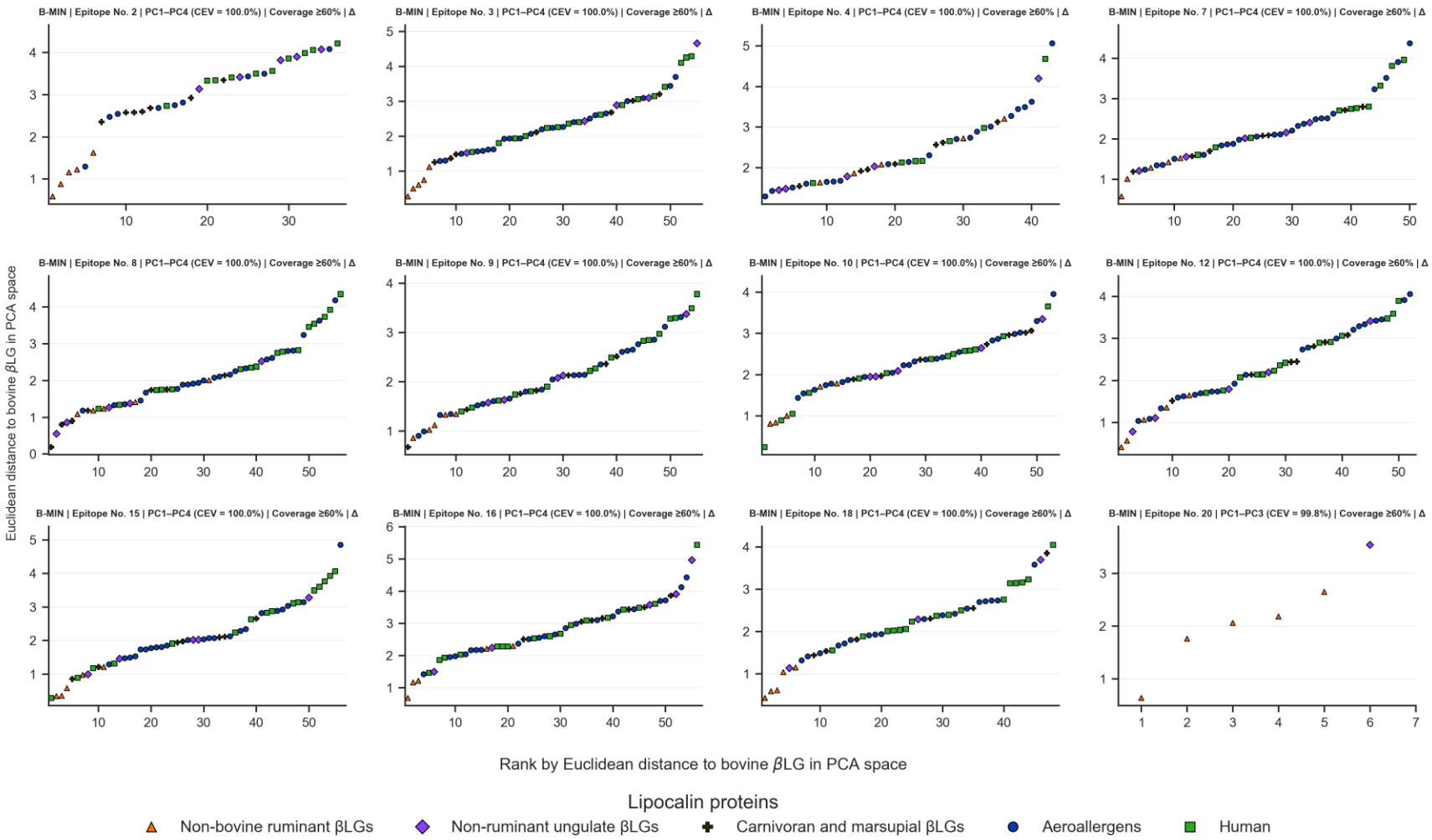

Figure S3. 27. MIN nearest-neighbor rankings for MLERs.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the matched Bos d 5 reference in retained PCA space for each mapped linear epitope region (MLER) under the MIN protocol after structural coverage filtering. Points are colored by biological group. Lower ranks indicate greater similarity to the Bos d 5 reference region.

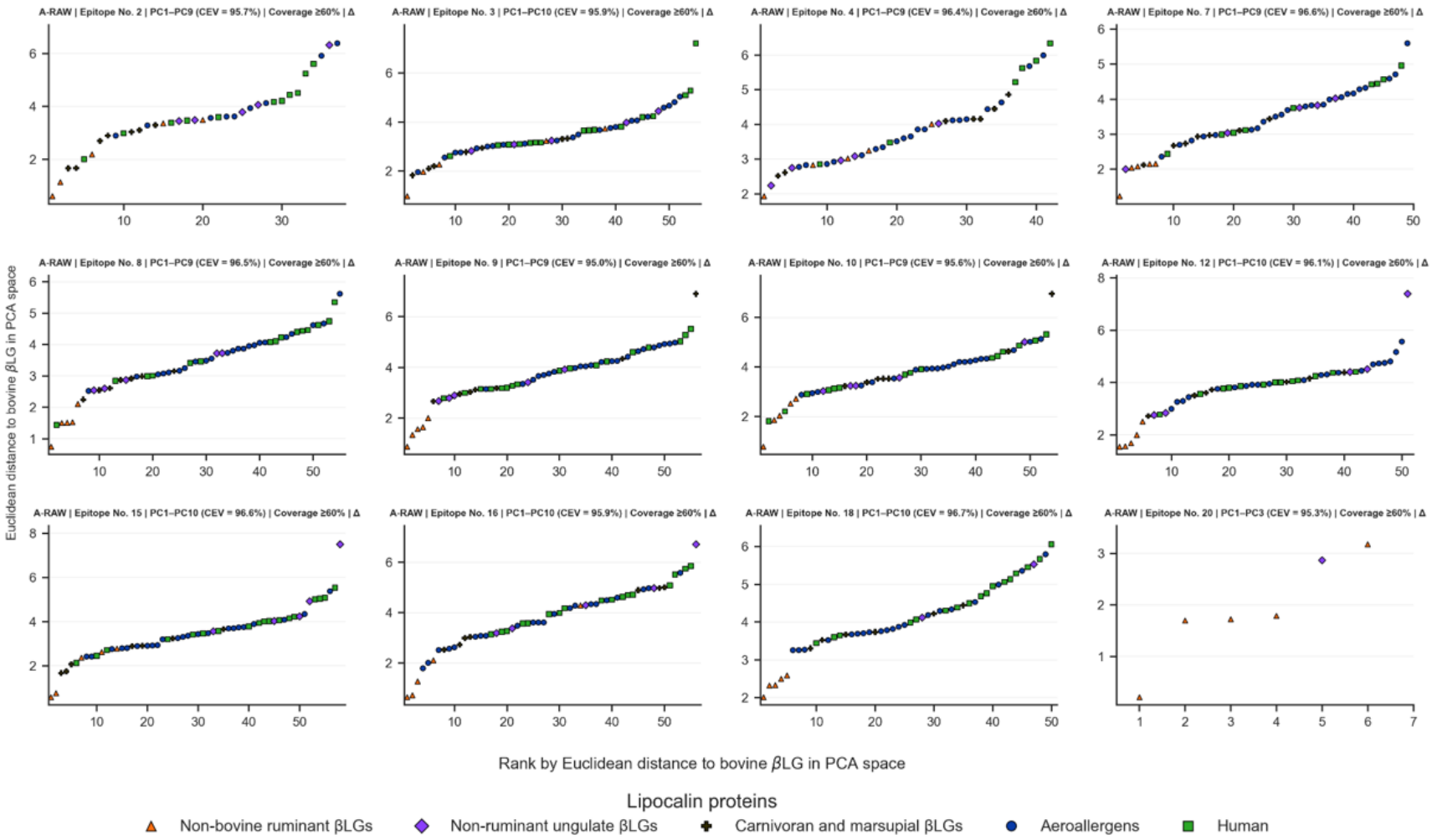

Figure S3. 28. RAW nearest-neighbor rankings for MLERs.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the matched Bos d 5 reference in full PCA space for each mapped linear epitope region (MLER) under the RAW protocol. Points are colored by biological group. Lower ranks indicate greater similarity to the Bos d 5 reference region.

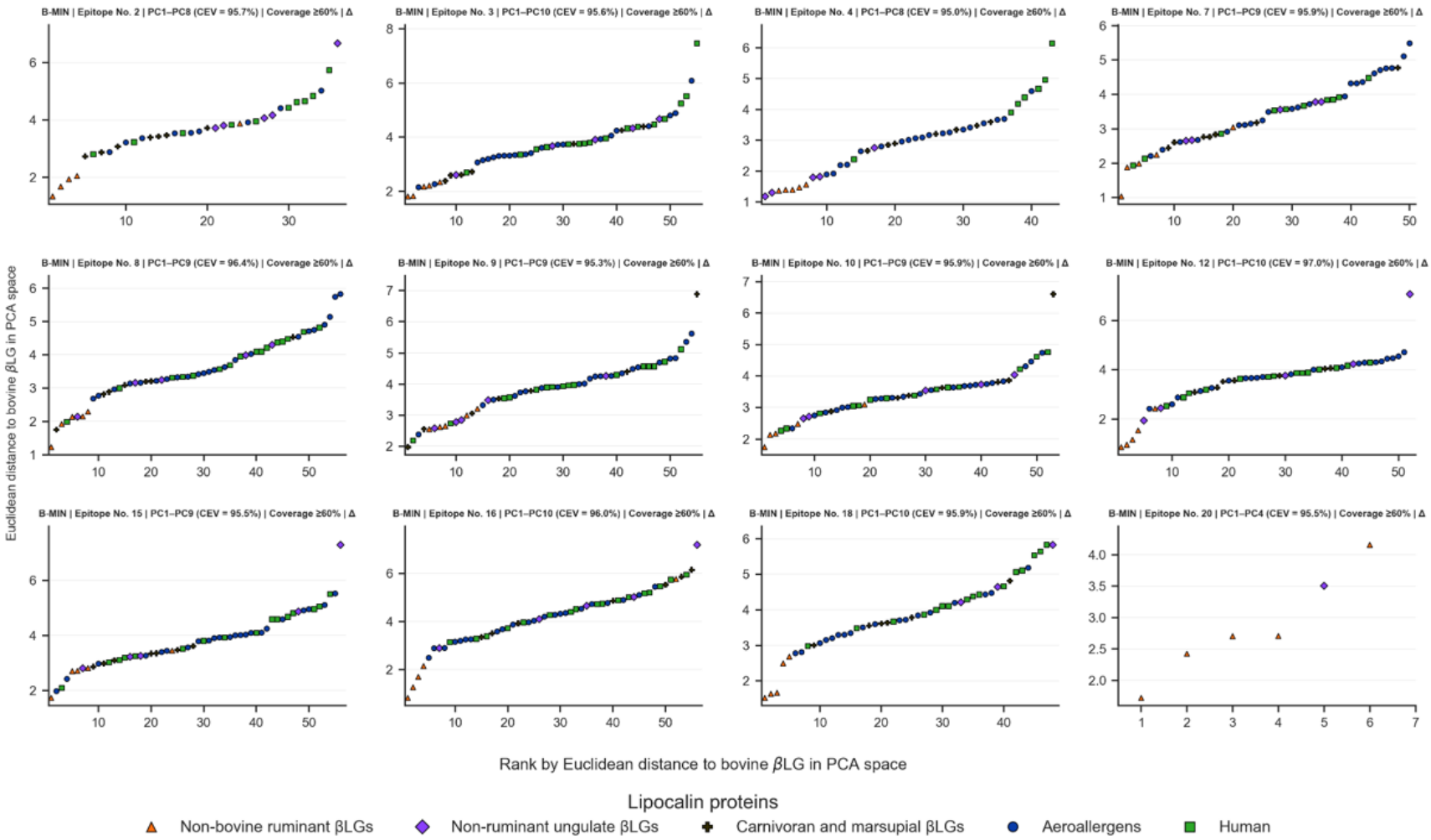

Figure S3. 29. MIN nearest-neighbor rankings for MLERs.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the matched Bos d 5 reference in full PCA space for each mapped linear epitope region (MLER) under the MIN protocol. Points are colored by biological group. Lower ranks indicate greater similarity to the Bos d 5 reference region.

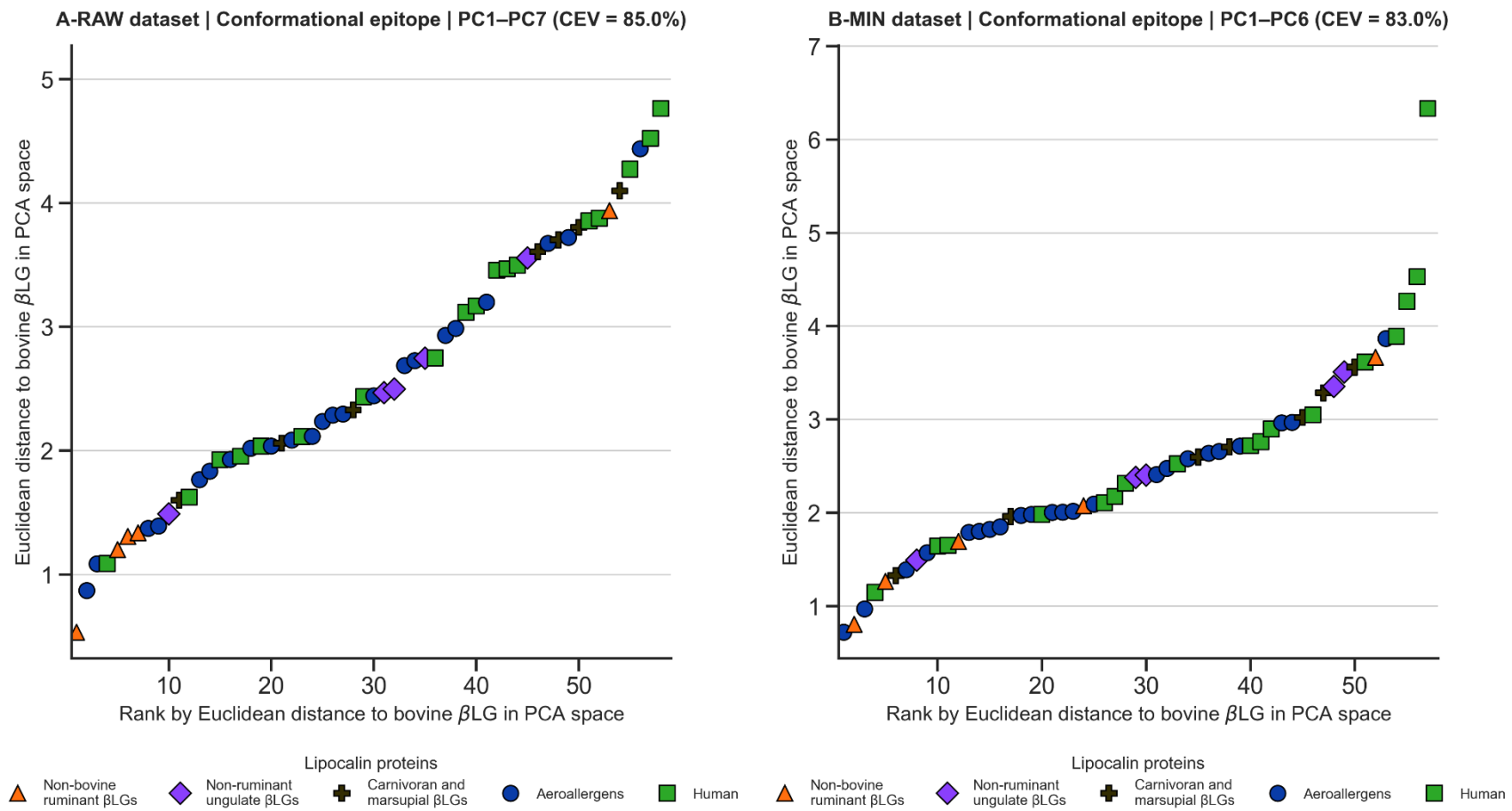

Figure S3. 30. Nearest-neighbor rankings for the MCER.

Rankings for mapped conformational epitope region (MCER). Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in full PCA space for the RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater similarity to Bos d 5.

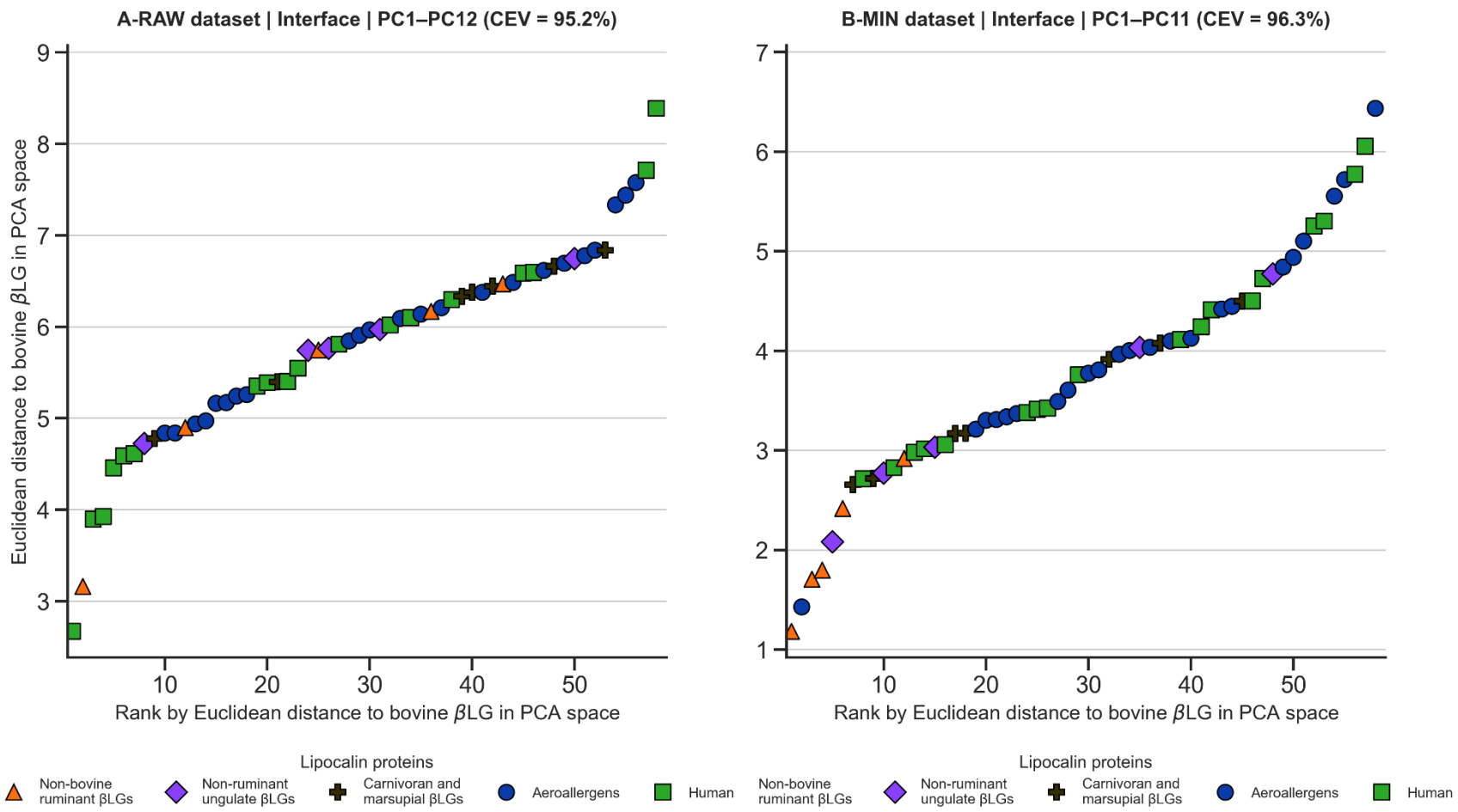

Figure S3. 31. Nearest-neighbor rankings for interface descriptor set.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in full PCA space. Rankings are shown for interface descriptor set under RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater descriptor-space similarity to Bos d 5.

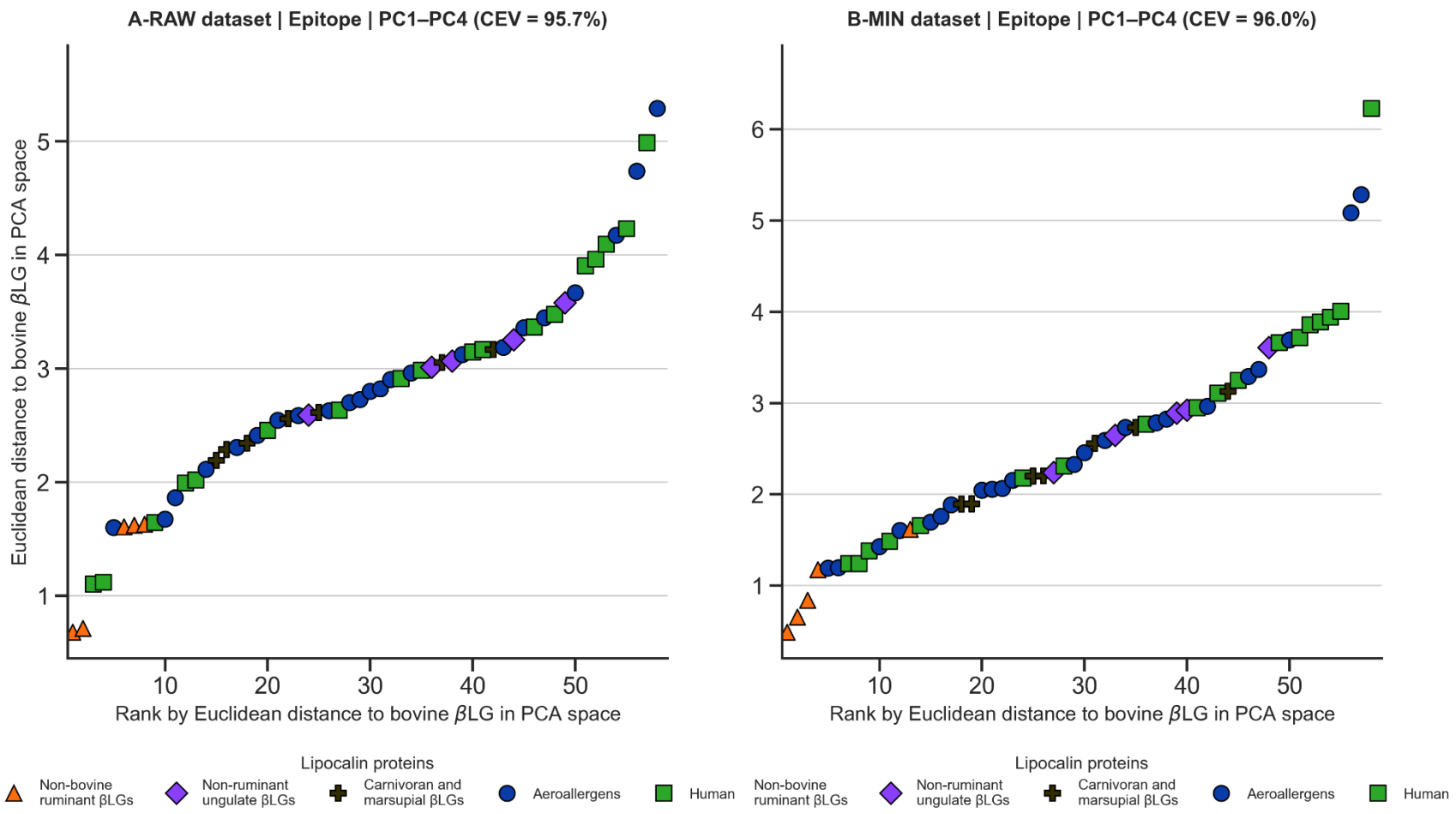

Figure S3. 32. Nearest-neighbor rankings for epitope descriptor set.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in full PCA space. Rankings are shown for epitope descriptor set under RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater descriptor-space similarity to Bos d 5.

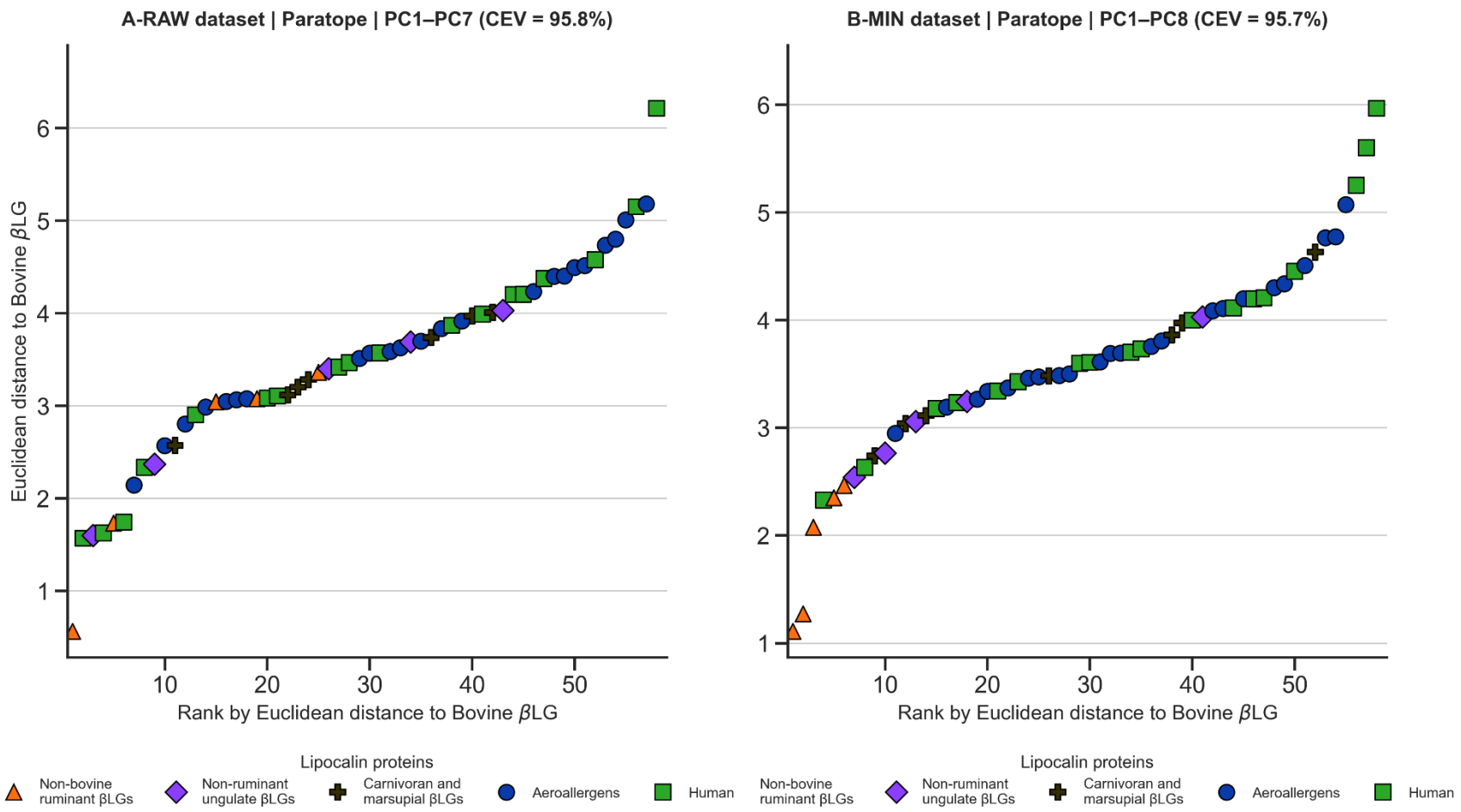

Figure S3. 33. Nearest-neighbor rankings for paratope descriptor set.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in full PCA space. Rankings are shown for paratope descriptor set under RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater descriptor-space similarity to Bos d 5.

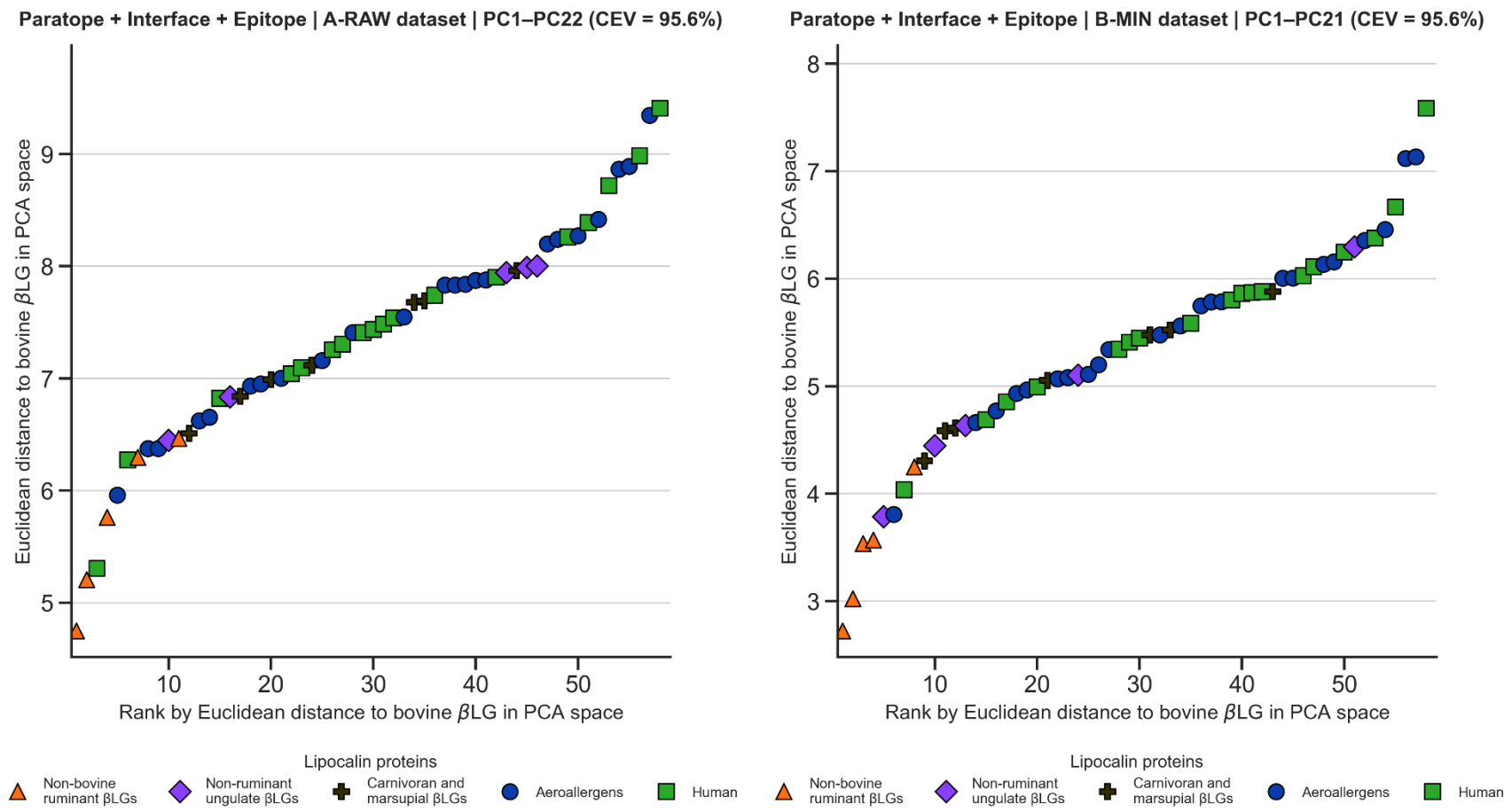

Figure S3. 34. Nearest-neighbor rankings for composite descriptor sets.

Rank-distance plots show query structures ordered by ascending Euclidean distance from the Bos d 5 reference profile in full combined PCA space. Rankings are shown for interface, interface, epitope, paratope, and composite descriptor sets under RAW and MIN protocols. Points are colored by biological group. Lower ranks indicate greater descriptor-space similarity to Bos d 5.

Pairwise group comparison

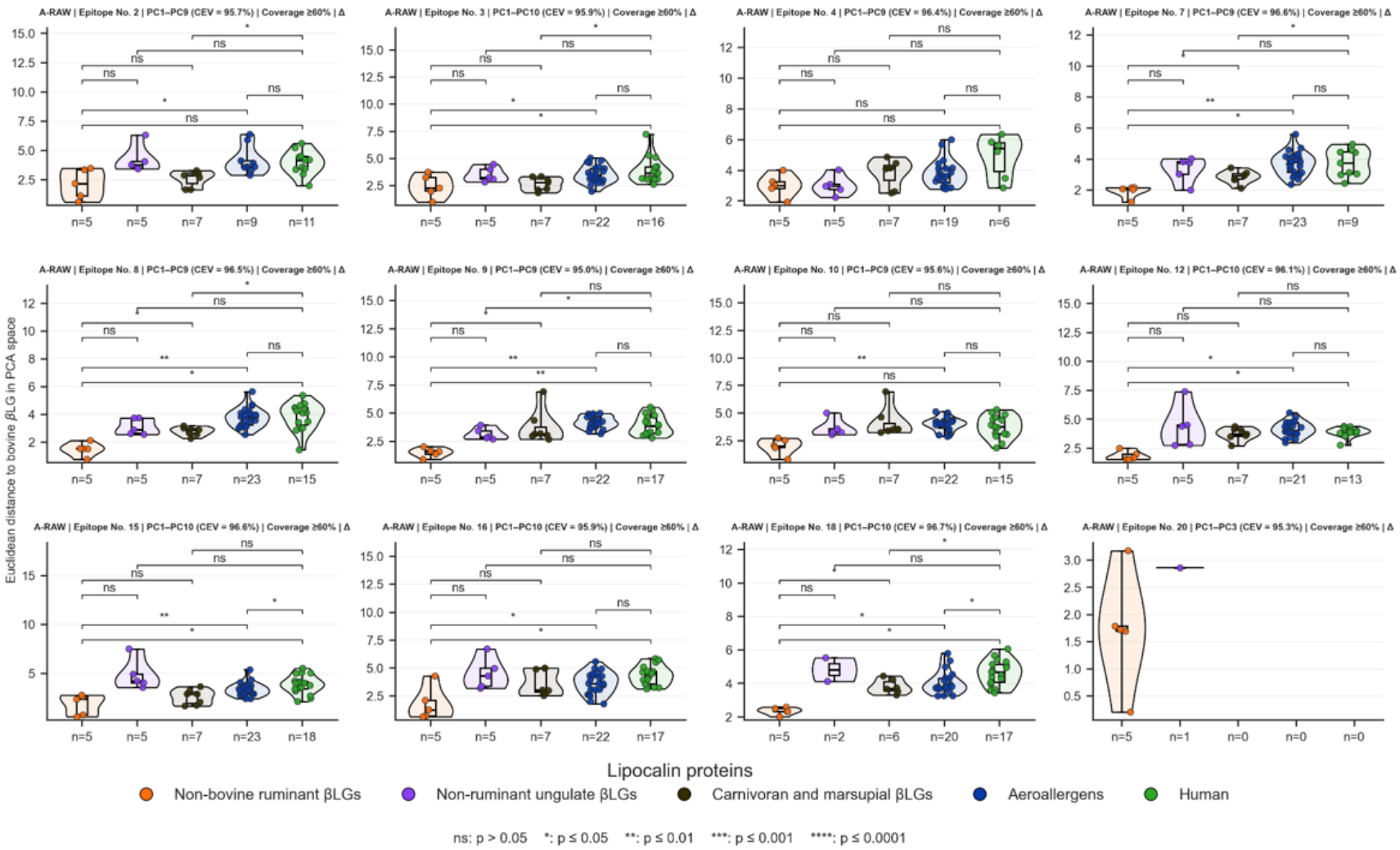

Figure S3. 35. Group differences in RAW MLER Euclidean distances to Bos d 5.

Group comparisons of Bos d 5 mapped linear epitope regions (MLERs). Violin and box plots show Euclidean distances from query structures to the matched Bos d 5 reference profile in full reference-relative descriptor-difference PCA space. Panels correspond to mapped Bos d 5 linear IgE-binding regions. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after multiple-testing correction. Lower distances indicate greater descriptor-space similarity to Bos d 5.

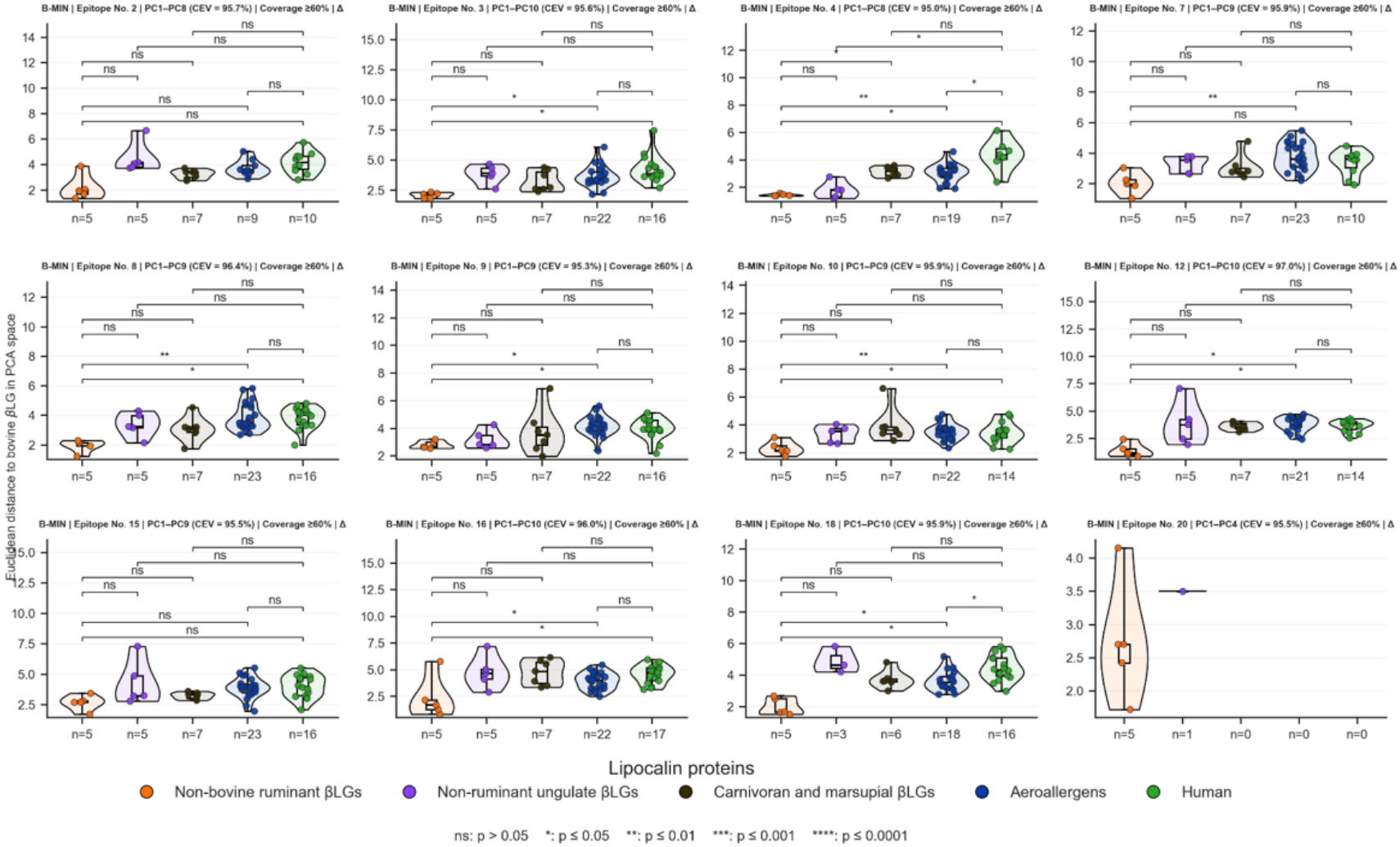

Figure S3. 36. Group differences in MIN MLER Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the matched Bos d 5 reference profile in full reference-relative descriptor-difference PCA space after structural coverage filtering. Panels correspond to mapped Bos d 5 linear IgE-binding regions. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after multiple-testing correction.

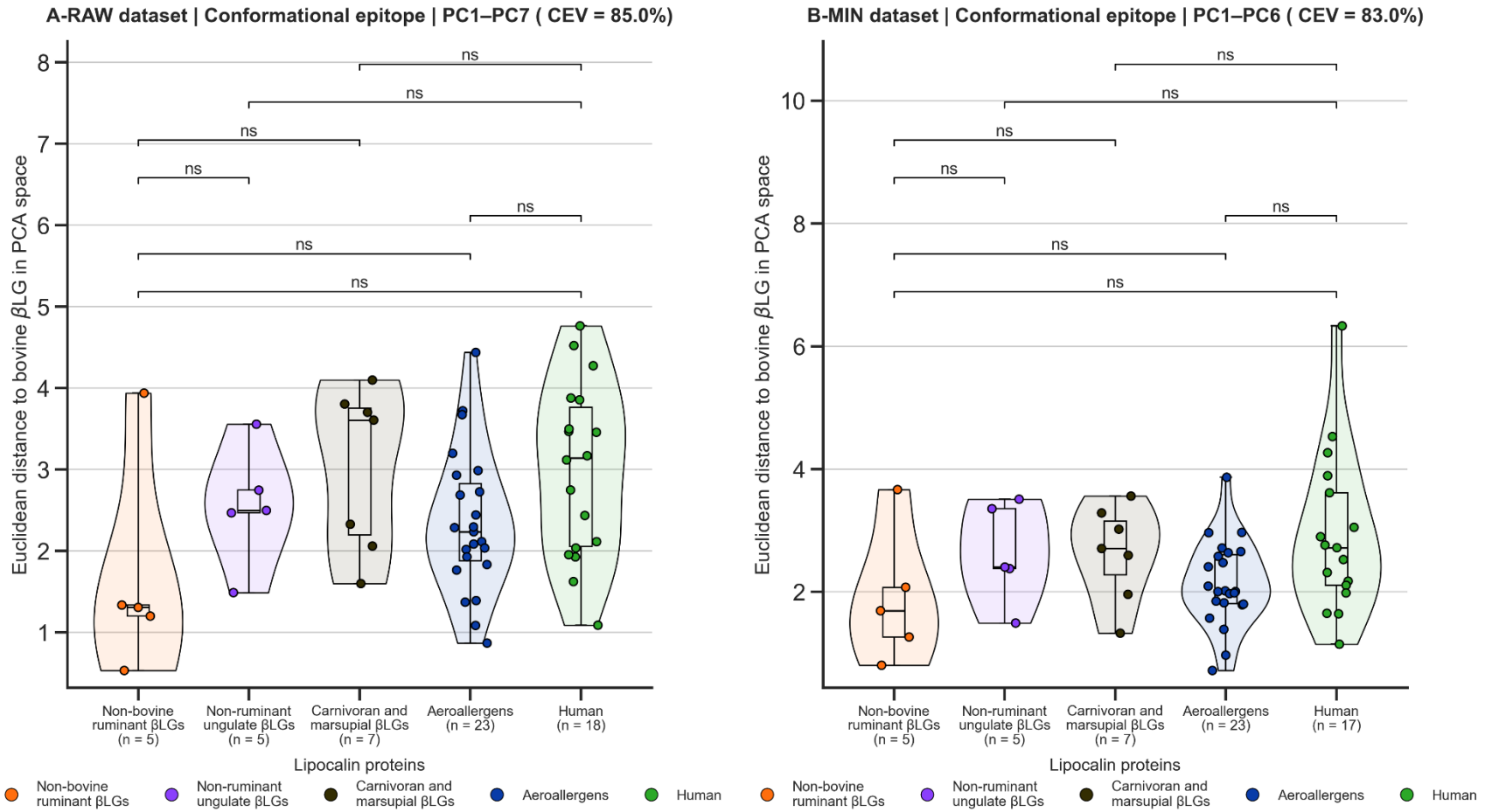

Figure S3. 37. Group differences in MCER Euclidean distances to Bos d 5.

Group comparisons of Bos d 5 PCA distances for the mapped conformational epitope region (MCER). Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in retained reference-relative descriptor-difference PCA space for the RAW and MIN protocols. Points are colored by biological group. Brackets indicate planned pairwise permutation-test results after FDR correction. Lower distances indicate greater descriptor-space similarity to Bos d 5

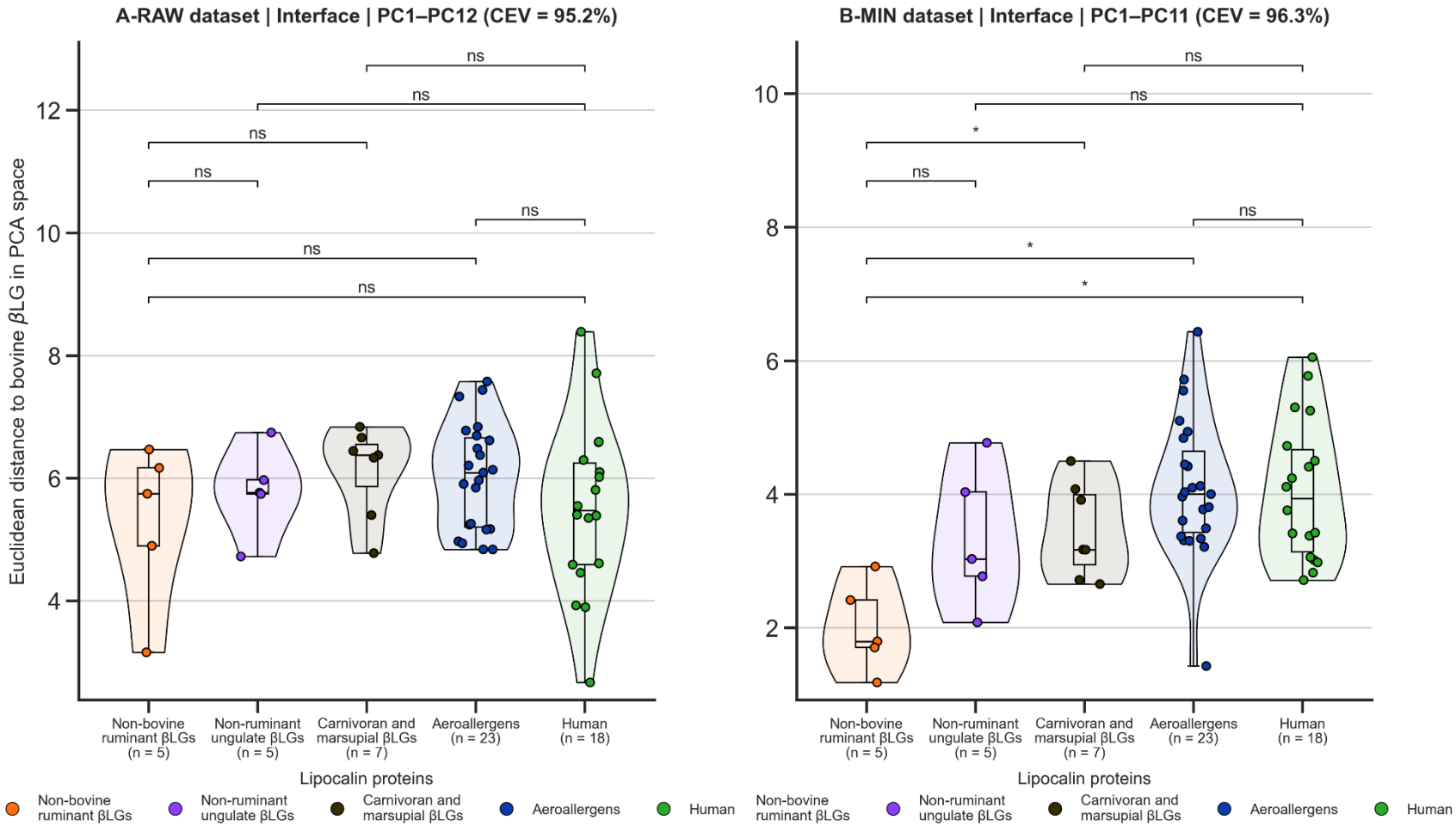

Figure S3. 38. Group differences in interface descriptor Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in full PCA space for interface descriptor sets. Results are shown for RAW and MIN protocols. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after FDR correction.

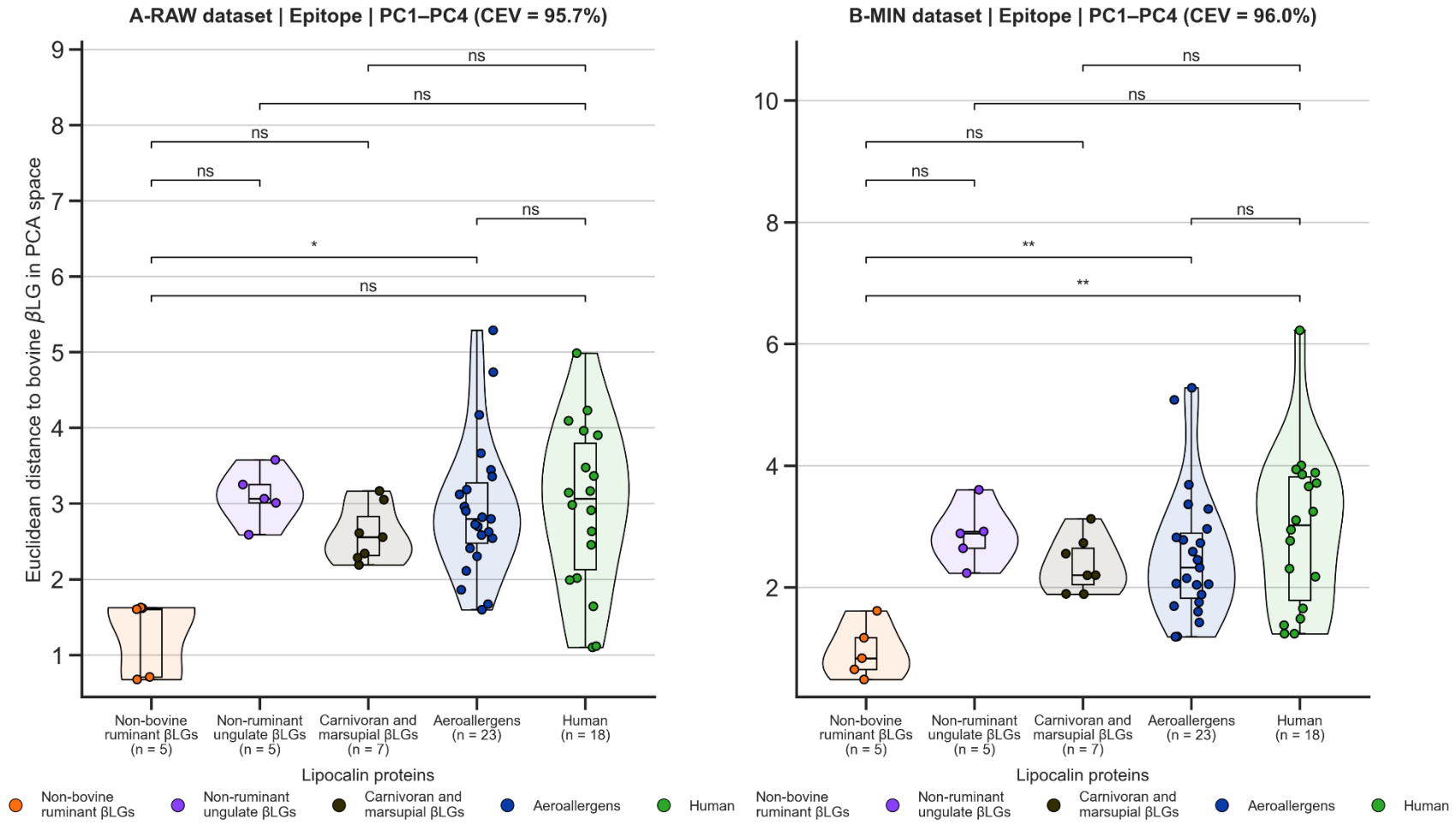

Figure S3. 39. Group differences in epitope descriptor Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in full PCA space for epitope descriptor sets. Results are shown for RAW and MIN protocols. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after FDR correction.

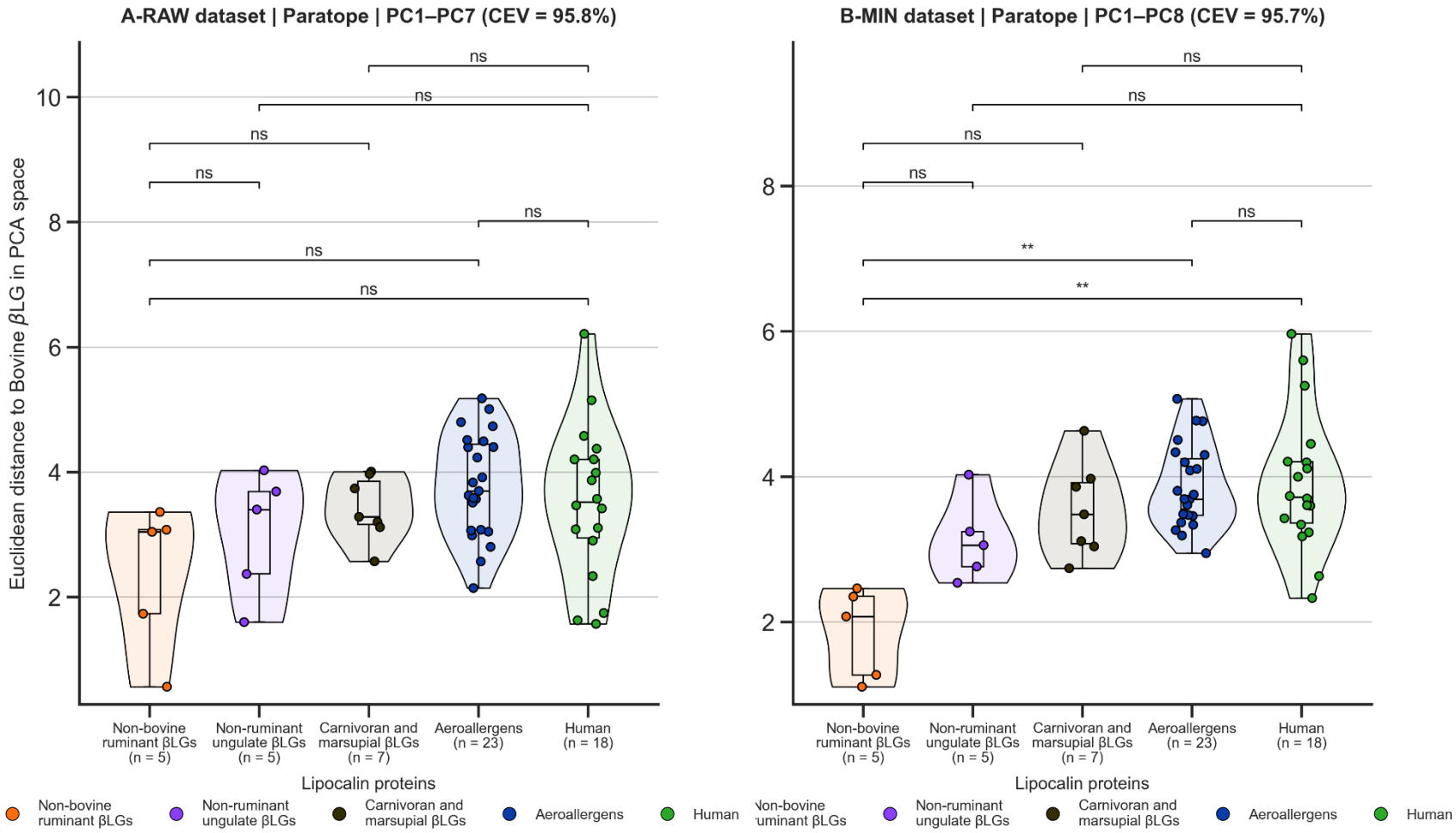

Figure S3. 40. Group differences in paratope descriptor Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in full PCA space for interface descriptor sets. Results are shown for RAW and MIN protocols. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after FDR correction.

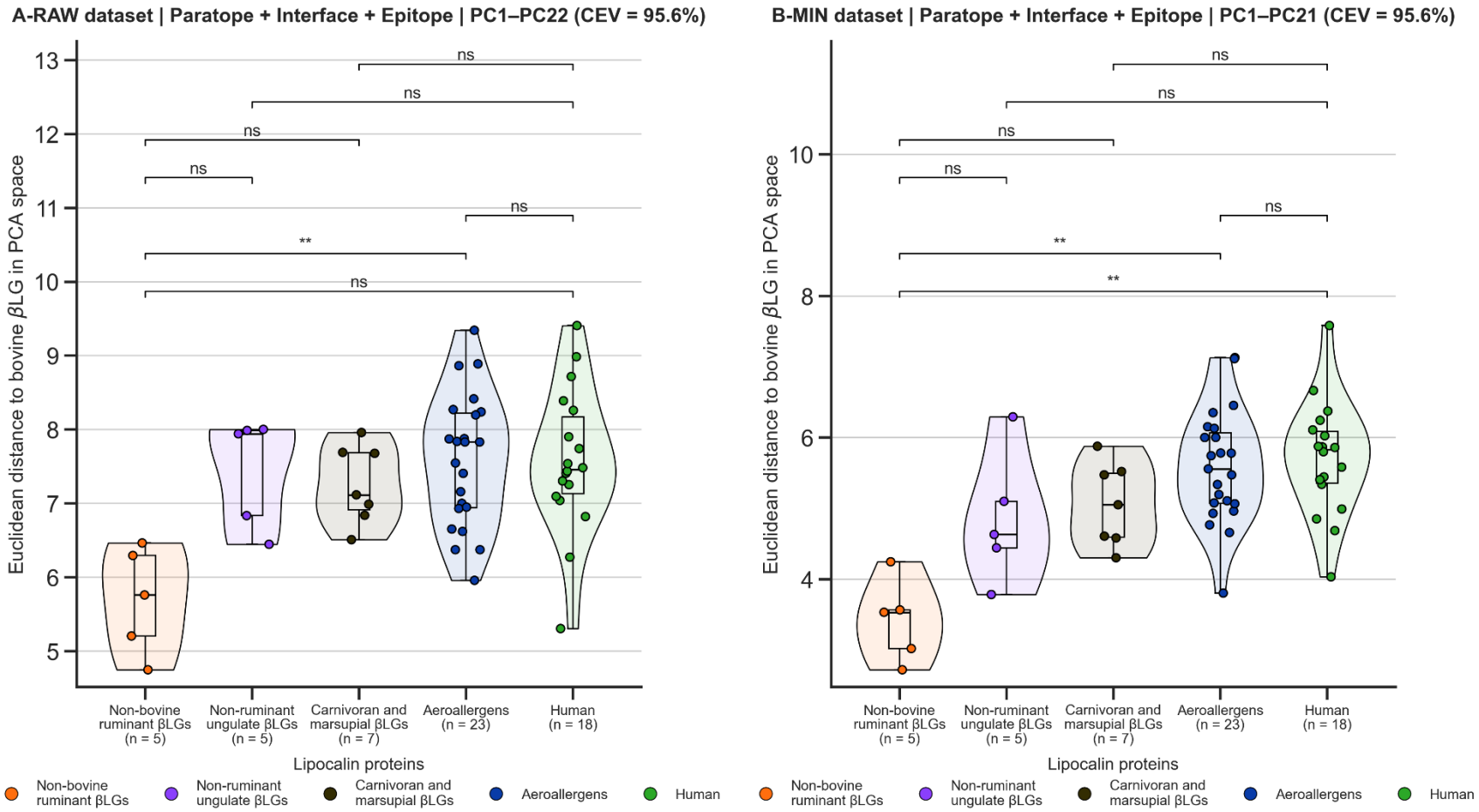

Figure S3. 41. Group differences in composite descriptor Euclidean distances to Bos d 5.

Violin and box plots show Euclidean distances from query structures to the Bos d 5 reference profile in full PCA space for comparative interface, epitope, paratope, and composite descriptor sets. Results are shown for RAW and MIN protocols. Points are colored by biological group, and brackets indicate planned pairwise permutation-test results after FDR correction.