38th EPS
European Peptide Symposium
6 - 11th of September 2026
Vienna, Austria
https://doi.org/10.17952/38EPS.2026.P266

Peptides involved in interaction with heavy metals for environmental remediation, an *in silico* approach.

X. C. Pulido[1], H. F. Cortes Hernandez[1] & C. Jaramillo-Páez[1]

[1] Grupo de Investigación en Química Aplicada a Procesos Ecológicos, Departamento de Química, Facultad de Ciencias, Universidad del Tolima, Ibagué, Colombia

xpulido@ut.edu.co

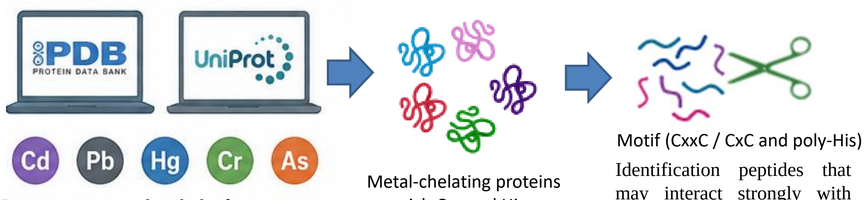
1. INTRODUCTION



Heavy metal (HM) pollution is an increasingly critical global concern with significant environmental and public-health implications. HM are non-biodegradable, accumulate along trophic chains, and remain in soils and water bodies for decades.



***In silico* approach for environmental remediation**



Proteins associated with the five target metals

Metal-chelating proteins rich Cys and His

Motif (CxxC / CxC and poly-His)

Identification peptides that may interact strongly with HM

The aim was to perform an *in silico* analysis of proteins associated with Cd, Pb, Hg, Cr and As deposited in the PDB (through its wwPDB portals) and in UniProt, and to derive, through controlled *in silico* digestion, a non-redundant catalogue of peptides with putative metal-binding properties, characterized by mass and absorption coefficient.

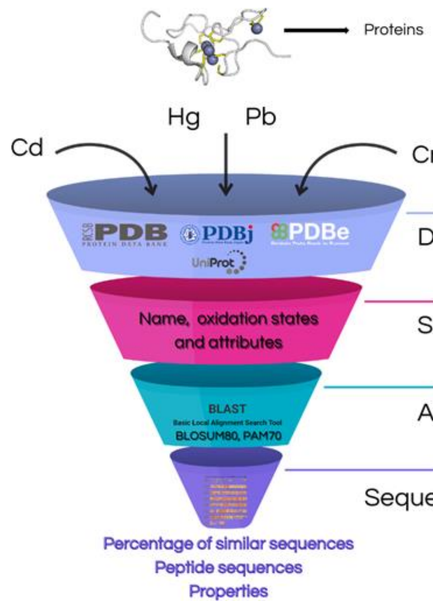


FIGURE 1. Flowchart of the *in silico* process: metal-based database queries feed structural and sequence retrieval, local alignment, and the selection of non-redundant sequences and candidate peptides.

2. MATERIALS AND METHODS

1. **Database search and data collection/** Worldwide Protein Data Bank (wwPDB), consulted through its PDB, PDBe and PDBj portals, for three-dimensional structures[8], and UniProtKB
2. **Sequence similarity and redundancy reduction/** local alignment using the BLAST algorithm[10], with BLOSUM and PAM substitution matrices
3. ***In silico* digestion/** PeptideCutter (ExPASy)[13] and the Protein Sequence Digestion Calculator
4. **Peptide analysis /** PeptideCutter (ExPASy)[13] and the Protein Sequence Digestion Calculator
5. **Mass and absorption coefficient determination/** Prot pi | Peptide Tool
6. **Motif and binding-site identification /** cysteine- and histidine-rich patterns (CxxC, CxC, HxH and poly-His) Si se encontraron péptidos con HxH y poli-His
7. **Gene Ontology (GO) Analysis /** UniProt database
8. **Statistical analysis**

3. RESULTS AND DISCUSSION

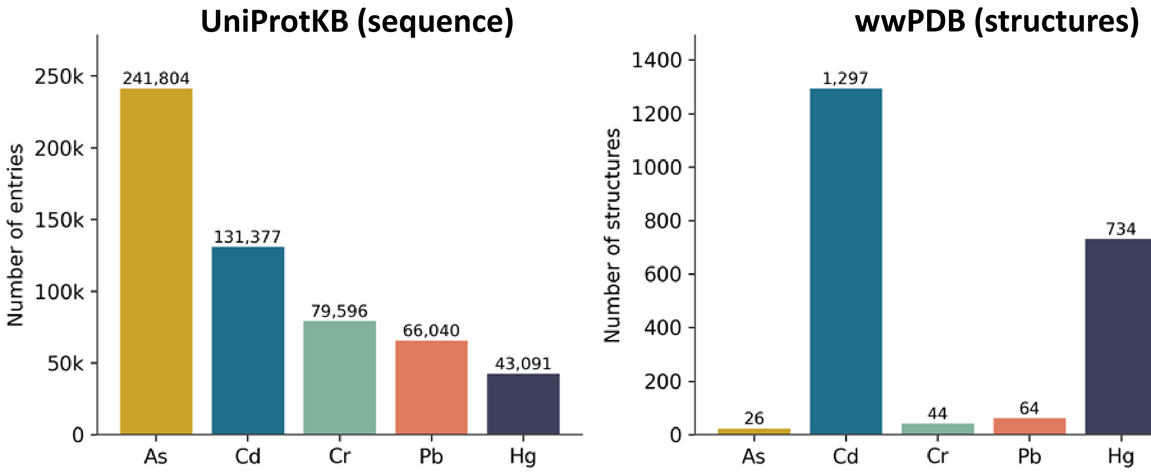


FIGURE 2. Number of macrostructures retrieved per metal in the sequence resource (UniProtKB, left, all entries) and the structural archive (wwPDB, right). wwPDB structures are counted once against the shared archive.

The raw retrieval returned 561,908 UniProtKB entries and 2,165 wwPDB structures across the five metals (Figure 2). In UniProt, the abundance order was As > Cd > Cr > Pb > Hg, with arsenic alone accounting for 241,804 sequence entries, reflecting the broad annotation of arsenate/arsenite resistance and metabolism systems in prokaryotic genomes. In the structural archive, the order was reversed toward cadmium (1,297 structures), consistent with the extensive crystallographic characterization of Cd-substituted metalloproteins. Reporting sequence and structural counts separately, and counting the shared wwPDB archive only once, avoids the portal-by-portal overcounting that would otherwise inflate the totals several-fold.

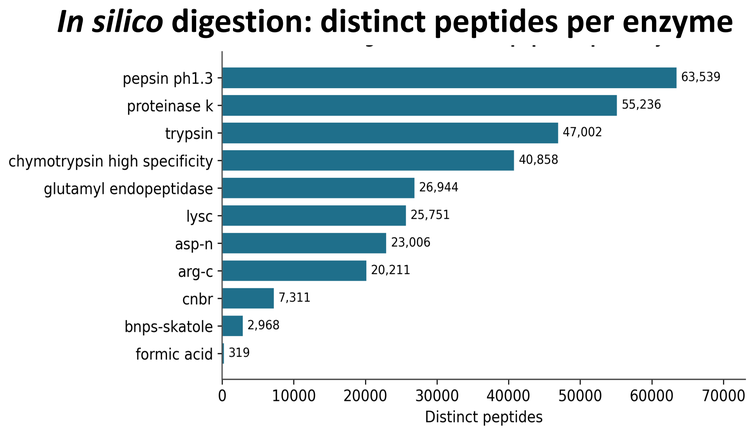


FIGURE 3. Distinct peptides produced by each enzyme or chemical agent during *in silico* digestion of the non-redundant protein set.

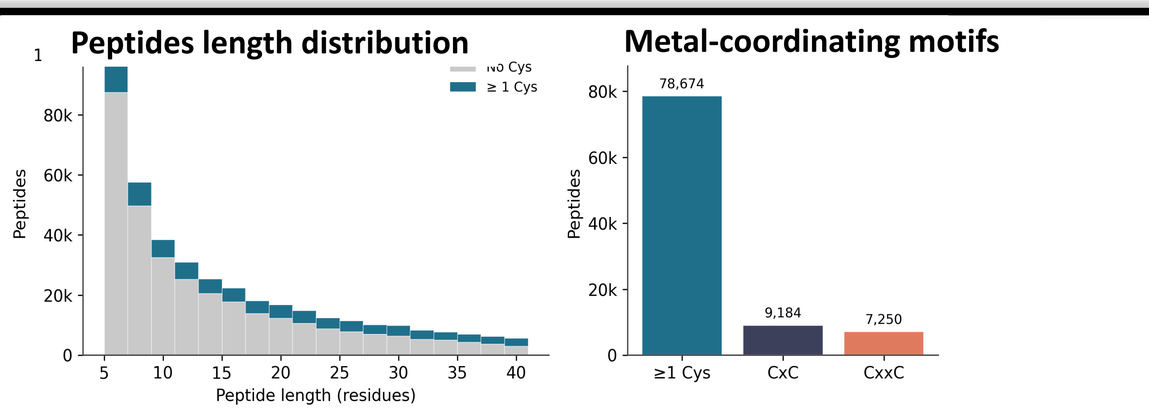


FIGURE 4. Peptide length distribution stratified by cysteine content (left) and abundance of metal-coordinating cysteine motifs (right).

Table 1. Top ten candidate peptides ranked by metal-coordination potential. Cysteine residues (thiolate donors) are highlighted in red; masses are monoisotopic.

Peptide (Cys highlighted)	UniProt	Len	Cys	CxxC / CxC	Mass (Da)	pI
LCCCFCCYETCEHCLECFCCCCKKDN	A3A4E0	26	12	6 / 5	3047.98	4.83
GCLYACLFTLCCCFCCYETCECCLECLCCC	Q5VRD7	30	14	5 / 5	3363.10	4.05
SCKCSNCQCACCKKSCCSCCPSGCSKCA	O93571	28	12	5 / 4	2826.94	8.50
CCYETCECCLDCLCCCCN	Q5VSB5	18	10	4 / 4	2025.54	4.05
CTNCSCKCCKKSCCSCCPSGCSKCA	Q92044	25	11	5 / 3	2538.85	8.56
PNCSACCAGSCCSPDCCSCCKPNCSCCK	Q67UU9	28	12	4 / 2	2778.84	7.83
GSCKCKNCRCQSCRKSCCSCCPASCSNCA	P15787	29	11	5 / 3	3022.08	8.74
DSCKCEGCKCTSCKKSCCSCCPAECEKCAK	P25713	30	11	5 / 3	3167.16	8.07
DKCKCKGCKCTNCKKSCCSCCPAGCEKCAK	P28184	30	11	5 / 3	3162.27	8.84
GDSCSCAGSCKCKNCRCQSCRKSCCSCCPA	P15787	30	11	4 / 4	3080.09	8.57

Motif counts (CxxC / CxC) include overlapping occurrences. Source proteins (UniProt accessions) correspond to metallothioneins and metallothionein-like sequences.

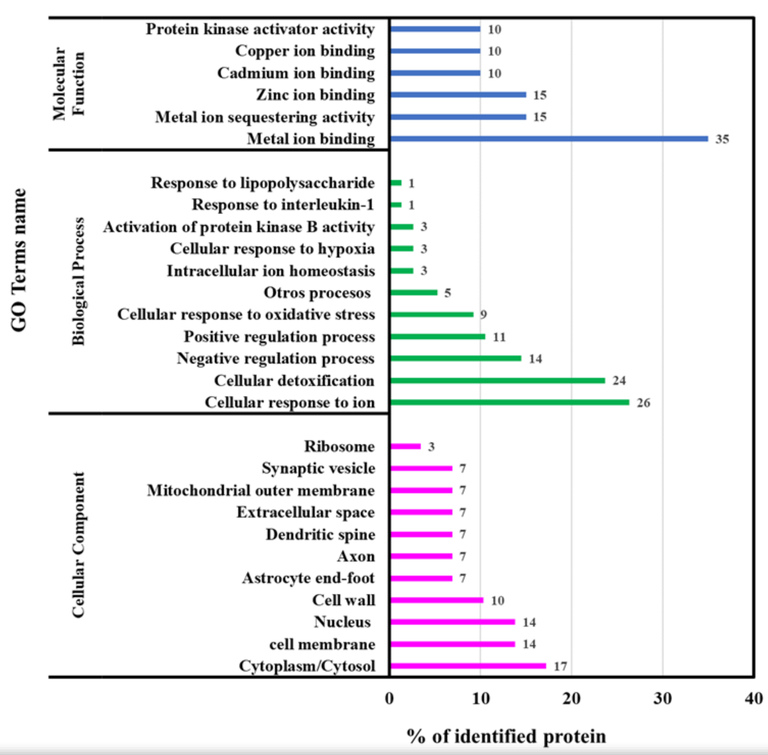


FIGURE 5. Gene Ontology (GO) classification of the nine proteins from which the candidate peptides were selected. This functional analysis divided proteins into three groups: Molecular Function, Biological Process and Cellular Component.

CONCLUSION

An *in silico* pipeline combining database mining, curation, enzymatic fragmentation, and physicochemical characterization reduced a raw retrieval of 561,908 UniProtKB entries to a non-redundant, high-confidence set of 5,156 metal-associated proteins, from which it derived 275,933 distinct candidate peptides using eleven cleavage specificities. Nearly one in five peptides carried a cysteine, and thousands reproduced the CxxC/CxC motifs that coordinate soft divalent ions, providing a well-defined priority pool of Cd-, Pb-, and Hg-binding candidates. Future work will validate the prioritized cysteine- and histidine-rich peptides experimentally and quantify their metal-binding affinity, advancing the design of peptide-based biosorbents for environmental remediation.

ACKNOWLEDGMENTS

We thank to the University of Tolima for its financial support under the research project, code: 470121.

REFERENCE

1. M. Balali-Mood, K. Naseri, Z. Tahergorabi,M. R. Khazdair, & M. Sadeghi. Frontiers in pharmacology, **2021**, 12, 643972. 1-19.
2.The Protein Data Bank (2000) Nucleic Acids Research, **2000**, 28, 235-242.
3. R. Ochoa and K. Deibler, J Pept Sci, **2025**, 31, e3666, 1-9.