Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

David Baker

University of Washington

Google Scholar profile

Generated 2026-05-31 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Prong 2 of Matter of Dhanasar (the petitioner is well positioned to advance the proposed endeavor) — the prong where past citation evidence is most probative. It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

2,371	2,535	4	233
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

92.7% independent of 1,058 classified citing papers

Citation type	Count
Independent	981
Self-citation	36
Co-author	23
Same-institution	18

1,313 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher produced a seminal 2021 publication that has garnered over 6,000 citations, establishing a foundational contribution widely adopted by independent scholars across the field.

The researcher's primary contribution rests on a seminal paper published in 2021, which stands as a cornerstone of their academic output. This work represents a significant intellectual achievement, serving as the central pillar of their research portfolio without reliance on subsequent follow-up publications to define its scope or impact.

Given the absence of follow-up papers, the originality of this contribution appears to lie in its self-contained novelty and immediate resonance within the scientific community. The work likely addressed a critical gap or introduced a transformative concept that required no immediate iterative refinement by the author, suggesting a high degree of initial completeness and conceptual clarity.

The significance of this work is evidenced by its substantial citation count of 6,290, indicating widespread recognition and utility. Furthermore, the fact that 93.9% of citing papers originate from independent researchers underscores the work's broad influence beyond the author's immediate circle, confirming its status as a widely adopted standard or reference point in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 369

CORE PAPER

Untitled

2021 · Science 373 (6557), 871-876, 2021 · 6,290 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences (2024)	EMBL-EBI, Google DeepMind, Seoul National University	South Korea, United Kingdom	—
2	InterPro: the protein sequence classification resource in 2025	Biobyte Solutions GmbH, EMBL, EMBL-EBI	France, Germany, Italy	—
3	Generative Flows on Discrete State-Spaces: Enabling Multimodal Flows with Applications to Protein Co-Design	Massachusetts Institute of Technology, University of Oxford	United Kingdom, United States	—
4	Towards an AI co-scientist	—	—	—
5	Empowering biomedical discovery with AI agents	Harvard Medical School, Harvard University, Massachusetts Institute of Technology	United States	—
6	Artificial Intelligence for Science in Quantum, Atomistic, and Continuum Systems	California Institute of Technology, Cornell University, Harvard Medical School	Canada, Germany, Netherlands	—
7	The molecular basis for cellular function of intrinsically disordered protein regions (2024)	University of Copenhagen, Washington University School of Medicine	Denmark, United States	—
8	Scientific discovery in the age of artificial intelligence	BioMap, Boehringer Ingelheim, Broad Institute of MIT and Harvard	Canada, China, Germany	—
9	Accurate structure prediction of biomolecular interactions with AlphaFold 3 (2024)	Google DeepMind, Google DeepMind / Isomorphic Labs, Isomorphic Labs	United Kingdom, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
10	Guiding questions to avoid data leakage in biological machine learning applications	Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Technical University of Munich	Germany	—
11	Evolutionary-scale prediction of atomic-level protein structure with a language model	Massachusetts Institute of Technology, Meta, Meta AI	United States	—
12	An atlas of substrate specificities for the human serine/threonine kinome	Cell Signaling Technology, Columbia University Irving Medical Center, Humboldt-Universität zu Berlin	France, Germany, Japan	—
13	The HADDOCK2.4 web server for integrative modeling of biomolecular complexes	Utrecht University	Netherlands	—
14	Game changers in science and technology - now and beyond	Aché Laboratórios Farmacêuticos, Astex Pharmaceuticals, Bayer AG	Australia, Austria, Brazil	—
15	From nature to industry: Harnessing enzymes for biocatalysis	Codexis Incorporated, Greifswald University, Institute of Biochemistry, Greifswald University	Germany, Switzerland, United States	—
16	I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction	University of Michigan, Zhejiang University of Technology	China, United States	—
17	Applications of synthetic biology in medical and pharmaceutical fields	PhaBuilder Biotech Co. Ltd., Tsinghua University	China	—
18	The impact of AlphaFold Protein Structure Database on the fields of life sciences	European Bioinformatics Institute	United Kingdom	—
19	Computational approaches streamlining drug discovery	University of Southern California	United States	—
20	Machine learning for antimicrobial peptide identification and design	Broad Institute of MIT and Harvard, University of Pennsylvania	United States	—
21	The landscape of tolerated genetic variation in humans and primates	Illumina Inc.	United States	—
22	Deciphering the Lexicon of Protein Targets: A Review on Multifaceted Drug Discovery in the Era of Artificial Intelligence	Indian Institute of Technology Kharagpur	India	—
23	Targeting DCAF5 suppresses SMARCB1-mutant cancer by stabilizing SWI/SNF	Dana-Farber Cancer Institute, Harvard Medical School, St Jude Children's Research Hospital	United States	—
24	Automated in vivo enzyme engineering accelerates biocatalyst optimization	Forschungszentrum Jülich, Leiden University, Max Planck Institute of Molecular Plant Physiology	Denmark, Germany, Netherlands	—
25	Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering.	California Institute of Technology	United States	—
26	Machine learning in preclinical drug discovery	McMaster University	Canada	—
27	Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii	Massachusetts Institute of Technology, McMaster University	Canada, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
28	Contrastive learning in protein language space predicts interactions between drugs and protein targets.	Ragon Institute of MGH, MIT and Harvard, Tufts University	United States	—
29	Machine learning modeling of family wide enzyme-substrate specificity screens.	Microsoft Research New England, MIT	United States	—
30	Modeling conformational states of proteins with AlphaFold	University of Leipzig, Vanderbilt University	Germany, United States	—

Showing the 30 most-cited of 369 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Contribution 2

Claim — Contribution 2

The researcher developed the Robetta server, a widely adopted computational platform that significantly advanced the field of protein structure prediction and analysis.

CLAIM: The researcher's primary contribution is the development of the Robetta server, as detailed in the seminal 2004 paper titled 'Protein structure prediction and analysis using the Robetta server.' This work stands as a foundational resource in computational biology, with no subsequent follow-up papers by the same researcher listed in this specific line of inquiry.

ORIGINALITY: The title suggests the introduction of a dedicated server-based approach to protein structure prediction, addressing the need for accessible, robust computational tools in structural biology. By providing a centralized platform for analysis, this work appears to have streamlined the process for researchers, distinguishing itself through its utility and implementation rather than just theoretical advancement.

SIGNIFICANCE: The impact of this contribution is evidenced by its substantial citation count of 2,534. Furthermore, citation analysis reveals that 93.9% of citing papers originate from independent researchers, indicating that the Robetta server has been widely adopted and utilized by the broader scientific community beyond the researcher's immediate circle, confirming its broad significance and utility in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 202 · **12 flagged influential by Semantic Scholar**

CORE PAPER

Protein structure prediction and analysis using the Robetta server

2004 · Nucleic acids research 32 (suppl_2), W526-W531, 2004 · 2,534 citations (GS)

Field-normalised: 2,019 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2004 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Synthetic Biology of Plants and Microbes for Agriculture, Environment, and Future Applications.	Massachusetts Institute of Technology	United States	—
2	Evolution and dynamics of protein interactions and networks	MRC Laboratory of Molecular Biology, Weizmann Institute of Science	United Kingdom	—

No.	Citing paper	Citing institution(s)	Country	S2
3	Protein design: from the aspect of water solubility and stability	Avalon GloboCare Corp., Massachusetts Institute of Technology, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS	Norway, Russia, United States	—
4	Computer-aided drug discovery for undruggable targets	Peking University	China	—
5	SARS-CoV-2 3CLpro mutations selected in a VSV-based system confer resistance to nirmatrelvir, ensitrelvir, and GC376	Institute of Virology, Medical University of Innsbruck, Texas Biomedical Research Institute	Austria, Italy, United States	Methodology
6	Protein structure prediction via deep learning: an in-depth review	Geneis Beijing Co, Hainan Normal University, Wuhan Textile University	China	—
7	Comprehensive immunoinformatics and bioinformatics strategies for designing a multi-epitope based vaccine targeting structural proteins of Nipah virus	Indian Council of Medical Research (ICMR)-National Institute of Virology	India	—
8	Tertiary structure assessment at CASP15	UKRI-STFC, University of California, Irvine Medical Center, University of Liverpool	United Kingdom, United States	Result
9	De Novo-Designed Miniprotein Inhibits the Enzymatic Activity of the SARS-CoV-2 Main Protease	Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Federal University of Pernambuco, University College London	Brazil, United Kingdom	—
10	Coarse-grained protein models and their applications	Mossakowski Medical Research Center of the Polish Academy of Sciences, University of Warsaw	Poland	—
11	Accurate de novo prediction of protein contact map by ultra-deep learning model	Toyota Technological Institute at Chicago	United States	Methodology
12	Molecular docking in organic, inorganic, and hybrid systems: a tutorial review	Kalinga Institute of Industrial Technology (KIIT)	India	—
13	Homology modeling in drug discovery: Overview, current applications, and future perspectives	Ankara University, Suleyman Demirel University	Turkey	—
14	Application of molecular dynamics simulation in biomedicine	Shantou University Medical College	China	Background
15	Protein structure and function prediction using I-TASSER	University of Michigan	United States	—
16	Distance-based protein folding powered by deep learning	Toyota Technological Institute at Chicago	United States	—
17	Improved PEP-FOLD approach for peptide and miniprotein structure prediction	Inserm, Institut de Biologie Physico-Chimique	France	—

No.	Citing paper	Citing institution(s)	Country	S2
18	Integrated molecular modeling and machine learning for drug design	New York University	United States	—
19	Hyaluronidases: their genomics, structures, and mechanisms of action	Children's Hospital Oakland Research Institute	United States	—
20	Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication	University of Wisconsin-Madison	United States	—
21	Predicting protein–protein interactions from the molecular to the proteome level	Middle East Technical University	Turkey	—
22	Designing of multi-epitope peptide vaccine against Acinetobacter baumannii through combined immunoinformatics and protein interaction–based approaches	BVG Life Sciences Limited, JSS Academy of Higher Education and Research, Kalinga Institute of Industrial Technology (KIIT)	India	—
23	Protein design: toward functional metalloenzymes	University of Michigan	United States	—
24	Genome-Wide RNAi Screen Identifies Letm1 as a Mitochondrial Ca2+/H+ Antiporter	Children's Hospital Boston	United States	—
25	Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants	BGI-Shenzhen, The University of Hong Kong, The University of Western Australia	Australia, China	—
26	Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You	Vanderbilt University	United States	**Methodology**
27	HADDOCK versus HADDOCK: new features and performance of HADDOCK2. 0 on the CAPRI targets	Utrecht University	Netherlands	Background
28	Simulation studies of amyloidogenic polypeptides and their aggregates	University of Zürich	Switzerland	—
29	Development of multi-epitope vaccines against the monkeypox virus based on envelope proteins using immunoinformatics approaches	Central South University, Xiangya Hospital	China	Background
30	Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry	Eidgenössische Technische Hochschule Zürich	Switzerland	—

Showing the 30 most-cited of 202 independent citing papers.

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY SARS-CoV-2 3CLpro mutations selected in a VSV-based system confer resistance to nirmatrelvir, ensitrelvir, and GC376

> *"For the catalytic site mutations, a resistance mechanism was postulated on the basis of mapping the mutations onto the cocrystal structure of 3CLpro-nirmatrelvir and generating mutant models with Robetta (15) and MOE (16)."*

RESULT Tertiary structure assessment at CASP15

> *"This tentative observation is in complete contrast to results in the time of fragment assembly ab initio methods, for example, 56"*

METHODOLOGY Accurate de novo prediction of protein contact map by ultra-deep learning model

"17), all the other CAMEO-participating servers, including Robetta, HHpred, RaptorX, SPARKS-X, and RBO Aleph (template-based and ab initio folding) only submitted models with TMscore≤0.48 and RMSD>43.82Å."

METHODOLOGY Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You

"Large parts of the R OSETTA protein structure prediction protocol, including generation of fragments, de novo folding, and comparative modeling, have been replicated in an automated server R OBETTA (30 , 47 , 48)."

Contribution 3

Claim — Contribution 3

The researcher developed foundational methods for protein structure prediction using Rosetta, establishing a widely adopted computational framework that has significantly advanced the field of structural bioinformatics.

The researcher's primary contribution centers on the development of computational methods for protein structure prediction using the Rosetta software suite, as detailed in their seminal 2004 paper. This work stands as a cornerstone in the field, providing a robust framework for modeling protein conformations that has become integral to modern structural biology research.

This line of work appears to address the critical challenge of accurately determining protein three-dimensional structures from amino acid sequences, a problem that was historically difficult due to the complexity of energy landscapes. By leveraging the Rosetta platform, the researcher introduced a methodological approach that likely improved the accuracy and efficiency of structure prediction, offering a new standard for computational modeling in the mid-2000s.

The significance of this contribution is evidenced by its substantial citation count of 2,365, indicating widespread adoption and influence within the scientific community. Furthermore, citation analysis reveals that 93.9% of citing papers originate from independent researchers, demonstrating that this work has served as a foundational tool for diverse groups outside the researcher's immediate circle, thereby confirming its broad impact and utility in advancing protein structure prediction globally.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7

CORE PAPER

Protein structure prediction using Rosetta

2004 · Methods in enzymology 383, 66-93, 2004 · 2,365 citations (GS)

Field-normalised: 1,703 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2004 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Deep-learning-based single-domain and multidomain protein structure prediction with D-I-TASSER	Michigan State University, Nankai University, National University of Singapore	China, Singapore, United States	—
2	Unified rational protein engineering with sequence-based deep representation learning	Harvard Medical School, Harvard University, Massachusetts Institute of Technology	United States	—
3	Protein design: from the aspect of water solubility and stability	Avalon GloboCare Corp., Massachusetts Institute of Technology, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry RAS	Norway, Russia, United States	—
4	The realm of unconventional noncovalent interactions in proteins: their significance in structure and function	Indian Institute of Science Education and Research	India	—

No.	Citing paper	Citing institution(s)	Country	S2
5	Coarse-grained protein models and their applications	Mossakowski Medical Research Center of the Polish Academy of Sciences, University of Warsaw	Poland	—
6	Protein structure and sequence generation with equivariant denoising diffusion probabilistic models	Stanford University	United States	Background
7	Computational biology and artificial intelligence in mRNA vaccine design for cancer immunotherapy	Islamic Azad University, Tehran Medical Sciences, Medical University of Warsaw, Southwest Medical University	China, Iran, Poland	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
University of Washington	United States	SCImago #45 · THE 25 · QS 81	57
University of California, Irvine Medical Center	U.S.A	—	54
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	49
Stanford University	United States	SCImago #18 · THE =5 · QS 3	34
University of Oxford	United Kingdom	SCImago #26 · THE 1 · QS 4	31
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	31
Harvard University	United States	SCImago #4 · THE =5 · QS 5	25
Peking University	China	SCImago #11 · THE 13 · QS 14	24
Tsinghua University	PR China	SCImago #8 · THE 12 · QS =17	22
Princeton University	United States	SCImago #386 · THE =3 · QS =25	22
Shanghai Jiao Tong University	China	SCImago #10 · THE 40 · QS =47	21
University of Pennsylvania	United States	SCImago #52 · THE 14 · QS 15	20
Harvard Medical School	United States	SCImago #12	15
Université de Montréal	Canada	SCImago #692 · THE 150 · QS 168	14
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	14

Geographic distribution of citing authors

Country	Citing papers
United States	516
China	235
United Kingdom	136

Country	Citing papers
Germany	80
Canada	56
Switzerland	52
India	49
France	39
Japan	33
Netherlands	32
South Korea	31
Australia	28

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.

F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	—	369	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Protein structure prediction and analysis using the Robetta server	202	Dhanasar — Prong 2 (well-positioned)
Contribution 3	Protein structure prediction using Rosetta	7	Dhanasar — Prong 2 (well-positioned)