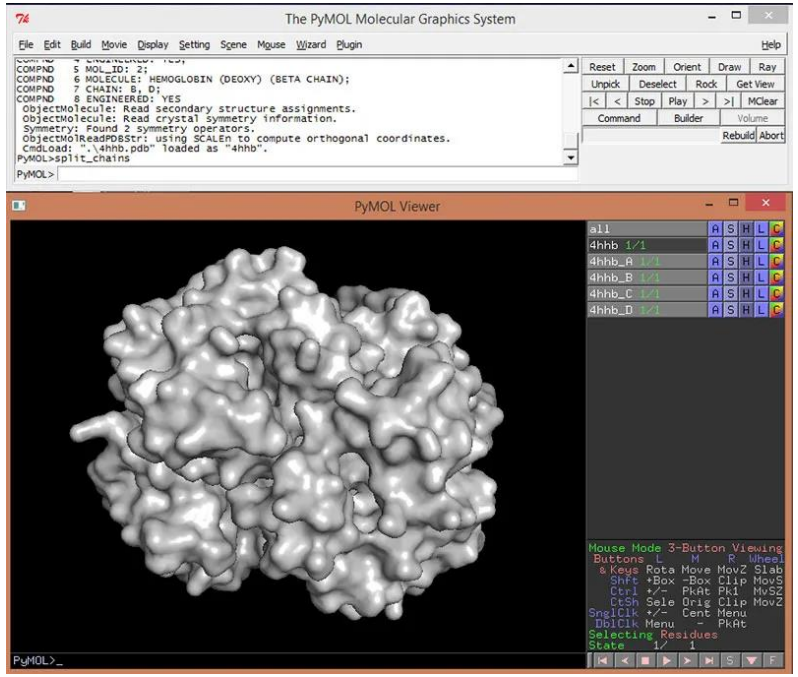
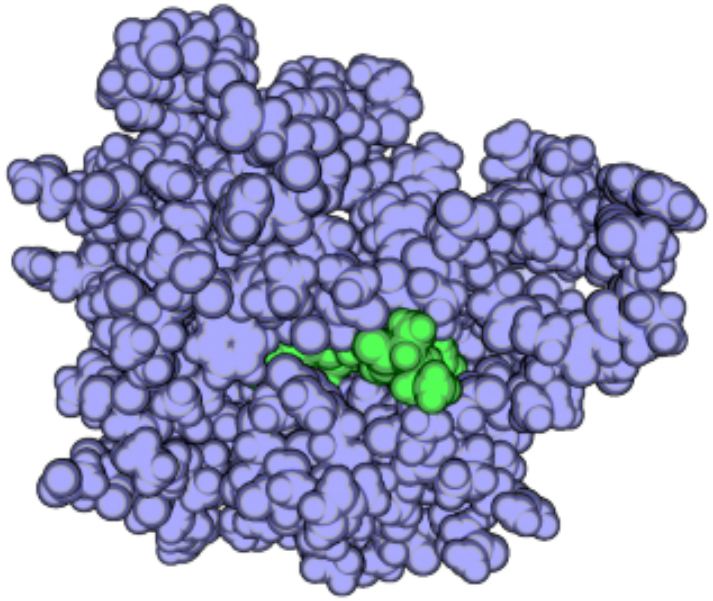
GoMol_3D Protein Analysis and Visualization Tool

Darin Boyes, Minhyek Jeon , Shivank Sadasivan, Tyler Katz,

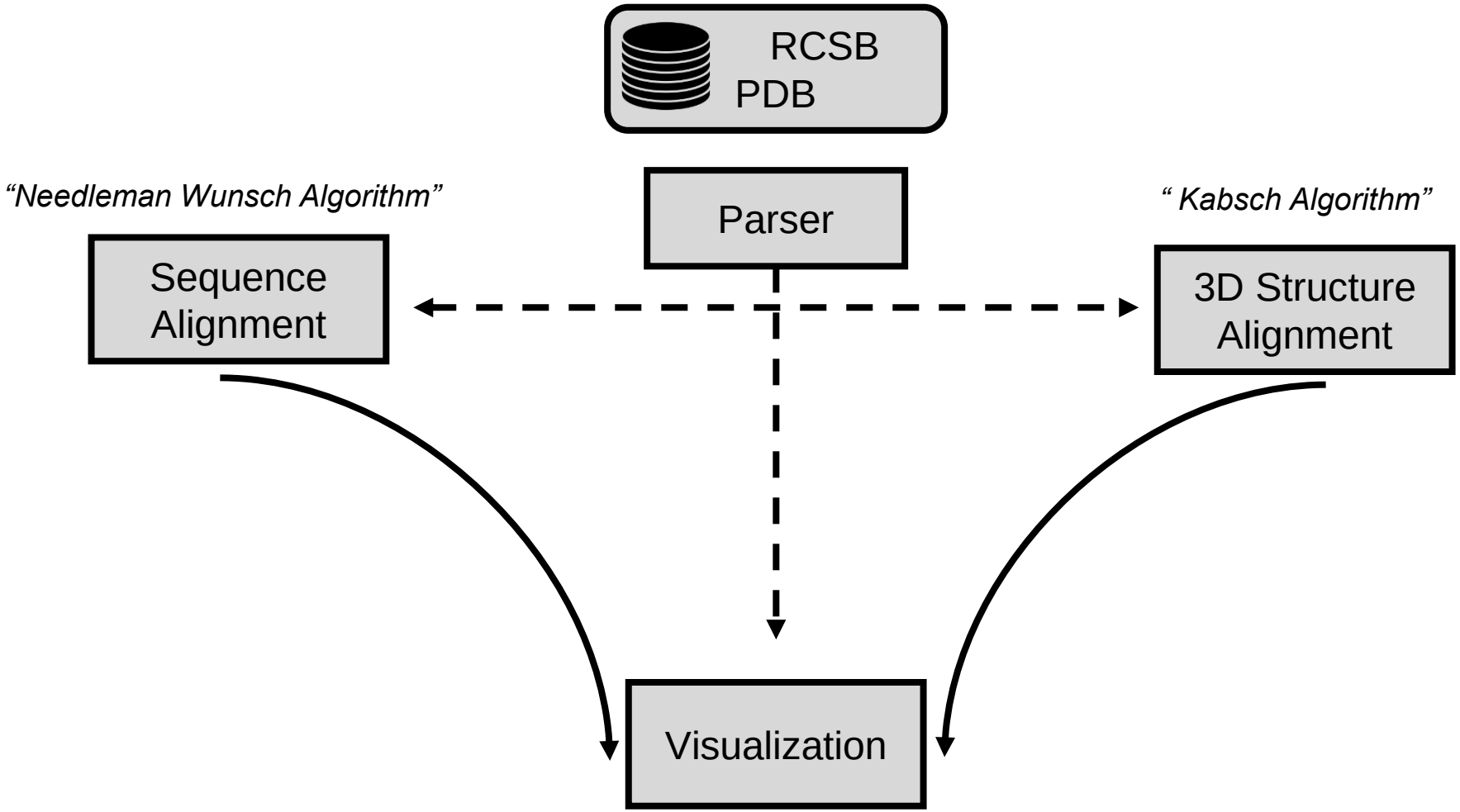
CONTENTS

“GoMol : A 3D Protein Analysis and Visualization Tool ”





- Proteins are an integral part of all biological systems
- Create a tool to compare sequence and 3D structure of protein
- Develop the tool by mostly using Go programming language


RCSB
PDB
Parser
"Needleman Wunsch Algorithm"
" Kabsch Algorithm"
Sequence Alignment
3D Structure Alignment
Visualization

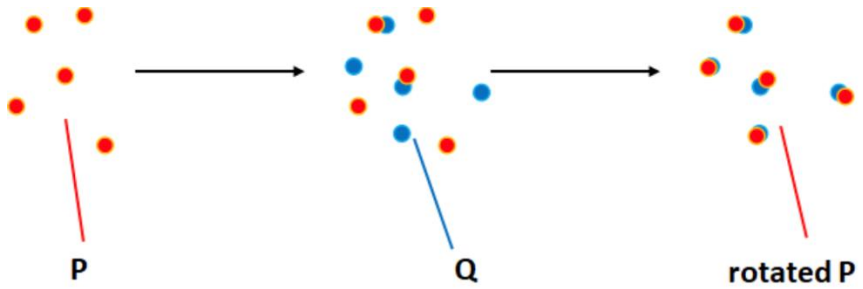
Sequence Alignment

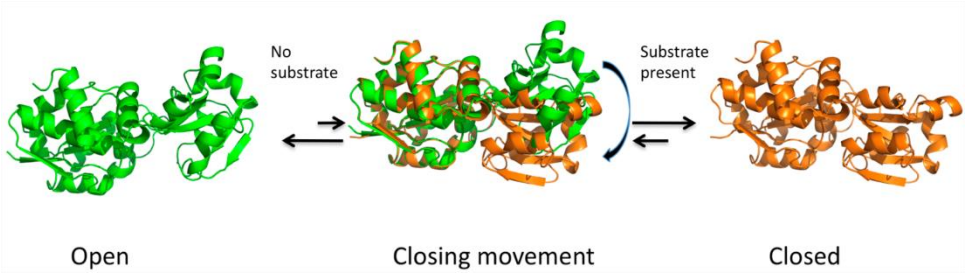
```
 * Blosum-62 substitution matrix
*  #   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V
*  A   4 -1 -2 -2  0 -1 -1  0 -2 -1 -1 -1 -1 -2 -1  1  0 -3 -2  0
*  R  -1  5  0 -2 -3  1  0 -2  0 -3 -2  2 -1 -3 -2 -1 -1 -3 -2 -3
*  N  -2  0  6  1 -3  0  0  0  1 -3 -3  0 -2 -3 -2  1  0 -4 -2 -3
*  D  -2 -2  1  6 -3  0  2 -1 -1 -3 -4 -1 -3 -3 -1  0 -1 -4 -3 -3
*  C   0 -3 -3 -3  9 -3 -4 -3 -3 -1 -1 -3 -1 -2 -3 -1 -1 -2 -2 -1
*  Q  -1  1  0  0 -3  5  2 -2  0 -3 -2  1  0 -3 -1  0 -1 -2 -1 -2
*  E  -1  0  0  2 -4  2  5 -2  0 -3 -3  1 -2 -3 -1  0 -1 -3 -2 -2
*  G   0 -2  0 -1 -3 -2 -2  6 -2 -4 -4 -2 -3 -3 -2  0 -2 -2 -3 -3
*  H  -2  0  1 -1 -3  0  0 -2  8 -3 -3 -1 -2 -1 -2 -1 -2 -2  2 -3
*  I  -1 -3 -3 -3 -1 -3 -3 -4 -3  4  2 -3  1  0 -3 -2 -1 -3 -1  3
*  L  -1 -2 -3 -4 -1 -2 -3 -4 -3  2  4 -2  2  0 -3 -2 -1 -2 -1  1
*  K  -1  2  0 -1 -3  1  1 -2 -1 -3 -2  5 -1 -3 -1  0 -1 -3 -2 -2
*  M  -1 -1 -2 -3 -1  0 -2 -3 -2  1  2 -1  5  0 -2 -1 -1 -1 -1  1
*  F  -2 -3 -3 -3 -2 -3 -3 -3 -1  0  0 -3  0  6 -4 -2 -2  1  3 -1
*  P  -1 -2 -2 -1 -3 -1 -1 -2 -2 -3 -3 -1 -2 -4  7 -1 -1 -4 -3 -2
*  S   1 -1  1  0 -1  0  0  0 -1 -2 -2  0 -1 -2 -1  4  1 -3 -2 -2
*  T   0 -1  0 -1 -1 -1 -1 -2 -2 -1 -1 -1 -1 -2 -1  1  5 -2 -2  0
*  W  -3 -3 -4 -4 -2 -2 -3 -2 -2 -3 -2 -3 -1  1 -4 -3 -2 11  2 -3
*  Y  -2 -2 -2 -3 -2 -1 -2 -3  2 -1 -1 -2 -1  3 -3 -2 -2  2  7 -1
*  V   0 -3 -3 -3 -1 -2 -2 -3 -3  3  1 -2  1 -1 -2 -2  0 -3 -1  4
```

```
MSSPVAKAARRVTHELHGVVVSAGLMDKTVKVRVGGQKWNKIVNKWFADPKHYLVHDPNSSLRTGDVVSIVPGWPTSQHKRHVIKHIIA
 ||| | |||||||||||||||||||| ||||||| ||||||||||  |||||||||||||||||||| | |||||| ||||| | |||
MSSQVSKAARRVTHELHGVVVSAGLMQKTVKVRVAGQKWNKIVNKFYADPKHYLVHDPNSSLRTGDVVAIAPGWPTSRHKRHVVKQIIA
Percentage Similarity: 69.38%
```

- To identify sequence of similarities and differences
- Similar length and significantly similar (evolutionary wise)
- Global Alignment
- Needleman Wunsch Algorithm (Dynamic Programming)
- BLOSUM62 (Substitution Matrix)

3D Alignment





「Kabsch Algorithm」

- Comparison of two similar proteins by RMSD
- Calculates optimal rotation matrix to minimize RMSD

「Applications of Kabsch」

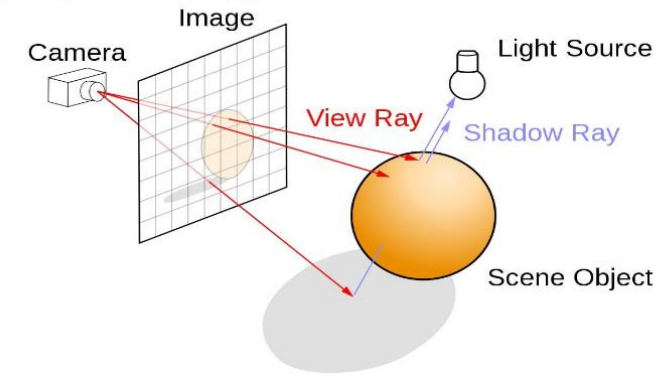
- Drug Discovery/Molecular Dynamics - Analysis of conformational states
- Evolutionary Biology - structural comparison gives insight to conservation

3D Visualization

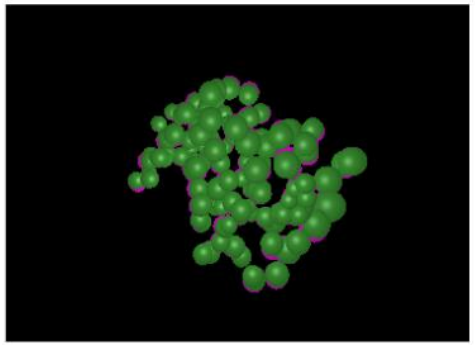
Input (PDB file)

Type	#	El	AA	Chain	AA#	x	y	z
ATOM	1	N	VAL	A	1	10.720	19.523	6.163
ATOM	2	CA	VAL	A	1	10.228	20.761	6.807
ATOM	3	C	VAL	A	1	8.705	20.714	6.878
ATOM	4	O	VAL	A	1	8.164	20.005	6.015
ATOM	5	CB	VAL	A	1	10.602	22.000	5.966
ATOM	6	CG1	VAL	A	1	10.307	23.296	6.700
ATOM	7	CG2	VAL	A	1	12.065	21.951	5.544

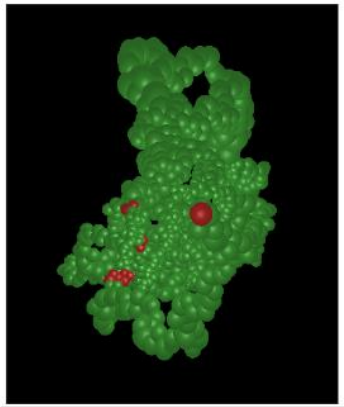
Ray Tracing



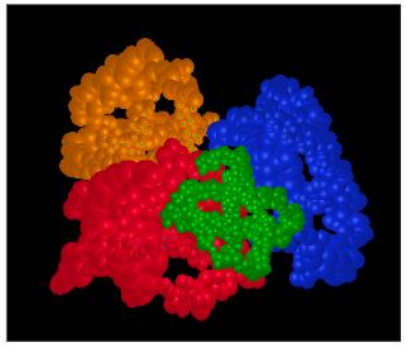
Output



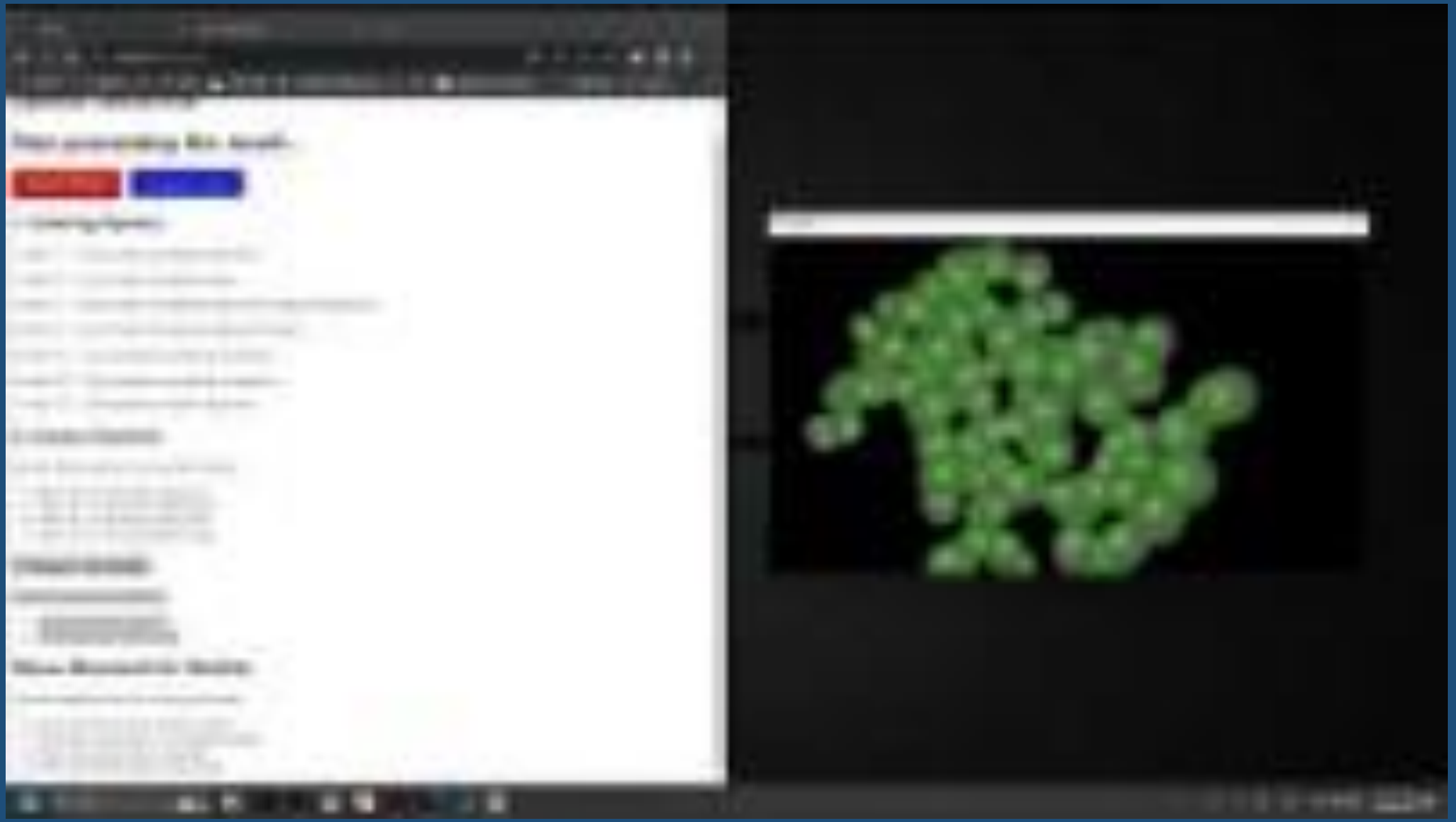
Visualization of Kabsch algorithm on 1LYZ and 2LYZ (2 lysozyme proteins)



Needleman-Wunsch visualization of 1A60 aligned with 4ANM (CK2)

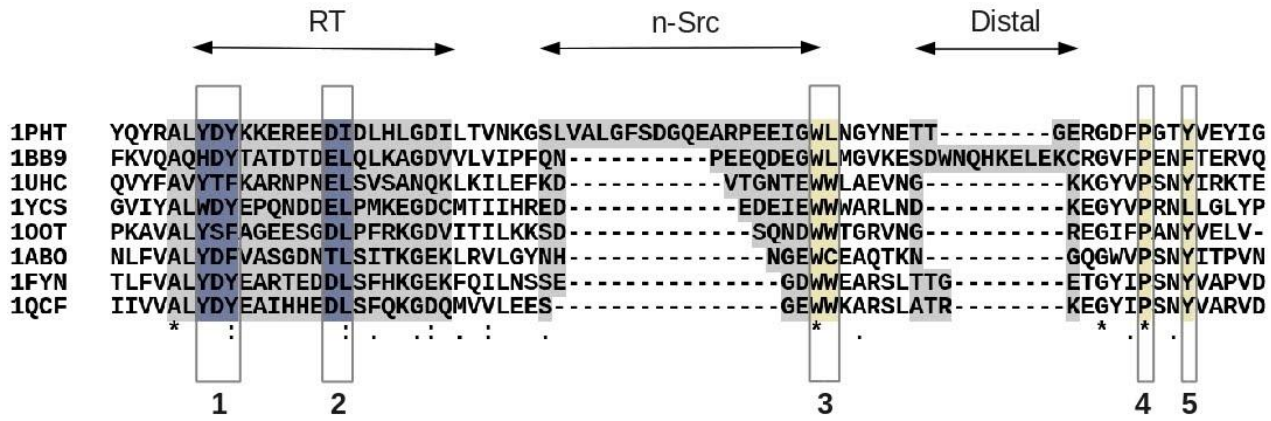


Side chain rendering of 1A3N (Hemoglobin)



Limitation & Future Directions

Figure-1



「Possible Improvements in Program 」

- Gap Costs
- Leverage sequence alignment results to broaden potential inputs for Kabsch
- Potentially include information about ligand binding sites from PDB

「Limitation of our Program」

- Slow Speed, pixel level
- Protein Multiple Sequence Alignment Not possible
- No Universal Parser
- No Local alignment

References

1.Libretexts. (2021, March 6). *9.5: Blosum62 scoring matrix for amino acid substitutions*. Biology LibreTexts. https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Investigations_in_Molecular_Cell_Biology_(O%27Connor)/09%3A_Protein_Conservation/9.05%3A_BLOSUM62_scoring_matrix_for_amino_acid_substitutions

1.Libretexts. (2021b, March 17). *3.3: Global alignment vs. local alignment vs. semi-global alignment*. Biology LibreTexts. https://bio.libretexts.org/Bookshelves/Computational_Biology/Book%3A_Computational_Biology_-_Genomes_Networks_and_Evolution_(Kellis_et_al.)/03%3A_Rapid_Sequence_Alignment_and_Database_Search/3.03%3A_Global_alignment_vs._Local_alignment_vs._Semi-global_alignment

1.NandiniUmbarkar. (2020, October 12). *Needleman-Wunsch algorithm*. Medium. https://medium.com/@nandiniumbarkar/needleman-wunsch-algorithm-7bba68b510db

1.Phong, B.. (1975, June 1). *Illumination for computer generated pictures*. Communications of the ACM. https://dl.acm.org/doi/10.1145/360825.360839

1.Compeau, P. (2022). Protein Structure Comparison. Biological Modeling. Retrieved December 3, 2023, from https://biologicalmodeling.org/coronavirus/accuracy

1.Kabsch algorithm. (2023). Wikipedia. Retrieved December 3, 2023, from https://en.wikipedia.org/wiki/Kabsch_algorithm

Vertrees, J. (2019, April 18). Kabsch. PyMOLWiki. Retrieved December 3, 2023, from https://pymolwiki.org/index.php/Kabsch

Thank you