

1/30/2020 Week 3 Module 2

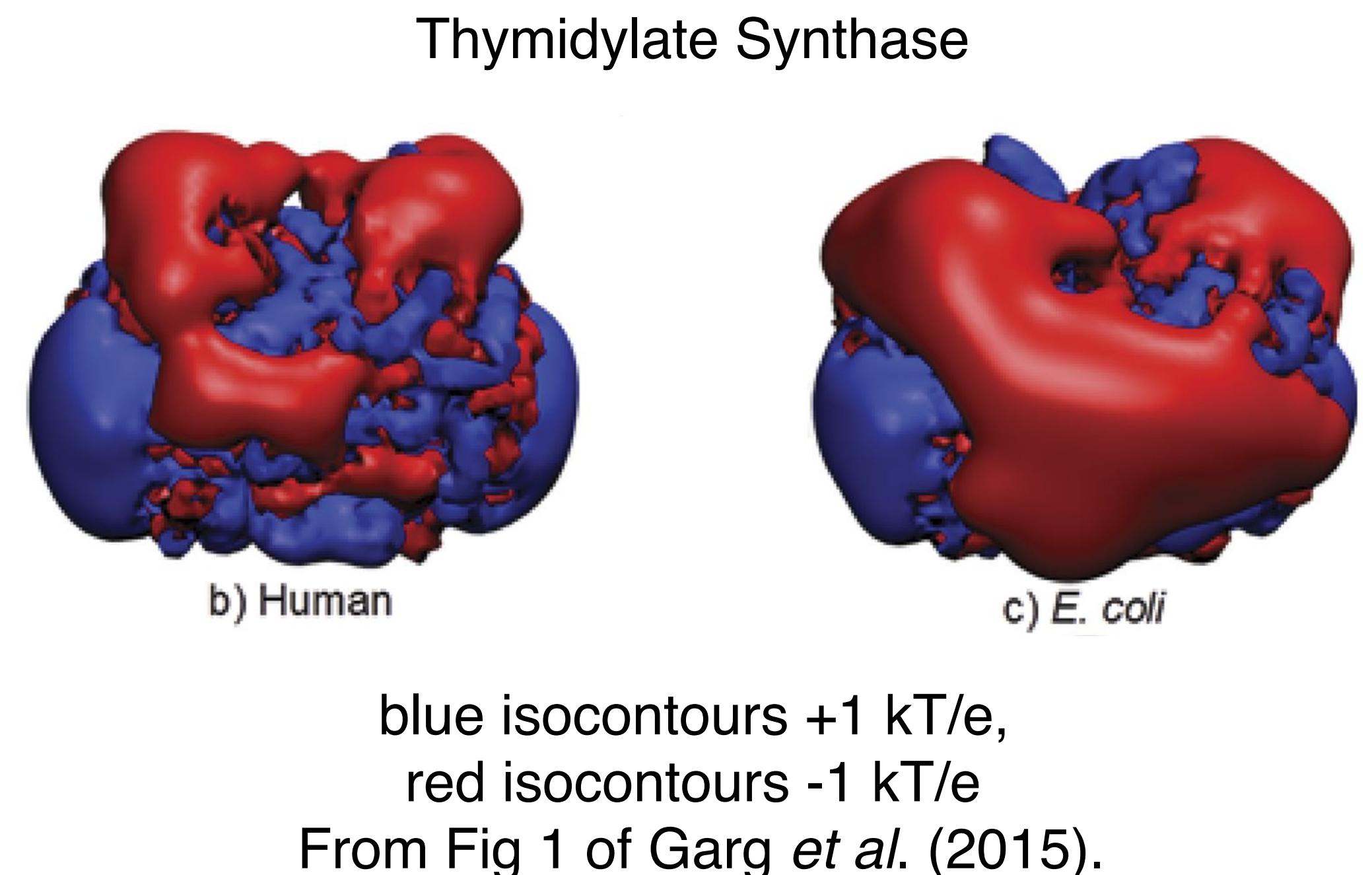
Structure Prediction and Electrostatics Calculations

- This module will consist of mini-lectures, a “tour” of structure prediction and electrostatics calculations, and a visualization exercise.
- We will reproduce some key results from this paper:

The image shows a screenshot of a scientific publication page. At the top right, the URL www.nature.com/scientificreports is visible. The title "SCIENTIFIC REPORTS" is prominently displayed in large black letters, with a red gear icon integrated into the letter "O". Below the title, the word "OPEN" is written in orange. The main title of the article is "Conservation and Role of Electrostatics in Thymidylate Synthase". The authors listed are Divita Garg^{1,2,3,†}, Stephane Skouloubris^{4,5}, Julien Briffotaux^{4,1}, Hannu Myllykallio⁴ & Rebecca C. Wade^{1,6,7}. The text "Received: 30 May 2015" is at the bottom left, followed by "Accepted: 28 October 2015" and "Published: 27 November 2015". A small number "1" is located at the bottom center of the page.

Electrostatics in Biomacromolecules

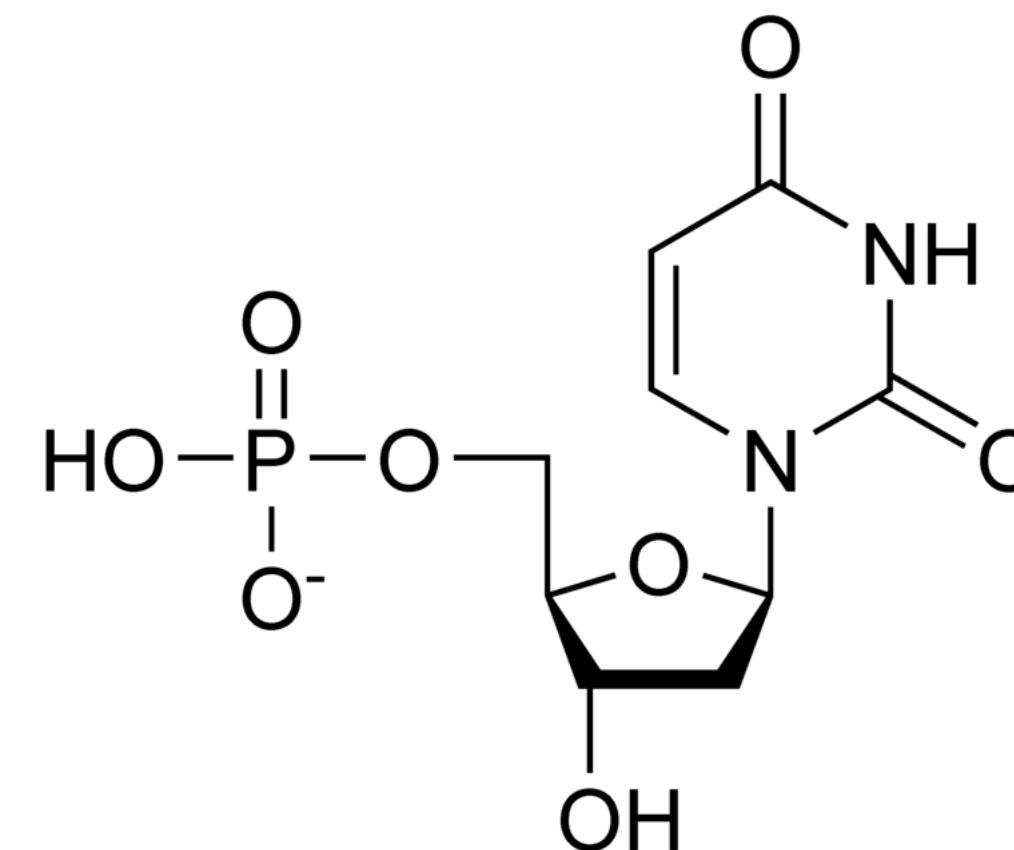
- Electrostatics important in (at least)
 - binding for
 - steering, facilitating approach of species
 - complexation, as complementarity means lower potential energy
 - enzyme catalysis, as electric potential stabilizes transition state
- Thus, electrostatic potential usually conserved near functional sites



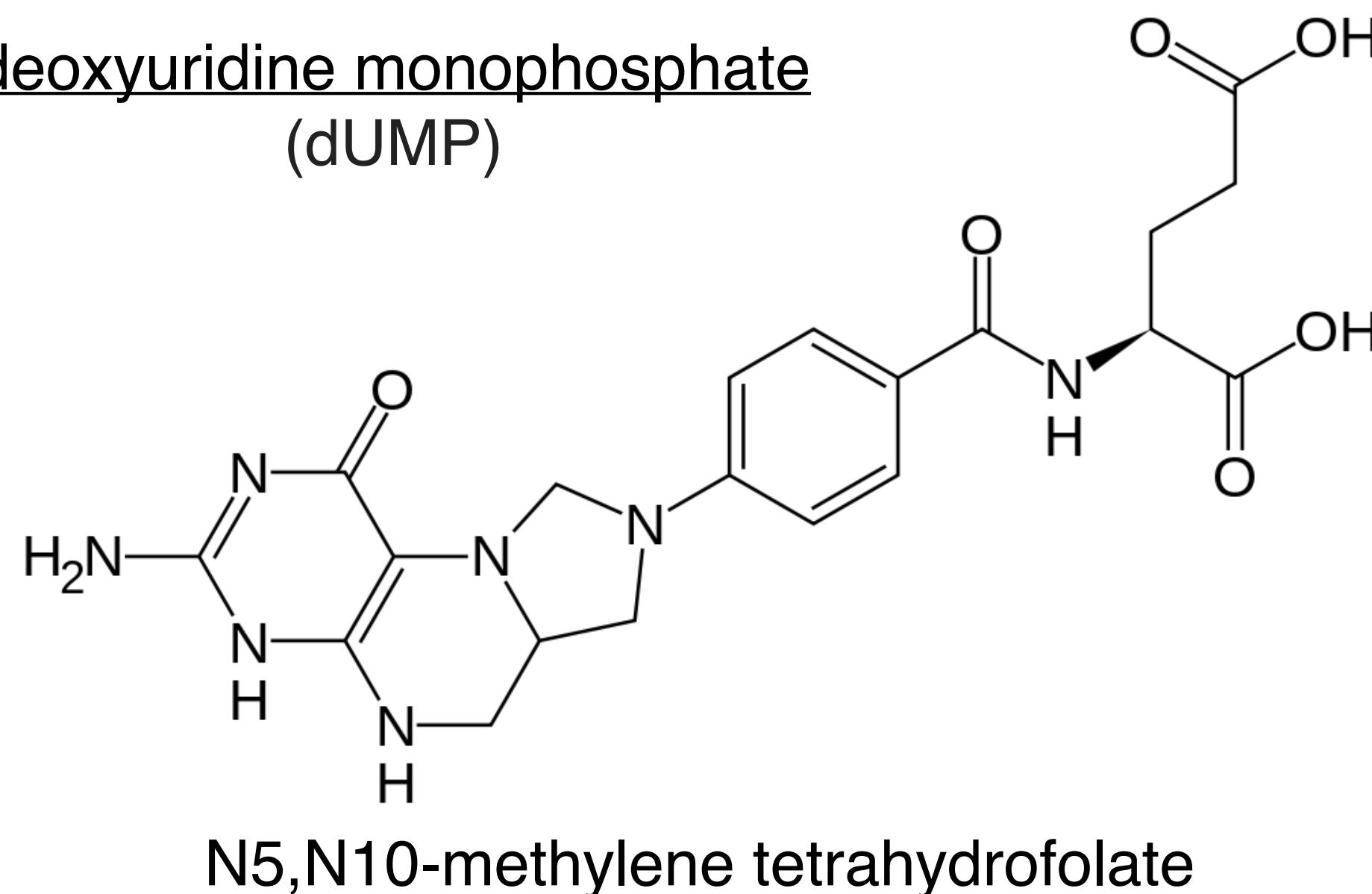
Modeling Electrostatics

- In biological macromolecules, the electrostatic potential is usually calculated based on the Poisson-Boltzmann equation
 - The Poisson equation describes the potential field due to a given charge distribution. Atoms in the biomolecule are assumed to have a fixed charge.
 - The Poisson-Boltzmann equation assumes that (infinitely small) ions surround a biomolecule in accordance with the Boltzmann distribution
- The PB equation is a partial differential equation that is solved numerically
- The equation is often linearized to be more numerically stable
- Chun Liu in Applied Math has worked on versions
 - that are time-dependent
 - account for the finite size of ions

Thymidylate Synthase Catalyzes

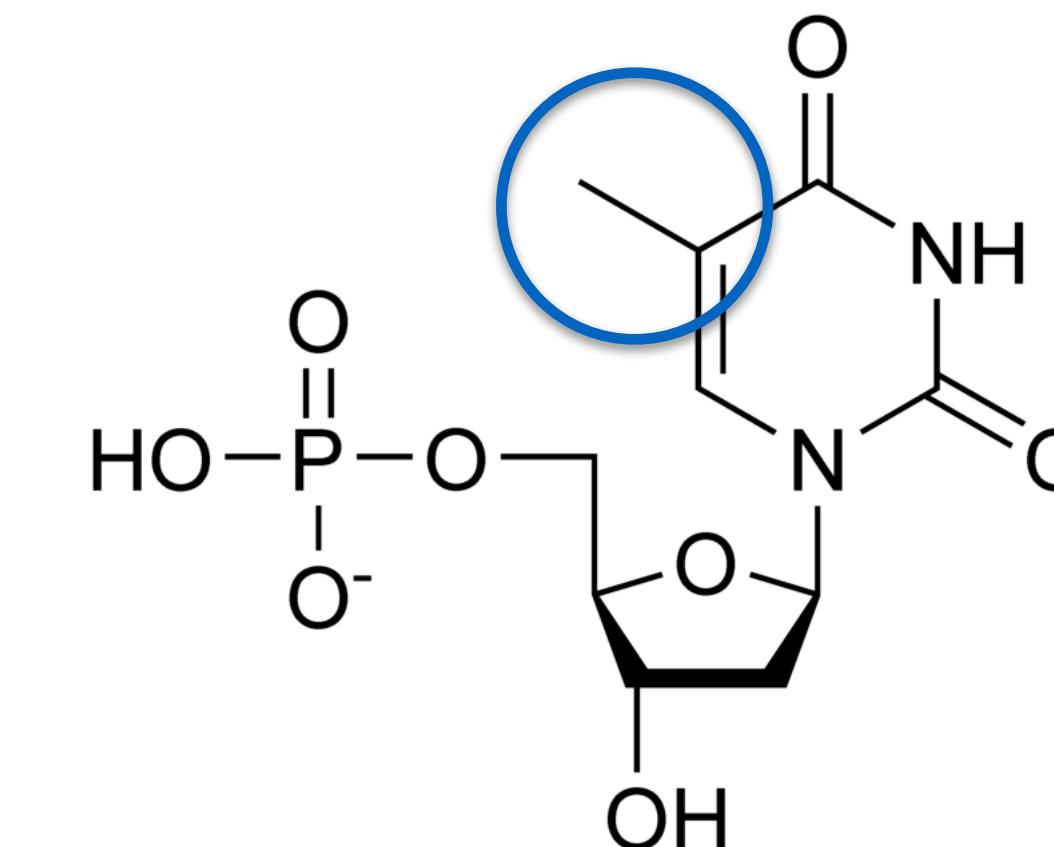


deoxyuridine monophosphate
(dUMP)

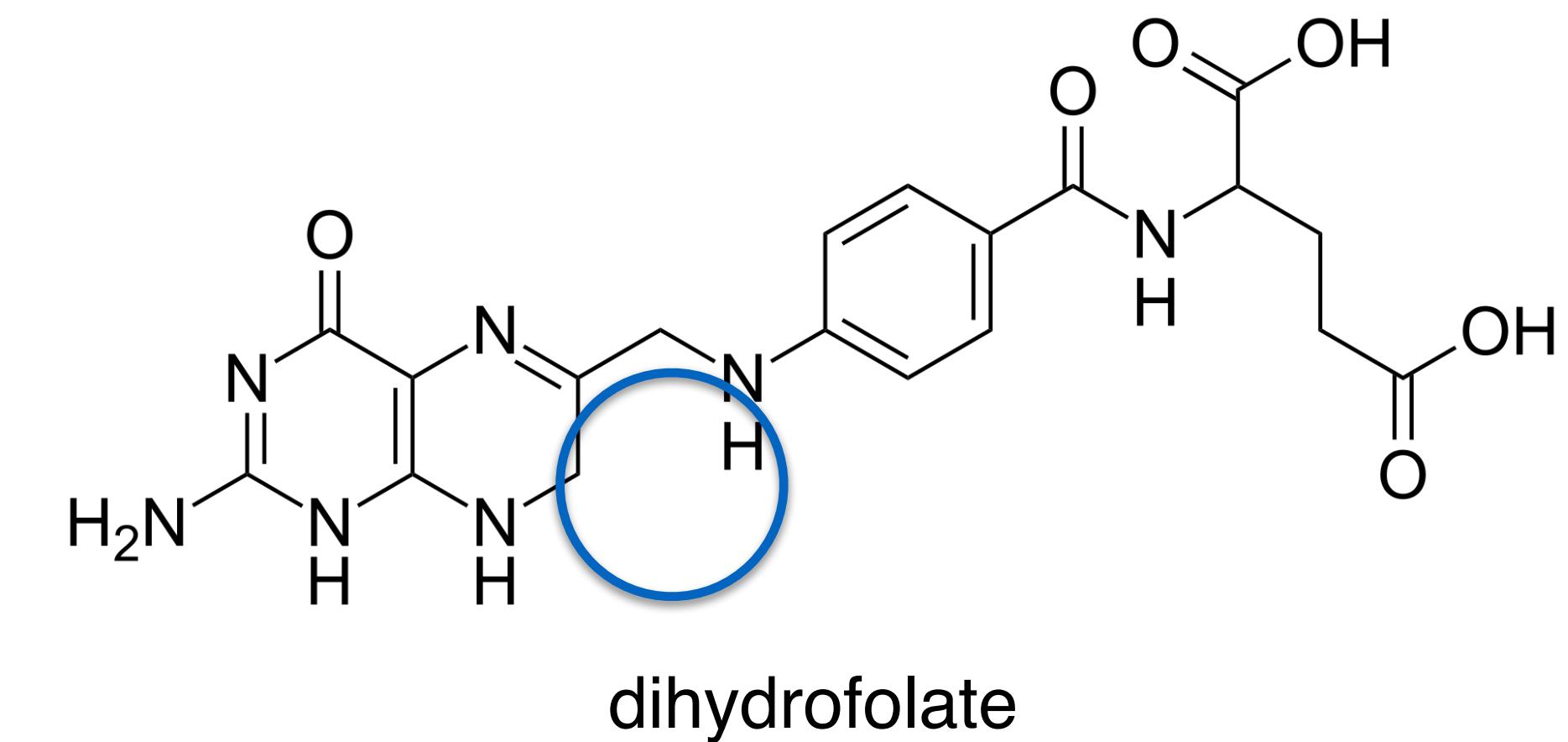


N5,N10-methylene tetrahydrofolate

to



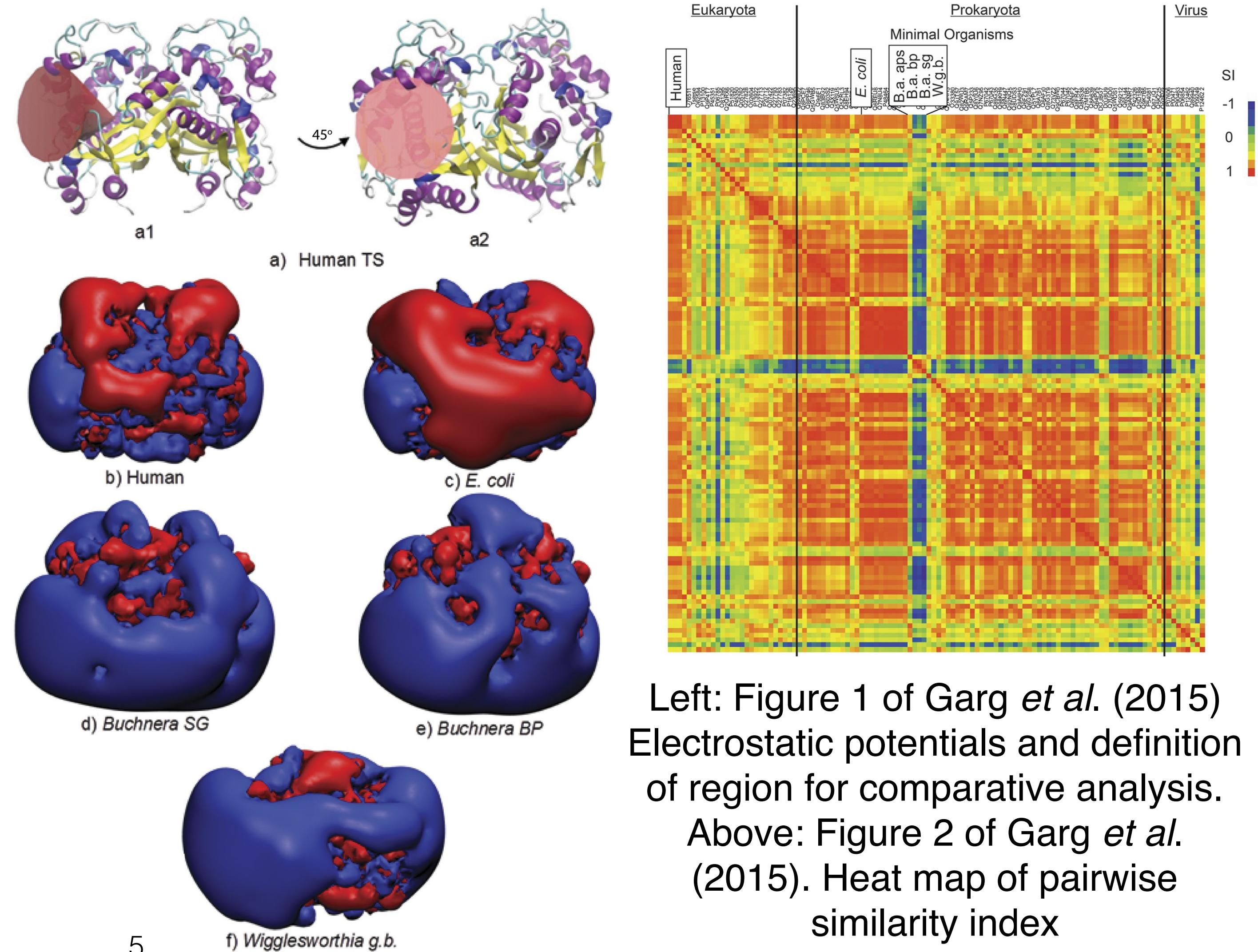
deoxythymidine monophosphate
(dTDP)



dihydrofolate

Summary of “Conservation and Role of Electrostatics in Thymidylate Synthase”

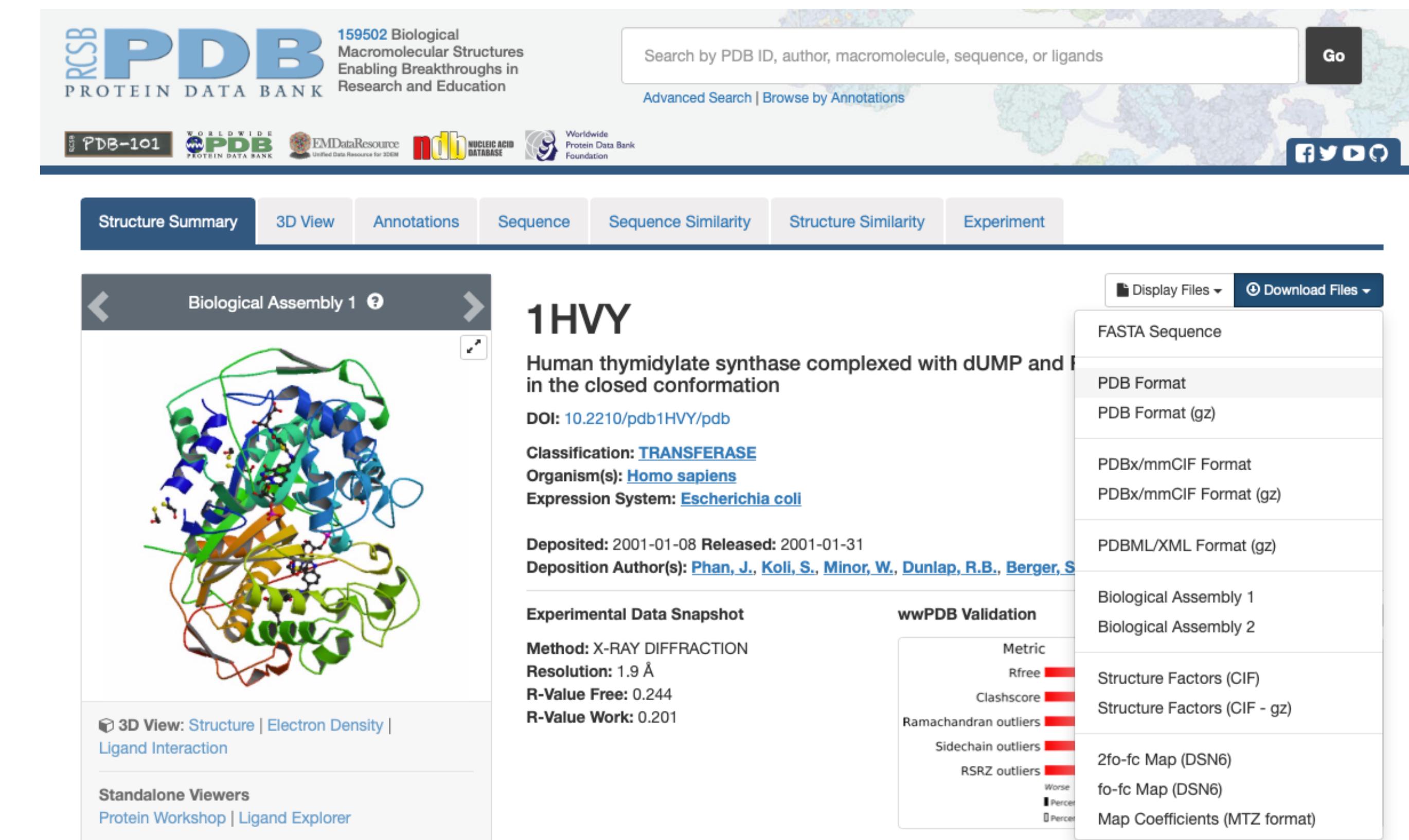
- Built 110 homology models
- Calculated and compared electrostatic potential of the enzyme across species
- Found minimal organisms, including *Wigglesworthia glossinidias brevipalpis* (W.g.b.), to have divergent potential
- Rationalized W.g.b. TS to be functional and unsuccessfully tried to express and purify it
- Mutated *E. coli* TS to be more like W.g.b. TS and found the enzyme to be less active



Plan for reproducing key results from “Conservation and Role of Electrostatics in Thymidylate Synthase”

- To reproduce the key results, we'll need computational models of thymidylate synthase from *homo sapiens*, *E. coli*, and W.g.b.
 - For the former two, there are many structures of TS in the PDB. Garg et. al. used PDB ID 1HVY for *homo sapiens* and 2G8O for *E. coli*
 - For the latter, we will use a model from the I-TASSER web server
- Next, we'll align the models with the MultiSeq module in VMD
- We'll then calculate the electrostatic potential with PDB2PQR and APBS
- Finally, we'll visualize the results in VMD
- In interest of time, I will not ask you to do every step. Instead, I will guide you through what I did and ask you to download results from previous calculations.

- Getting structures from the PDB (<https://www.rcsb.org/>) is pretty straightforward.
- For a particular crystal structure, you can just click on “Download Files” and select “PDB format”
- We will use the PDB files for 1HVY and 6NNR, which has superceded 2G8O.
- Next, let’s go through getting a model for W.g.b. thymidylate synthase.



Structure Prediction

Structure Prediction Principles

- **Homology modeling** is most useful for closer homology. Go through <http://www.bioinfo.rpi.edu/bystrc/courses/biol4550/lecture7/assets/player/KeynoteDHTMLPlayer.html#0> up to slide 11
- **Threading** is more useful for distant homologues. Watch <https://www.jove.com/video/3259/a-protocol-for-computer-based-protein-structure-function> up through minute 1
- Differences between approaches not completely distinct

Choosing Structure Prediction Software

- There are many software tools for protein structure prediction (see https://en.wikipedia.org/wiki/List_of_protein_structure_prediction_software)
- How should you decide which to use?
 - Ease of use
 - Web server - easier for sporadic use
 - Downloadable and scriptable - easier for large-scale applications
 - Accuracy
- The “Critical Assessment of protein Structure Prediction” (CASP) experiments are *blinded* tests of the ability to predict structure from sequence. (see <http://www.predictioncenter.org/index.cgi>)
- “I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, and CASP13 experiments.” So let’s use it!

- To reproduce the key results from Garg *et. al.*, we first need to find out what to model
- UniProt (<https://www.uniprot.org>) is a comprehensive biological sequence database
- A search for “thymidylate synthase” yields 50,099 results, including 566 that have been manually reviewed. This is much more than the 110 that were included by Garg *et. al.* in their 2015 paper!
- The two results for “thymidylate synthase wigglesworthia” is much more manageable. Select the result that has been reviewed.

UniProtKB results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q7W1A9	TYSY_BORPA	Thymidylate synthase	thyA BPP0787	Bordetella parapertussis (strain 12822 / ATCC BAA-587 / NCTC 13253)	323
C5BMA3	TYSY_TERTT	Thymidylate synthase	thyA TERTU_0365	Teredinibacter turnerae (strain ATCC 39867 / T7901)	277
Q8EV81	TYSY_MYCPE	Thymidylate synthase	thyA MYPE6870	Mycoplasma penetrans (strain HF-2)	289
A3PYA8	TYSY_MYCSJ	Thymidylate synthase	thyA Mjis_2099	Mycobacterium sp. (strain JLS)	266
P0C0M5	TYSY_STAES	Thymidylate synthase	thyA SE_1120	Staphylococcus epidermidis (strain ATCC 12228)	318
Q6FZ91	TYSY_BARQU	Thymidylate synthase	thyA BQ08630	Bartonella quintana (strain Toulouse) (Rochalimaea quintana)	264
A9IW82	TYSY_BART1	Thymidylate synthase	thyA BT_1568	Bartonella tribocorum (strain CIP 105476 / IBS 506)	264
Q6MID2	TYSY_BDEBA	Thymidylate synthase	thyA Bd3230	Bdellovibrio bacteriovorus (strain ATCC 15356 / DSM 50701 / NCIB 9529 / HD100)	264

UniProtKB results

UniProtKB consists of two sections:

- Reviewed (Swiss-Prot) - Manually annotated**
Records with information extracted from literature and curator-evaluated computational analysis.
- Unreviewed (TrEMBL) - Computationally analyzed**
Records that await full manual annotation.

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

Entry	Entry name	Protein names	Gene names	Organism	Length
Q8D2N4	TYSY_WIGBR	Thymidylate synthase	thyA WIGBR3200	Wigglesworthia glossinidia brevipalpis	264
H6Q4Z7	H6Q4Z7_WIGGL	Thymidylate synthase	thyA WIGMOR_0450	Wigglesworthia glossinidia endosymbiont of Glossina morsitans morsitans (Yale colony)	264

- To get the actual amino acid sequence, click on “Sequence” or scroll down and then click on FASTA. FASTA is a simple format for amino acid sequences based on one-letter codes.
- To run I-TASSER, you can visit the server interface at <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>, register, paste the sequence into the appropriate field, and press “Run I-TASSER”.
- But let's spare their servers and avoid the wait by not all submitting the same job.*

The top screenshot shows a protein sequence entry from UniProt. The sequence is:

```

10      20      30      40      50
MKEYLDLLNL ILKNQYPKID RTKTGTLSMF GYQIRINLNE GFPLLTTKYC
60      70      80      90      100
HFKSIVYELL WFLRGDTNIS FLKNNNISIW NKWADKNGNL GPIYCKQWRA
110     120     130     140     150
WEDKKNNNTID QIEIALNKLK KEPSSRRILV SSWNVGELDL MSIPPCHVLF
160     170     180     190     200
QLYVINNNKLS CQVYQRSCDI FGLGPFNIGS YALLTHIFAN QCDLLVEDLI
210     220     230     240     250
WTGGDIHLYK NHLNQAKLQL TRSPLPLPKI FIKKKPKNLF NYAFNDPLLI
260
DYNHHPKIKA PISI
  
```

The bottom screenshot shows the I-TASSER online server interface. The sequence input field contains:

```
>sp|Q8D2N4|ITYSY_WIGBR Thymidylate synthase OS=Wigglesworthia glossinidia brevipalpis  
OX=36870 GN=itysy PE=3 SV=1  
MKEYLDLLNLKQYKPIKRTKTGTLSMF GYQIRINLNE GFPLLTTKYC HFKSIVYELL  
WFLRGDTNIS FLKNNNISIW NKWADKNGNL GPIYCKQWRA EDKKNNNTID QIEIALNKLK  
KEPSSRRILV SSWNVGELDL MSIPPCHVLF QLYVINNNKLS CQVYQRSCDI FGLGPFNIGS  
YALLTHIFAN QCDLLVEDLI WTGGDIHLYK NHLNQAKLQL TRSPLPLPKI FIKKKPKNLF  
NYAFNDPLLI DYNHHPKIKA PISI
```

- I've taken the liberty of submitting the sequence for you. The results are available at
<http://zhanglab.ccmb.med.umich.edu/I-TASSER/output/S516679/>

[\[Home\]](#) [\[Server\]](#) [\[Queue\]](#) [\[About\]](#) [\[Remove\]](#) [\[Statistics\]](#)

I-TASSER results for job id S516679

(Click on [S516679_results.tar.bz2](#) to download the tarball file including all modeling results listed on this page. Click on [Annotation of I-TASSER Output](#) to read the instructions for how to interpret the results on this page. Model results are kept on the server for 60 days, there is no way to retrieve the modeling data older than 2 months)

Submitted Sequence in FASTA format

>protein
MKEYLDLLNLILKNGYKPIDRTKTGTLMSFGYQIRINLNNEGFPPLTTKYCHFKSIVYELL
WFLRGDTNISFLKKNNISIWNWKADKGNGLGPPIYGKQWRATEDKKNNNTIDQIEIALNKLK
KEPSSRRILVSSNVGELDDMSIPCPVHLFQLYVINNKLSQCQVYQRSCDIFLGLPFINIGS
YALLTHIFANQCDLLEVEDLIWTGGDIHLYKHNHLNQAKLQLTRSPPLPLPKIFIKKKPNLF
NYAENDELLIDYNHHPKIKADIST

Predicted Secondary Structure

Predicted Solvent Accessibility

- Here are the Top 10 threading templates. Based on the sequence identity, coverage, and Normalized Z-score, is thymidylate synthase from W.g.b. an easy or hard target?

Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB	Iden1	Iden2	Cov	Norm.	Download	Z-score	Align.	
1	3qj7A	0.56	0.55	0.98	4.34	Download			
2	1an5A	0.62	0.62	1.00	5.42	Download			
3	3qj7A	0.56	0.55	0.98	5.67	Download			
4	1vza	0.51	0.52	1.00	2.28	Download			
5	1vza	0.51	0.52	1.00	1.67	Download			
6	3qj7A	0.56	0.55	0.98	5.07	Download			
7	1vza	0.51	0.52	1.00	2.51	Download			
8	6pf8A	0.44	0.45	1.00	6.47	Download			
9	3qj7A	0.56	0.55	0.98	5.04	Download			
10	1qzfA	0.43	0.45	1.00	12.11	Download			

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. ([more about the colors used](#))
 (b) Rank of templates represents the top ten threading templates used by I-TASSER.
 (c) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
 (d) Iden2 is the percentage sequence identity of the whole template chains with query protein.
 (e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
 (f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
 (g) Download Align provides the 3D structure of the aligned regions of the threading templates.
 (h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
 1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH1 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: PROSPECT2 10: PRC

Top 10 threading templates used by I-TASSER

(I-TASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. I-TASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB	Iden1	Iden2	Cov	Norm.	Download	Z-score	Align.	
1	3qj7A	0.56	0.55	0.98	4.34	Download			
2	1an5A	0.62	0.62	1.00	5.42	Download			
3	3qj7A	0.56	0.55	0.98	5.67	Download			
4	1vza	0.51	0.52	1.00	2.28	Download			
5	1vza	0.51	0.52	1.00	1.67	Download			
6	3qj7A	0.56	0.55	0.98	5.07	Download			
7	1vza	0.51	0.52	1.00	2.51	Download			
8	6pf8A	0.44	0.45	1.00	6.47	Download			
9	3qj7A	0.56	0.55	0.98	5.04	Download			
10	1qzfA	0.43	0.45	1.00	12.11	Download			

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. ([more about the colors used](#))
 (b) Rank of templates represents the top ten threading templates used by I-TASSER.
 (c) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
 (d) Iden2 is the percentage sequence identity of the whole template chains with query protein.
 (e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
 (f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
 (g) Download Align provides the 3D structure of the aligned regions of the threading templates.
 (h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
 1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH1 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: PROSPECT2 10: PRC

- Here are the Top 10 threading templates. Based on the sequence identity, coverage, and Normalized Z-score, is thymidylate synthase from W.g.b. an easy or hard target?
- The sequence identity is not very high but in a reasonable range for homology modeling. The coverage is very high and Normalized Z-score is also high, so this not a hard target for iTASSER.

Top 10 threading templates used by iTASSER

(iTASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. iTASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB	Iden1	Iden2	Cov	Norm.	Download	Z-score	Align.
1	3qj7A	0.56	0.55	0.98	4.34	Download		
2	1an5A	0.62	0.62	1.00	5.42	Download		
3	3qj7A	0.56	0.55	0.98	5.67	Download		
4	1vza	0.51	0.52	1.00	2.28	Download		
5	1vza	0.51	0.52	1.00	1.67	Download		
6	3qj7A	0.56	0.55	0.98	5.07	Download		
7	1vza	0.51	0.52	1.00	2.51	Download		
8	6pf8A	0.44	0.45	1.00	6.47	Download		
9	3qj7A	0.56	0.55	0.98	5.04	Download		
10	1qzfA	0.43	0.45	1.00	12.11	Download		

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. ([more about the colors used](#))
 (b) Rank of templates represents the top ten threading templates used by iTASSER.
 (c) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
 (d) Iden2 is the percentage sequence identity of the whole template chains with query protein.
 (e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
 (f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
 (g) Download Align provides the 3D structure of the aligned regions of the threading templates.
 (h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
 1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH1 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: PROSPECT2 10: PRC

Top 10 threading templates used by iTASSER

(iTASSER modeling starts from the structure templates identified by LOMETS from the PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens of thousands of template alignments. iTASSER only uses the templates of the highest significance in the threading alignments, the significance of which are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. The templates in this section are the 10 best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading program, where the threading programs are sorted by the average performance in the large-scale benchmark test experiments.)

Rank	PDB	Iden1	Iden2	Cov	Norm.	Download	Z-score	Align.
1	3qj7A	0.56	0.55	0.98	4.34	Download		
2	1an5A	0.62	0.62	1.00	5.42	Download		
3	3qj7A	0.56	0.55	0.98	5.67	Download		
4	1vza	0.51	0.52	1.00	2.28	Download		
5	1vza	0.51	0.52	1.00	1.67	Download		
6	3qj7A	0.56	0.55	0.98	5.07	Download		
7	1vza	0.51	0.52	1.00	2.51	Download		
8	6pf8A	0.44	0.45	1.00	6.47	Download		
9	3qj7A	0.56	0.55	0.98	5.04	Download		
10	1qzfA	0.43	0.45	1.00	12.11	Download		

(a) All the residues are colored in black; however, those residues in template which are identical to the residue in the query sequence are highlighted in color. Coloring scheme is based on the property of amino acids, where polar are brightly coloured while non-polar residues are colored in dark shade. ([more about the colors used](#))
 (b) Rank of templates represents the top ten threading templates used by iTASSER.
 (c) Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence.
 (d) Iden2 is the percentage sequence identity of the whole template chains with query protein.
 (e) Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein.
 (f) Norm. Z-score is the normalized Z-score of the threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.
 (g) Download Align provides the 3D structure of the aligned regions of the threading templates.
 (h) The top 10 alignments reported above (in order of their ranking) are from the following threading programs:
 1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: HHSEARCH1 6: Neff-PPAS 7: HHSEARCH 8: pGenTHREADER 9: PROSPECT2 10: PRC

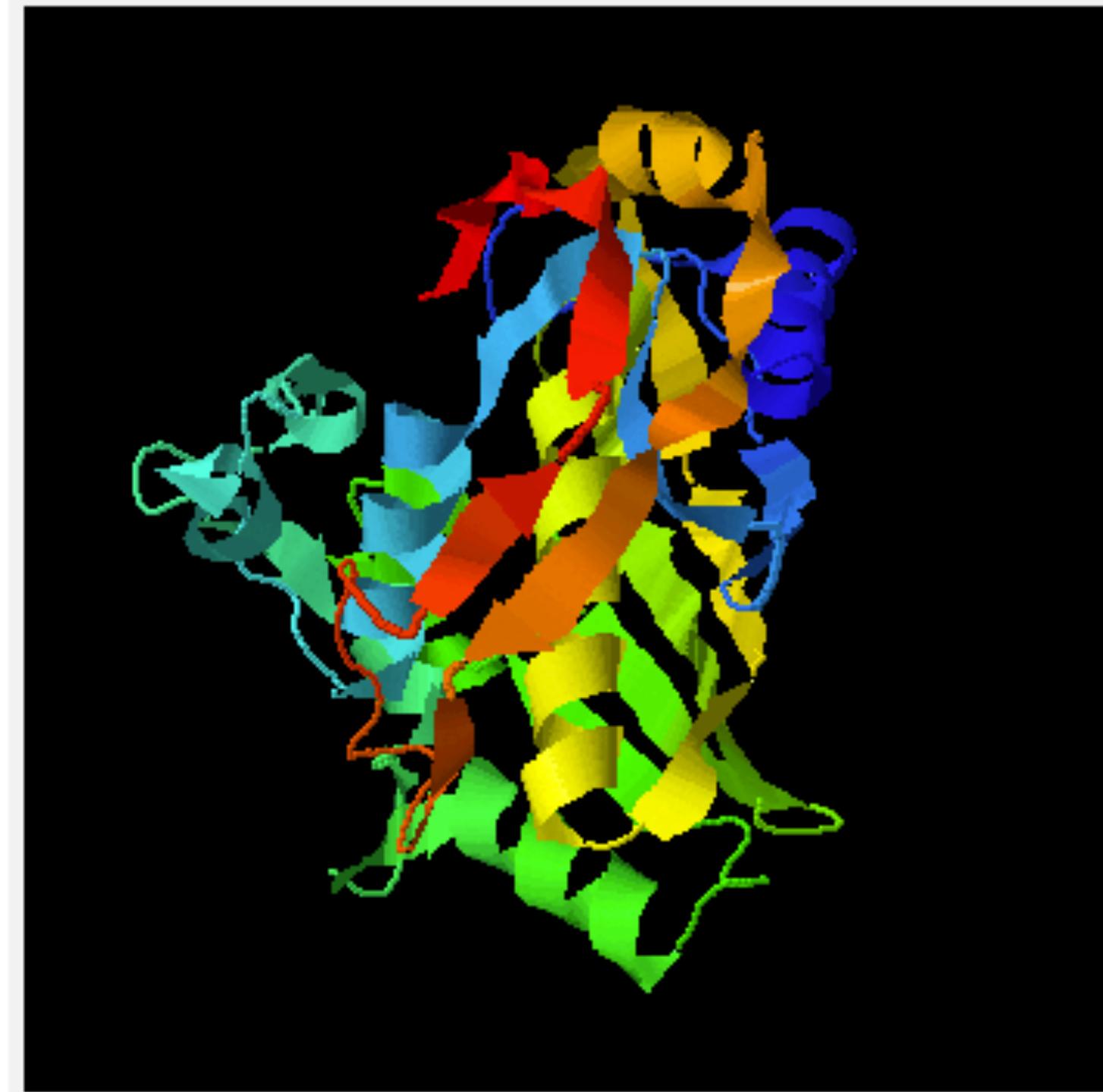
- The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length...
- Is I-TASSER confident about its final model?

Top 5 final models predicted by I-TASSER

(For each target, I-TASSER simulations generate a large ensemble of structural models. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length...)

- [More about C-score](#)
- [Local structure accuracy profile of the top five models](#)

(By right-click on the images, you can export image file or change the orientation)



Spin On/Off

- [Download Model 1](#)
- C-score=1.91 ([Read more about C-score](#))
- Estimated TM-score = 0.99±0.04
- Estimated RMSD = 2.2±1.7 Å

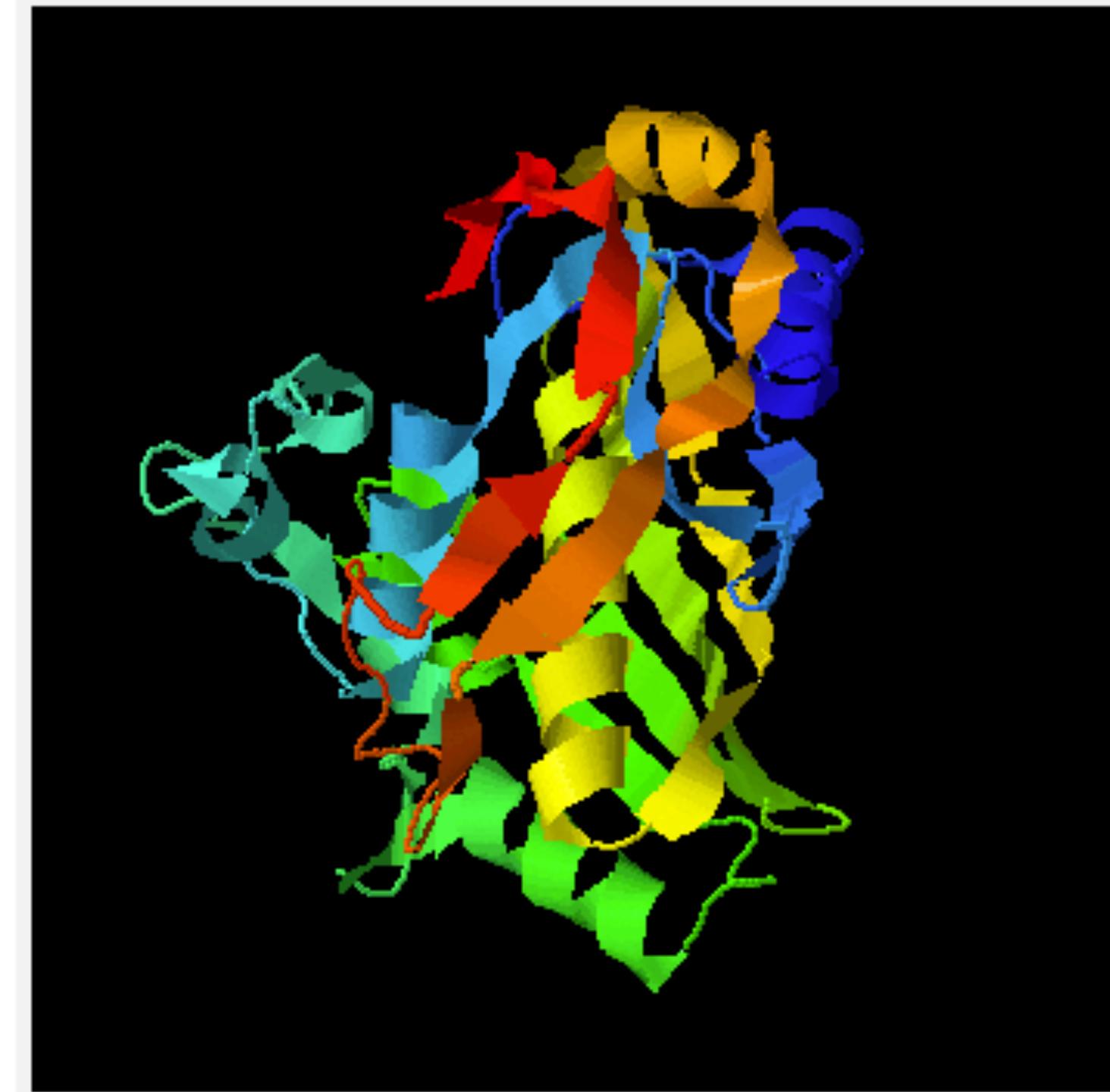
- The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length...
- Is I-TASSER confident about its final model? Yes.

Top 5 final models predicted by I-TASSER

(For each target, I-TASSER simulations generate a large ensemble of structural models. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length...)

- [More about C-score](#)
- [Local structure accuracy profile of the top five models](#)

(By right-click on the images, you can export image file or change the orientation)



Spin On/Off

- [Download Model 1](#)
- C-score=1.91 ([Read more about C-score](#))
- Estimated TM-score = 0.99 ± 0.04
- Estimated RMSD = $2.2 \pm 1.7 \text{ \AA}$

Structure Preparation

- None of the structures were ready for electrostatic potential calculation
- The human structure, 1HVY
 - has 4 chains, but we are only interested in 2
 - I used MultiSeq to write separate files for each chain, merged the files for chains A and B, and then modified “END” after chain A into “TER”
- The E. coli structure, 6NNR
 - is not aligned with the human structure
 - I used MultiSeq to write separate files for each chain, used MultiSeq STAMP structural alignment to superpose E. coli chains A and B on top of the respective human chains, wrote the aligned files, and merged and modified them as with 1HVY
- The W.g.b. homology model
 - is not aligned with the human structure
 - I used MultiSeq STAMP structural alignment to superpose W.g.b. chains A and B on top of the respective human chains, wrote the aligned files, and merged and modified them as with 1HVY

Electrostatics Calculations

- I ran electrostatic potential calculations with PDB2PQR and APBS (<http://www.poissonboltzmann.org>) on the PDB2PQR server (http://nbcr-222.ucsd.edu/pdb2pqr_2.1.1/). There were three steps:
 - Uploading the PDB file and selecting parameters. I used the defaults.
 - Running PDB2PQR by pressing “Submit”. This performs some basic structural preparation, e.g. adding missing atoms and optimizing hydrogen positions.
 - Running APBS by following “Click here to run APBS with your results”. This actually solves the PB equation.
- For TS, none of these calculations took a very long time.

PDB2PQR Server

Currently using PDB2PQR Version 2.1.1

Return to [the PDB2PQR homepage](#).

This server enables a user to convert PDB files into PQR files. PQR files are PDB files where the occupancy and B-factor columns have been replaced by per-atom charge and radius. pKa calculations are performed by PROPKA.

For more information on PDB2PQR please see the:

- [Home Page](#)
- [Register \(and help support PDB2PQR & APBS\)](#)
- [User Guide](#)
- [Examples](#)
- [Release Notes](#)

If you use the PDB2PQR service in a publication, please cite:

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* 32 W665-W667 (2004). [[Link](#)]

Note: In order to distribute server load, the PDB2PQR server currently is limited to a maximum size of 10000 atoms per protein. If you are interested in using PDB2PQR for larger proteins, you are encouraged to download a command line version of PDB2PQR from the [PDB2PQR download page](#). For additional limitations, please see the [PDB2PQR user guide](#).

Note: This server uses automatic refreshing to update the status of your PDB2PQR submission. Do not use the *back* button on your browser while the server is running.

Please enter either:

a PDB ID:
 upload a PDB file:

Status: complete ✓
Run time: 0:00:34
Current time: Mon Jan 20 13:52:26 2020

Here are the results:

- Input files
 - [TS_human.pdb](#)
- Output files
 - [TS_human.propka](#)
 - [TS_human.pqr](#)
 - [TS_human.in](#)
- Runtime and debugging information
 - [Program output \(stdout\)](#)
 - [Program errors and warnings \(stderr\)](#)

[Click here](#) to run APBS with your results.

Status: complete ✓
Run time: 0:01:35
Current time: Mon Jan 20 13:57:47 2020

Here are the results:

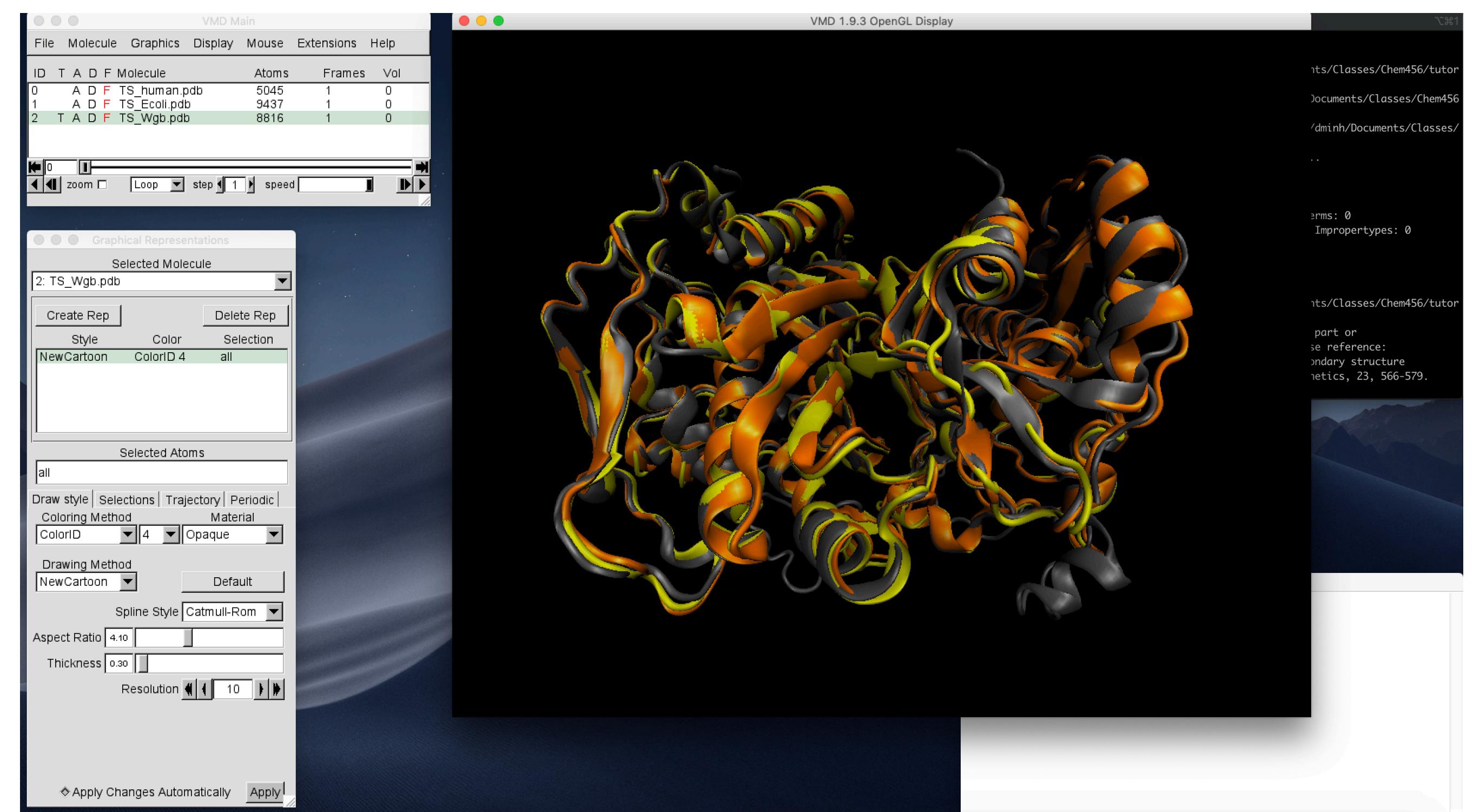
- Input files
 - [15795571126.pqr](#)
 - [apbsinput.in](#)
- Output files
 - [15795571126-pot-PE0.dx.gz](#)
 - [15795571126.cube.gz](#)
- Runtime and debugging information
 - [Program output \(stdout\)](#)
 - [Program errors and warnings \(stderr\)](#)

Visualize your results online:

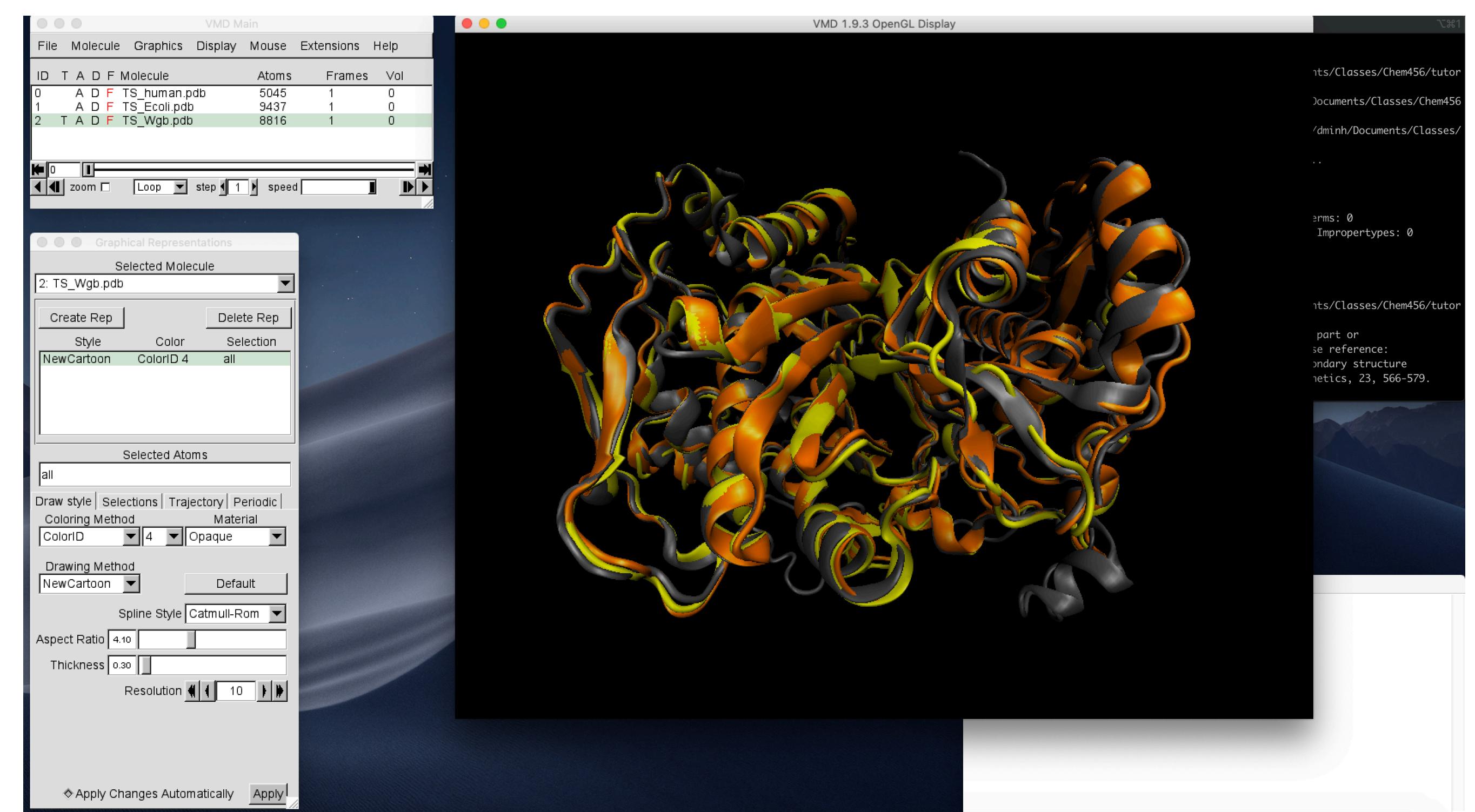
- [3Dmol](#)
- [Jmol](#)

Visualizing Electrostatic Potentials with VMD

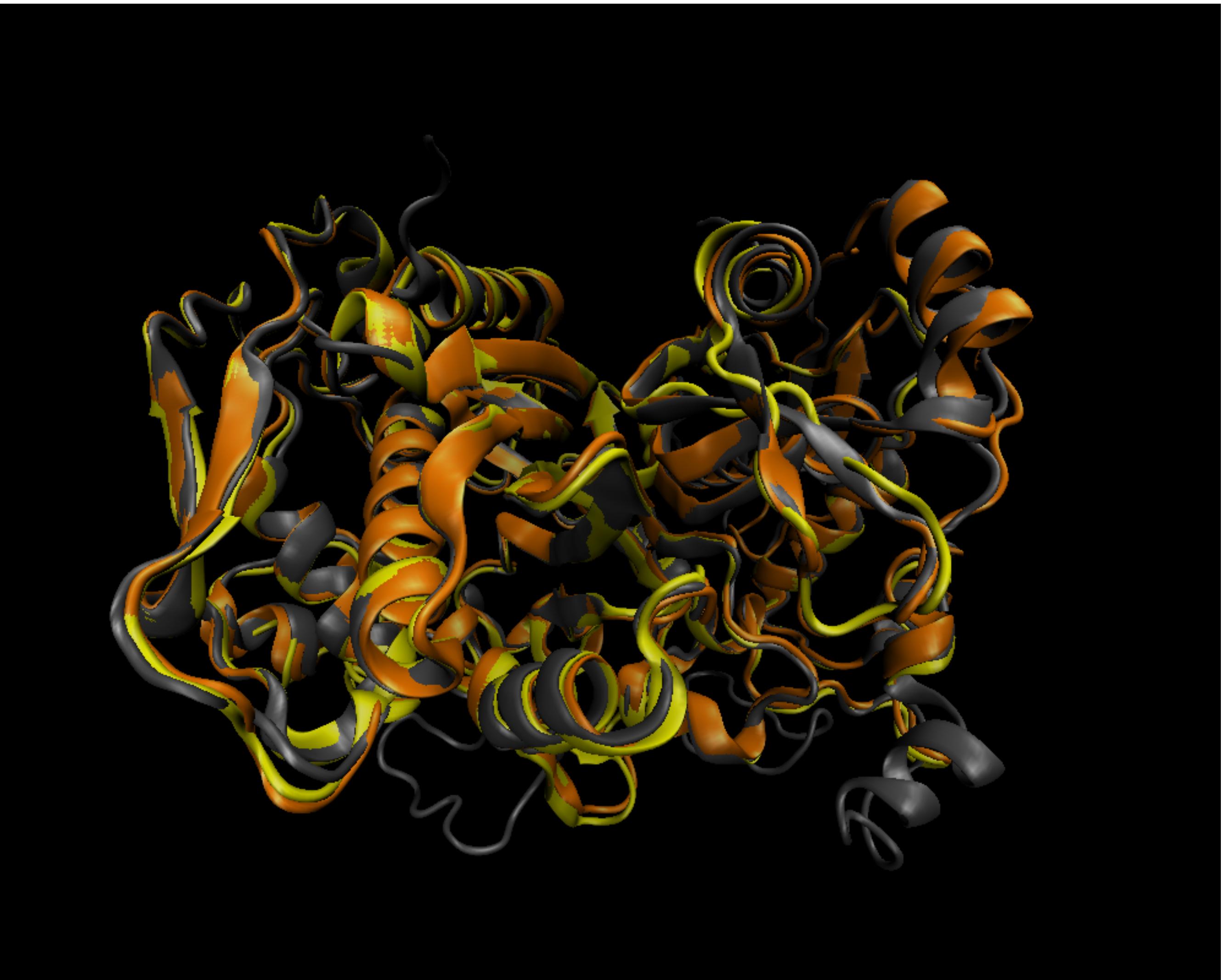
- Load the models of TS from human, E. coli, and W.g.b. The files are available on [github](#).
- Use the “NewCartoon” representation and color human gray (2), E. coli orange (3), and W.g.b. yellow (4)
- Your display should look something like the image to the right
- Which of the enzymes is least like the others?



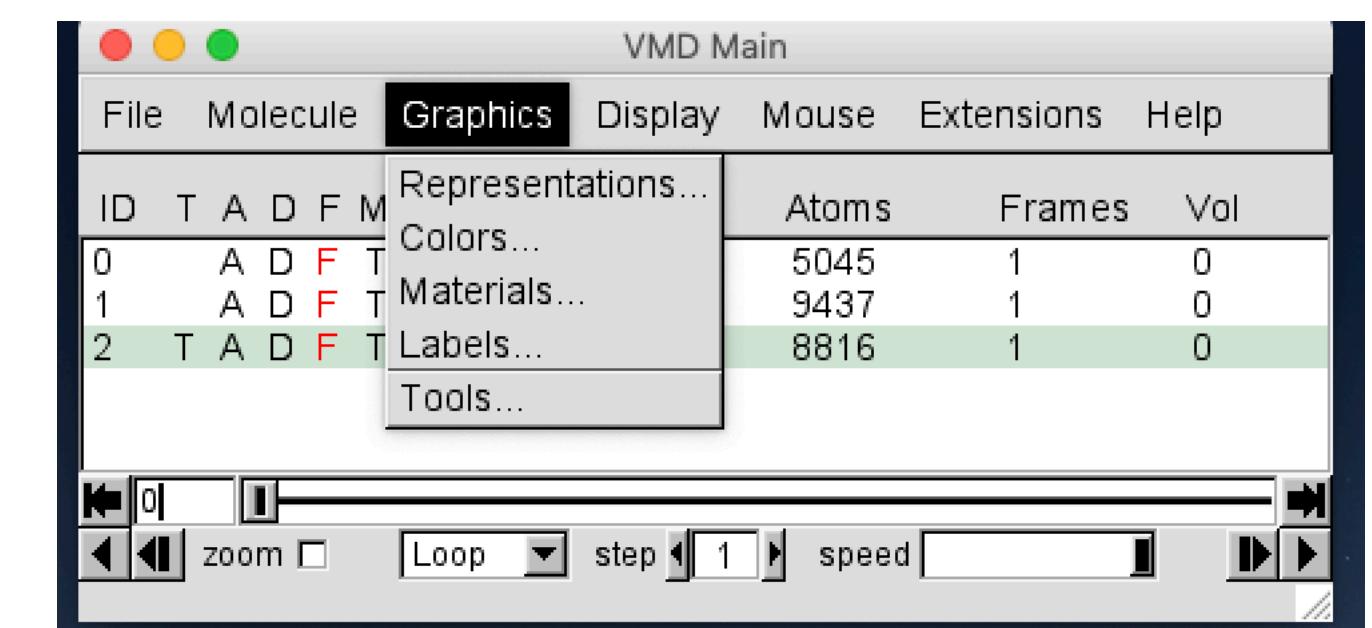
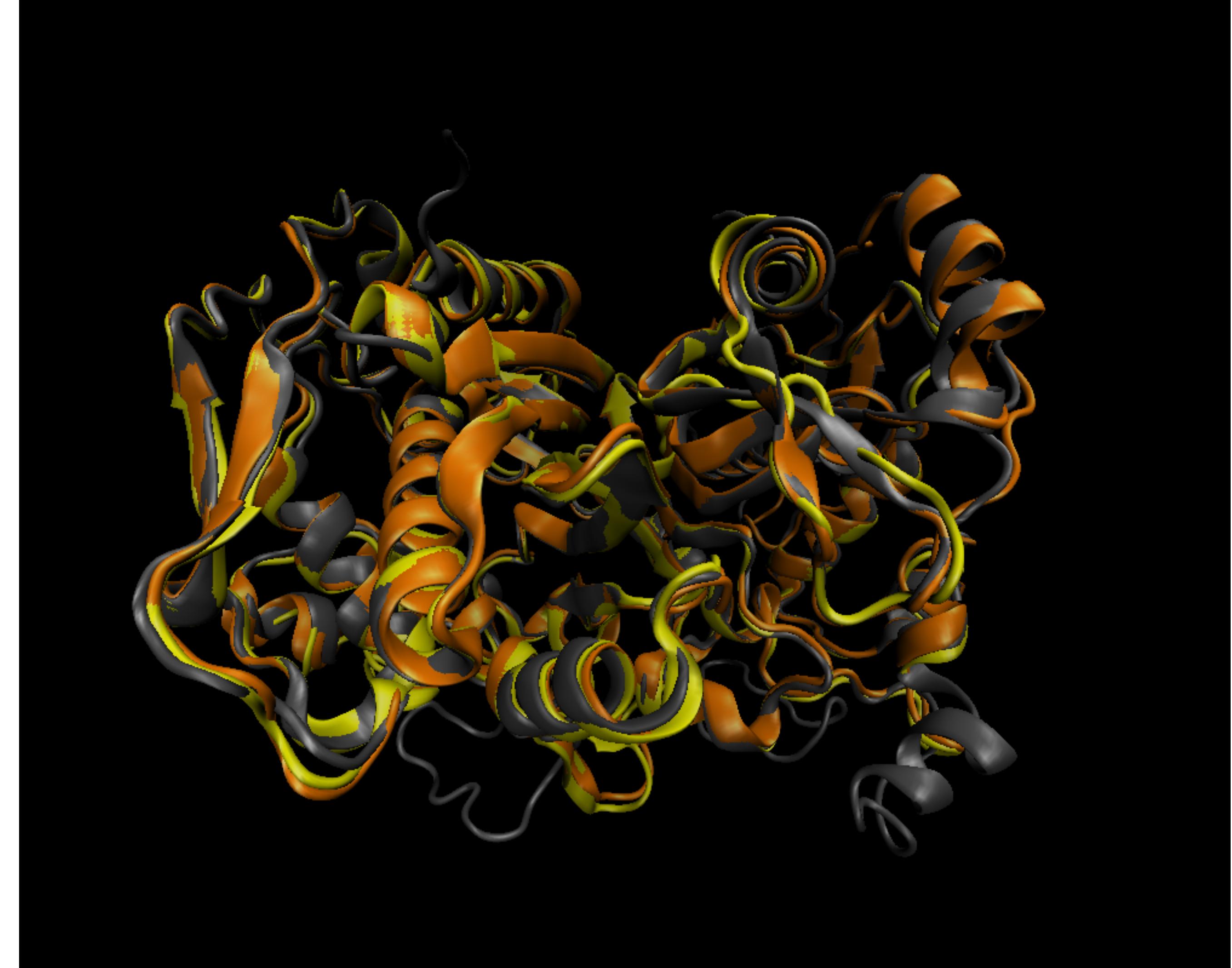
- Load the models of TS from human, E. coli, and W.g.b.
- Use the “NewCartoon” representation and color human gray (2), E. coli orange (3), and W.g.b. yellow (4)
- Your display should look something like the image to the right
- Which of the enzymes is least like the others?
- In the human structure, a short helix is extended and a coil added. A sheet is replaced by an extended loop.



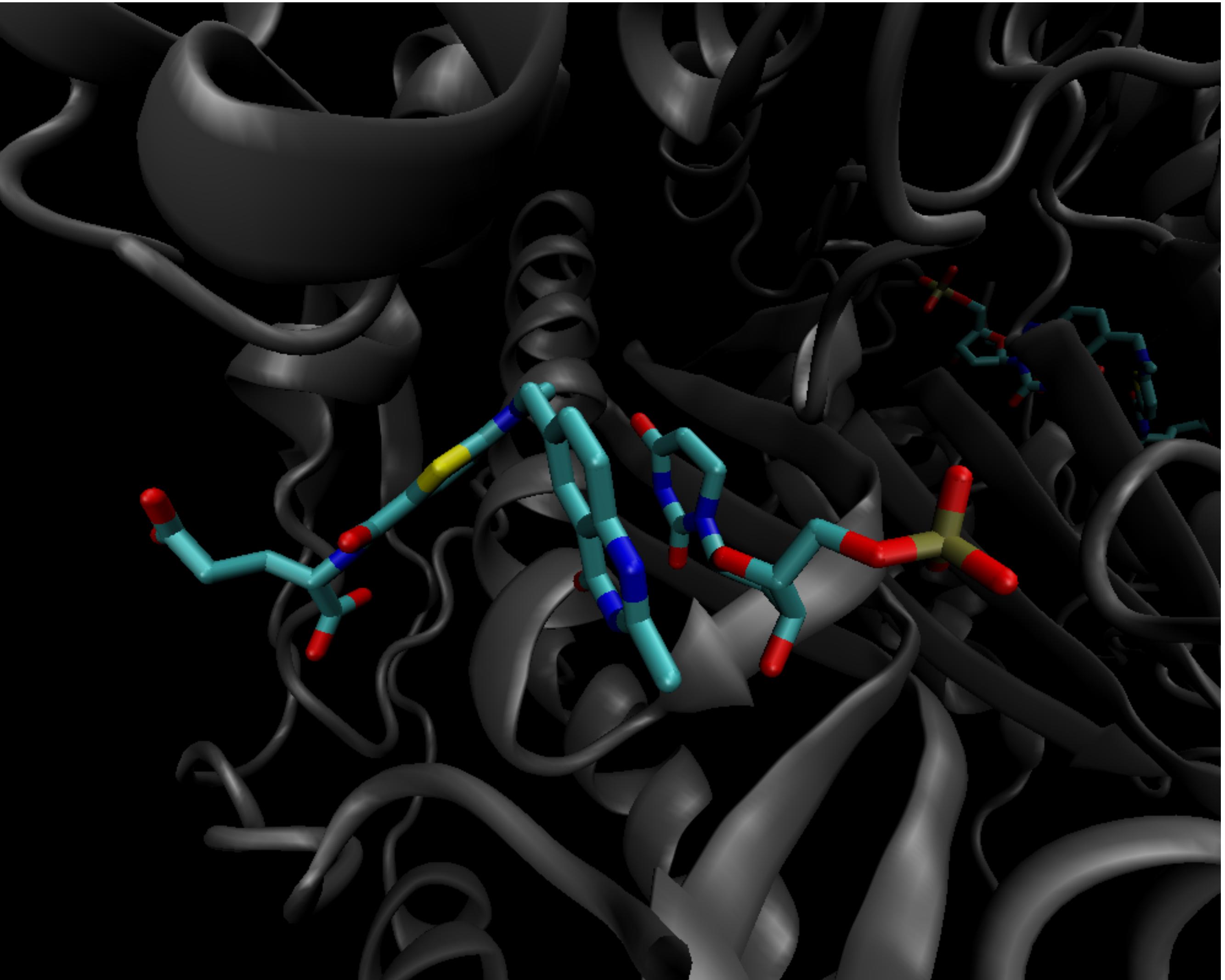
- Press 1 and then click to label specific atoms. Where (what residue numbers) are the unique parts of the human structure?



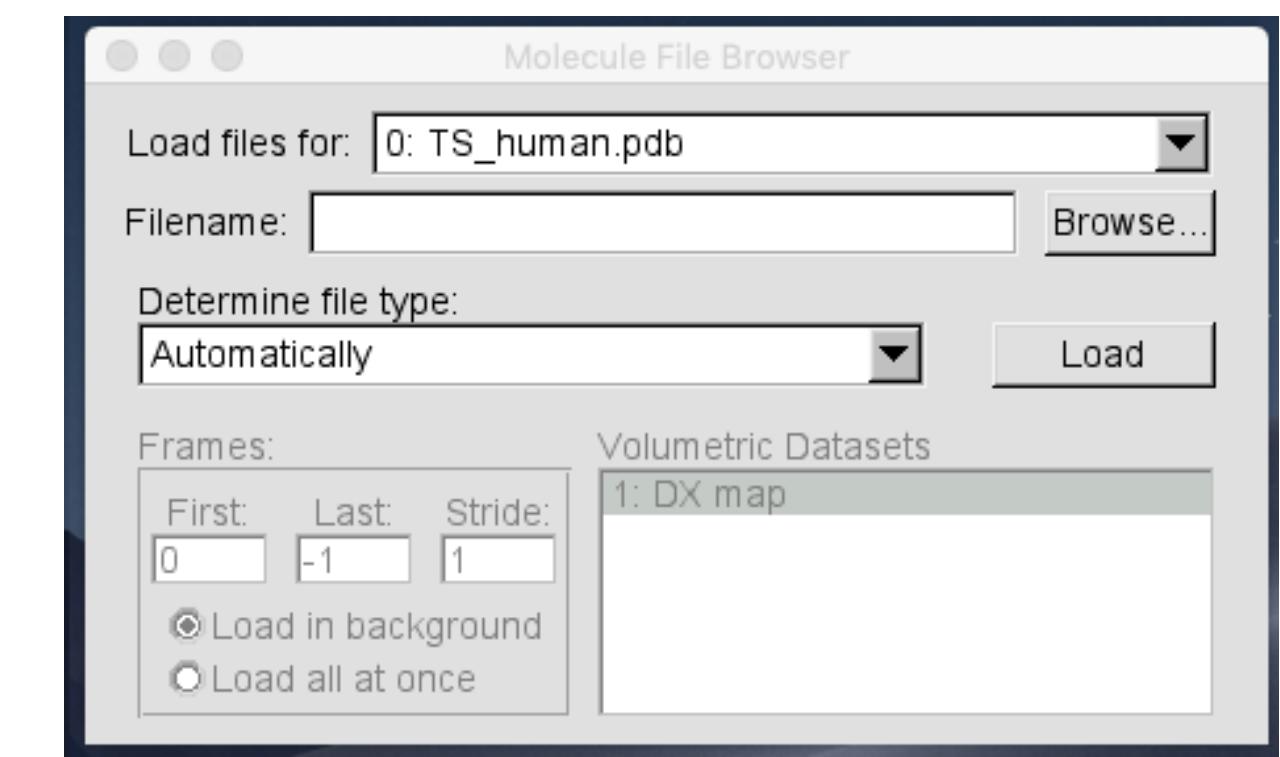
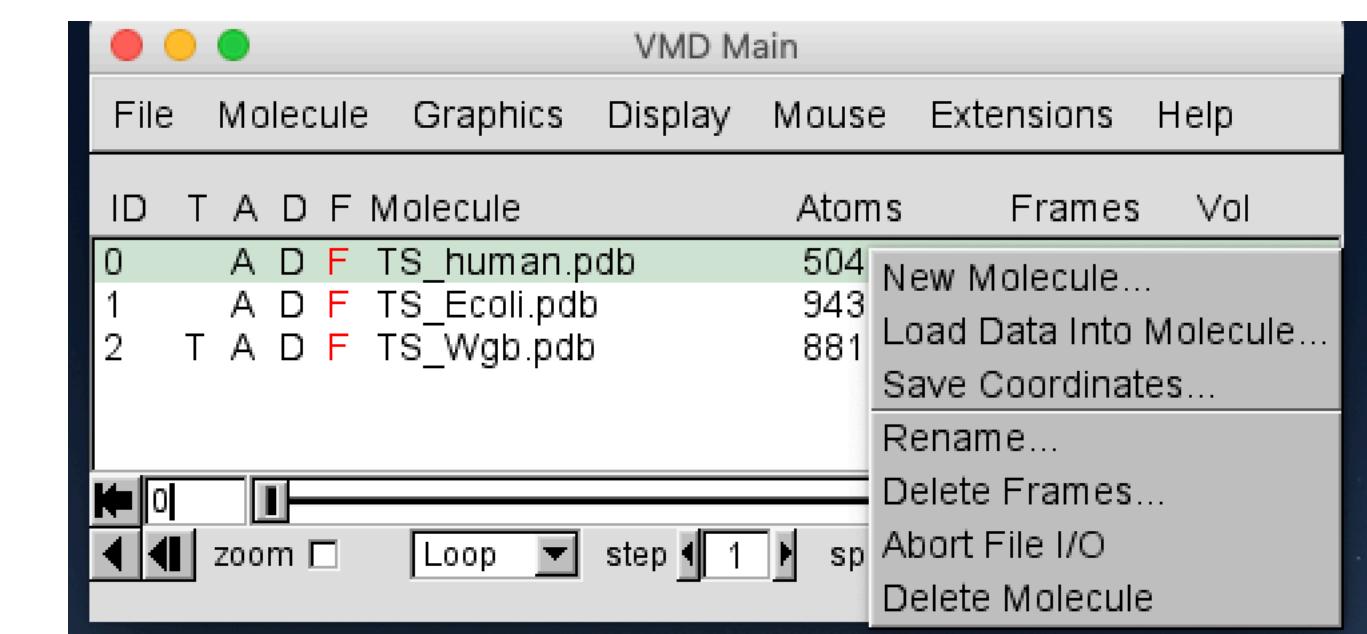
- Press 1 and then click to label specific atoms. Where (what residue numbers) are the unique parts of the human structure?
 - 114 to 125
 - 142 to 157
- You can remove the labels by going to the “Graphics” menu under “Labels”, selecting labels, and deleting them.



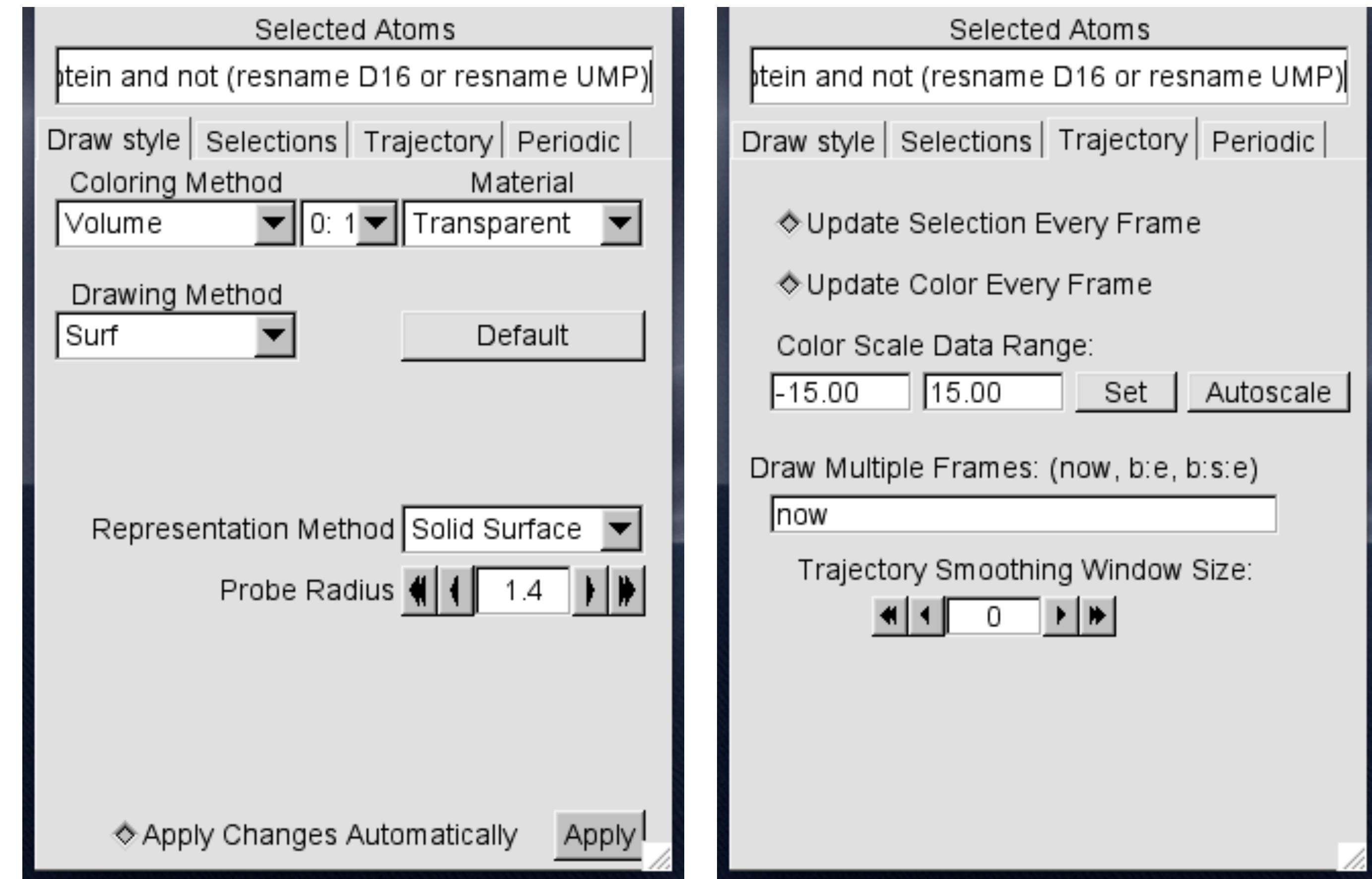
- The location of the active site is a lot more evident if you create a representation with Selected Atoms as “resname D16 or resname UMP”, Color Method “Element”, and Drawing Method “Licorice”



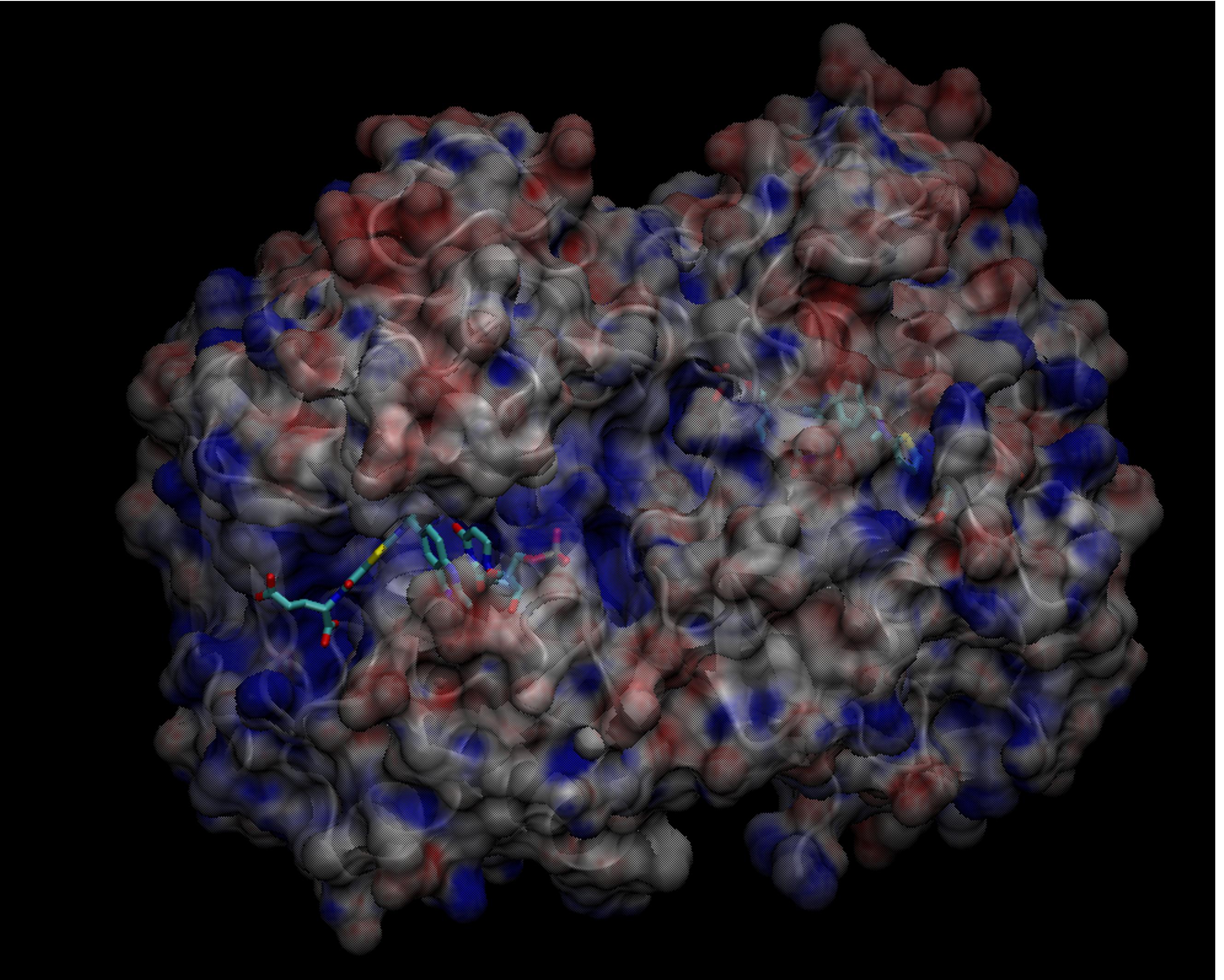
- Now let's actually load the electrostatic potential
- In the “VMD Main” window, right click on a molecule and select “Load Data Into Molecule”
- Browse to the .dx file in the same directory as the molecule of interest and load it
- It won't look like anything has happened except there will be a line under “Volumetric Datasets”
- Let's focus on human and hide the other structures by double-clicking on D in the VMD Main window



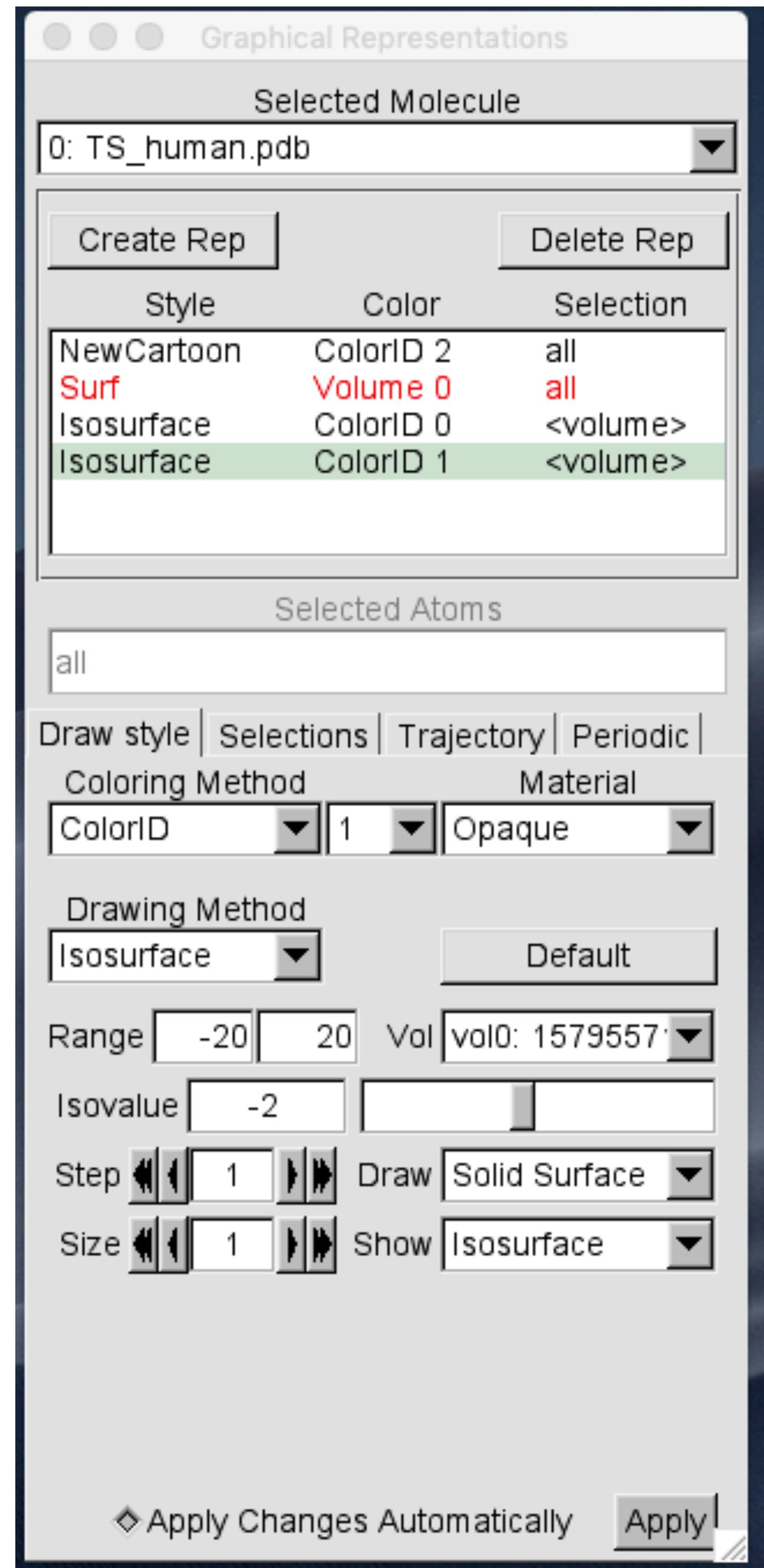
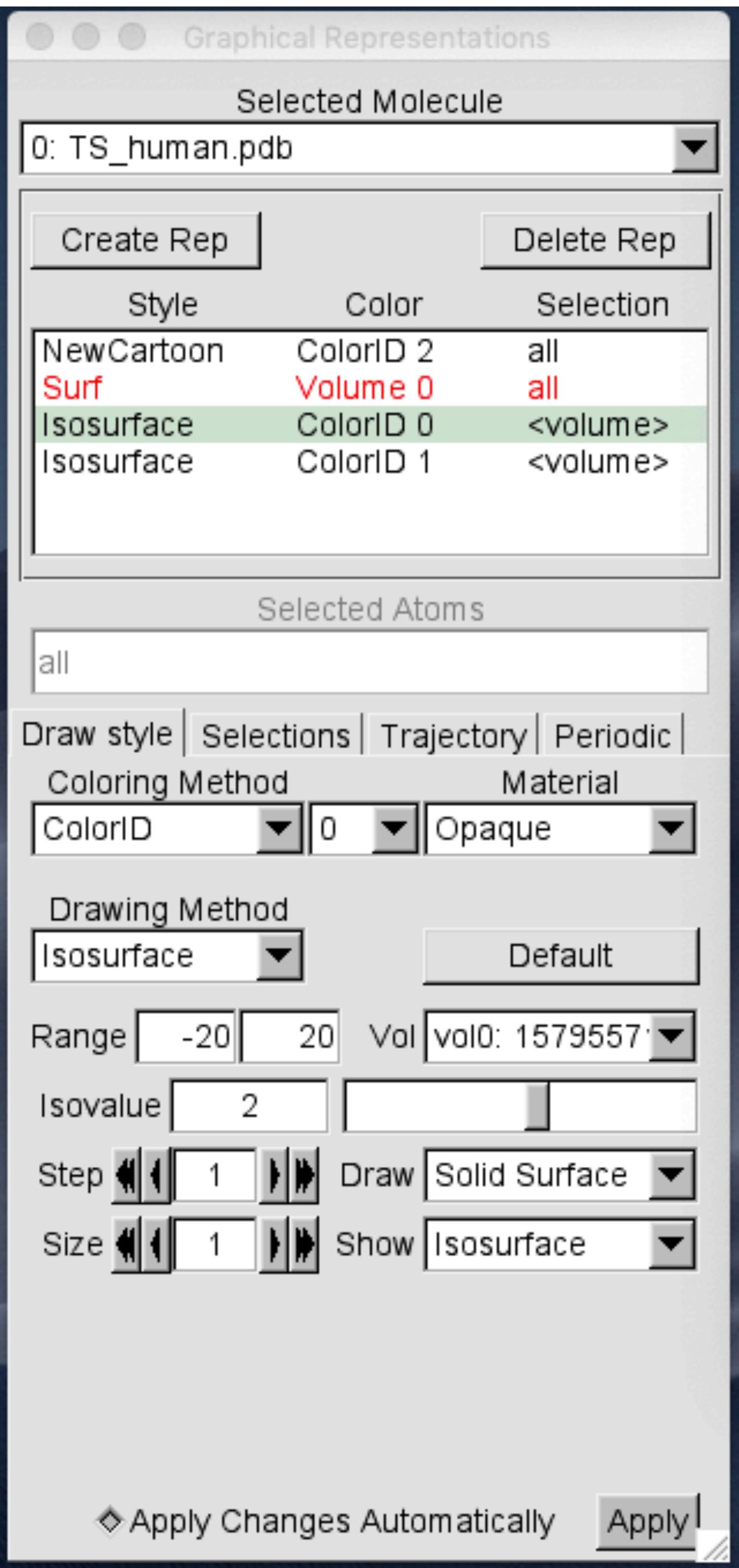
- In VMD, volumetric data can be displayed
 - by Isosurface, VolumeSlice, or FieldLines representations
 - Using the “Volume Coloring Method”
 - To color a surface based on the electrostatic potential, create a new representation selecting “protein and not (resname D16 or rename UMP)”, set the Drawing Method to “Surf”, Coloring Method to “Volume”, and Material to “Transparent”
 - Also click on the “Trajectory” tab and set Color Scale Data Range from -15 to 15. Otherwise everything will look white.



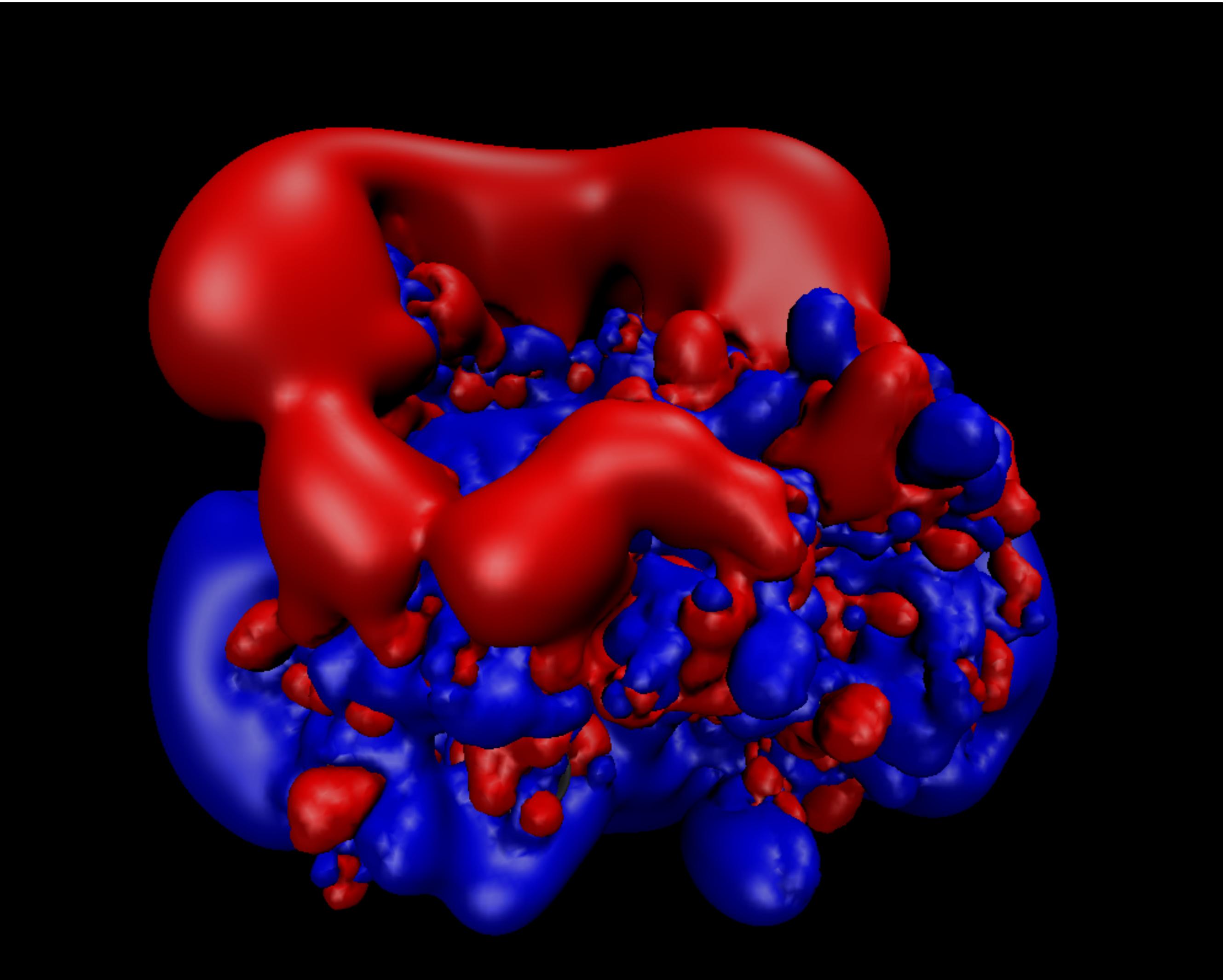
- On the surface, positive is blue and negative is red.
- Notice that there is a positive charge in the entire binding site cleft



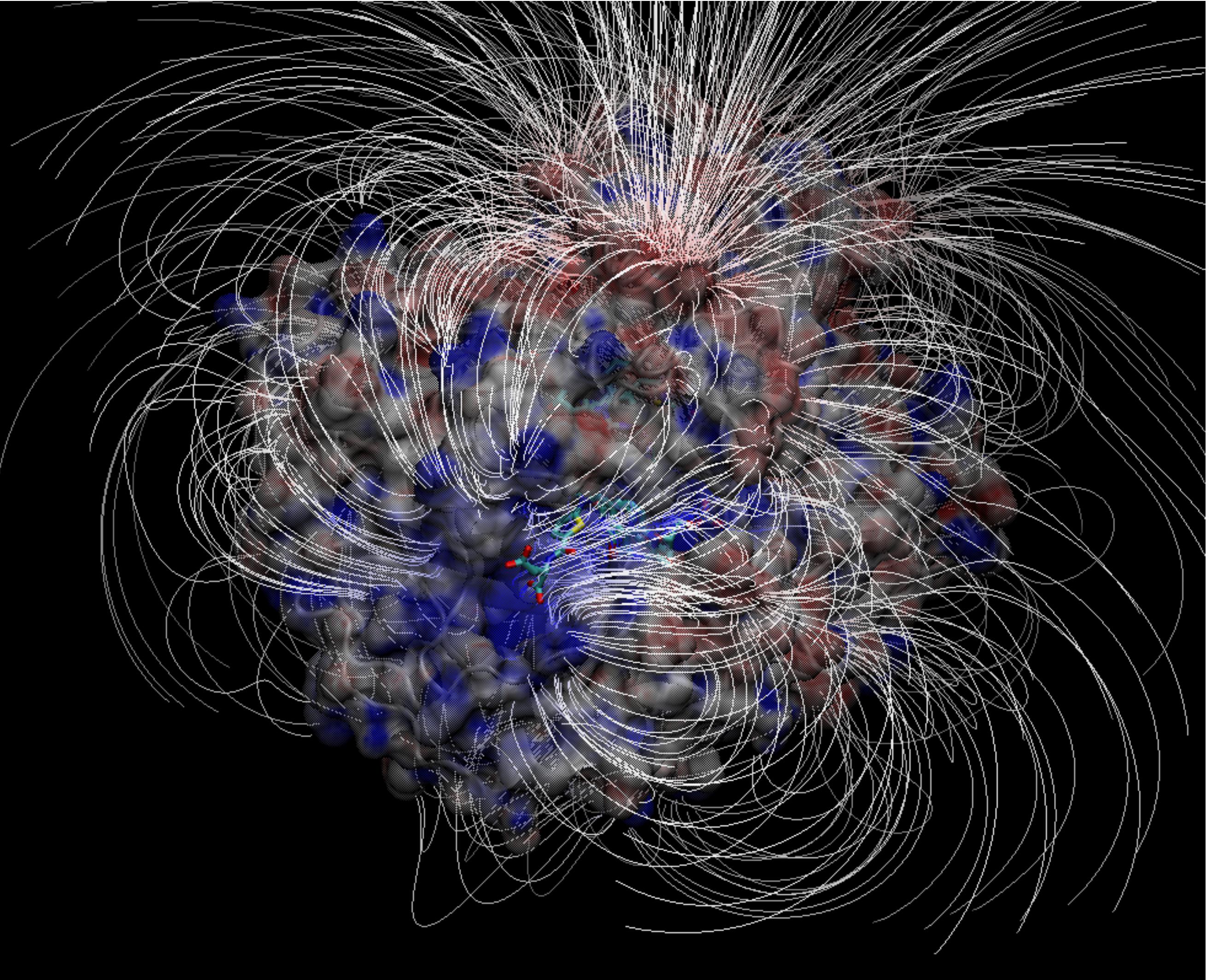
- Now let's try the isosurface representation. In this representation all points along the surface have the same electrostatic potential.
- Create two isosurfaces
 - a blue isosurface with isovalue 2
 - a red isosurface with isovalue -2
- Hide the solvent-accessible surface by double-clicking on "Surf"



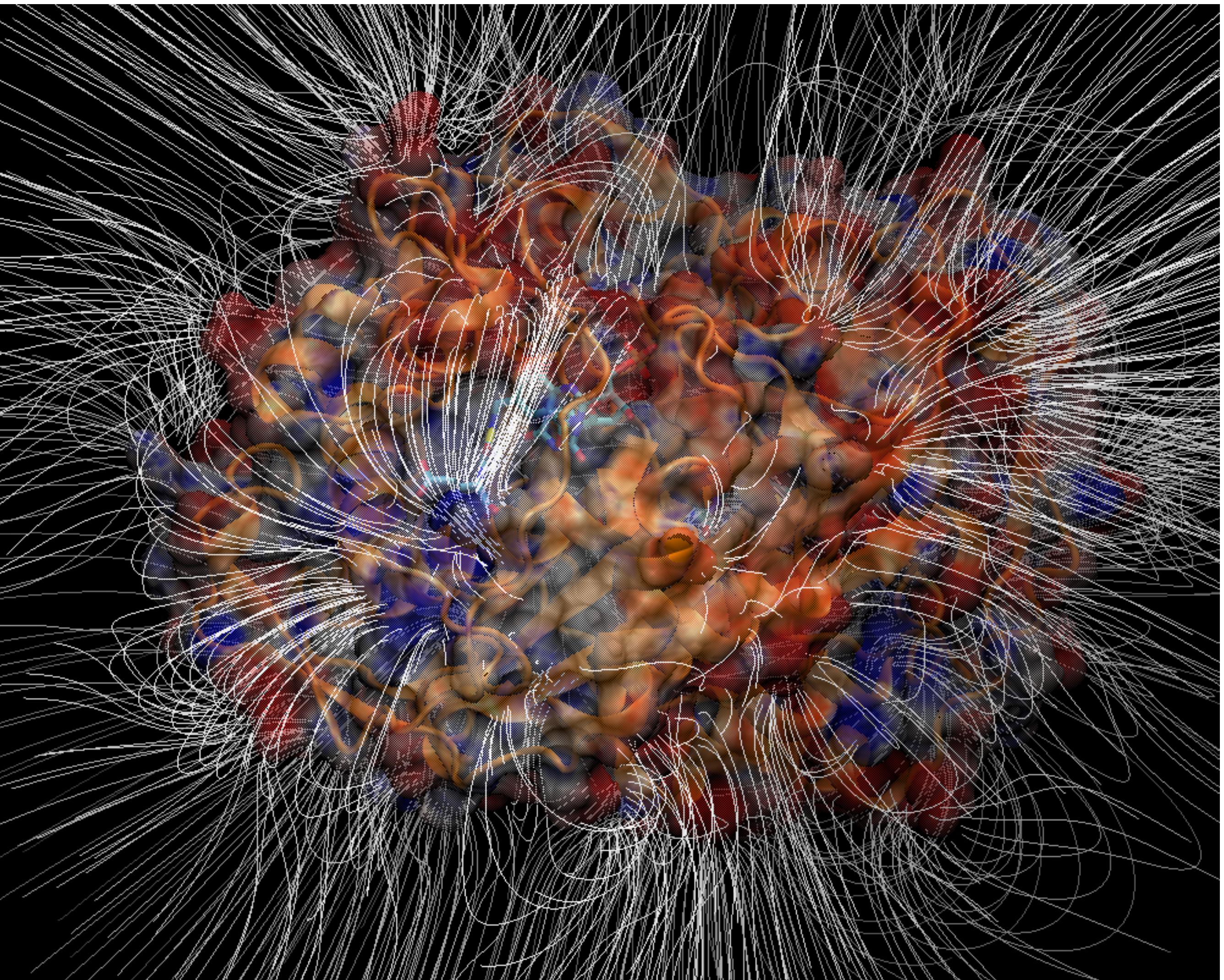
- The isosurfaces are a reasonable reproduction of Fig.1 from Garg *et al.* (2015).
- However, the charges here seem to have a larger magnitude than in the paper. This may be because of charge assignment or including the ligands in the PQR file.



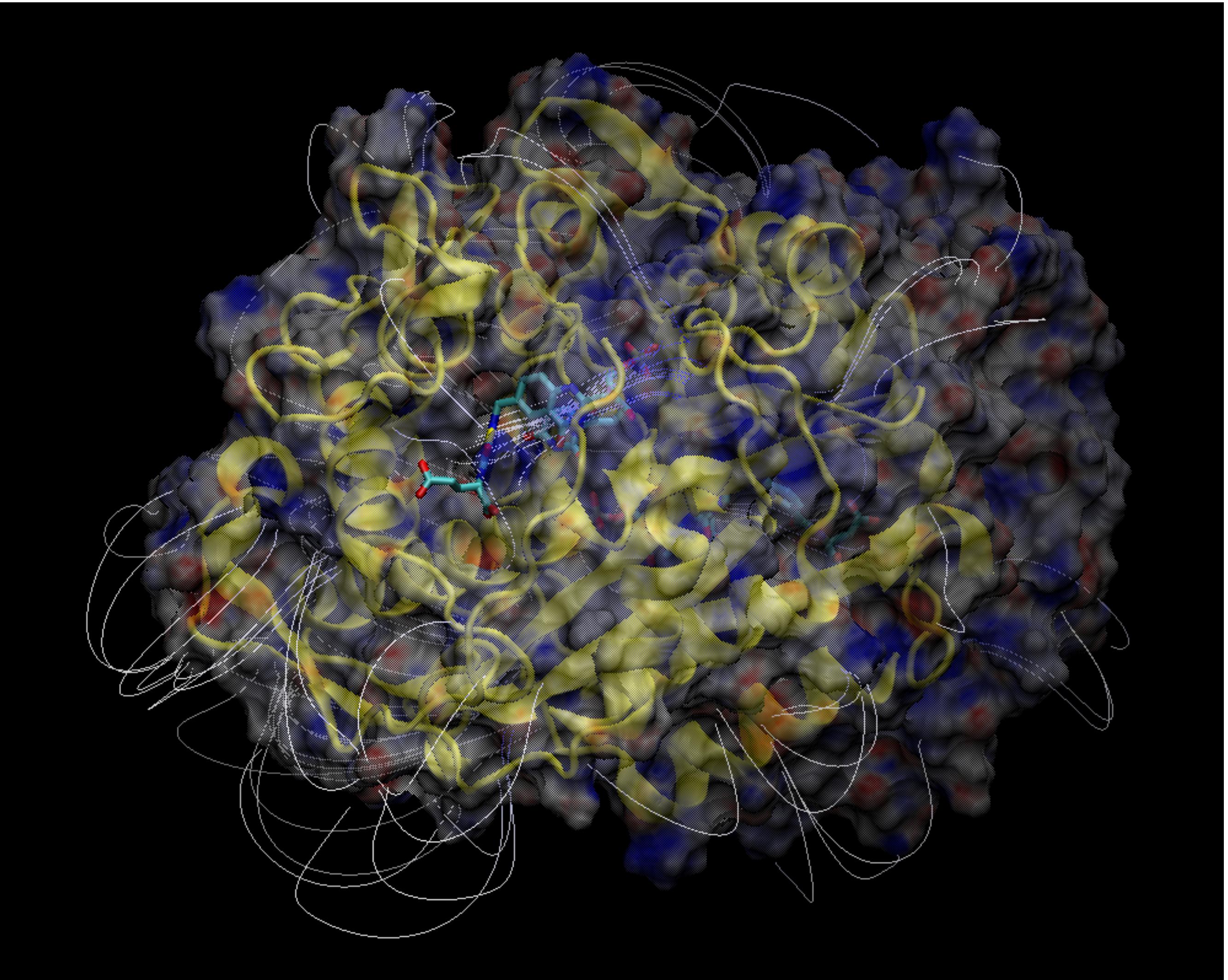
- Electric field lines show the path that a massless positive charge would take. The density of lines shows the intensity of the electric field.
- There is clearly a major electrostatic sink pulling negative ligands into the active site.



- The E. Coli enzyme has a similar positive patch and sink



- As a validation of Garg et. al. (2015), the W.g.b. electrostatic potential is much less pronounced.



References

- Garg, D.; Skouloubris, S.; Briffotaux, J.; Myllykallio, H.; Wade, R. C. Conservation and Role of Electrostatics in Thymidylate Synthase. *Sci Rep* 2015, 5 (1), 17356. <https://doi.org/10.1038/srep17356>, adapted under the CC BY 4.0 license.