Designing an ID2-like inhibitor of DNA binding *in silico*

By Joshua Sodicoff

ID2 Protein

- Found on chromosome 2
- 134 residues long
- A helix-loop-helix protein
- One of 4 proteins in the Inhibitor of DNA Binding family
- All are involved in the regulation of
  - Growth
  - Differentiation
  - Apoptosis
  - Angiogenesis
  - Tumorigenesis
- ID2 also plays a role in the circadian rhythm

A Divalent Ion is Crucial in the Structure and Dominant-Negative Function of Id Proteins, a Class of Helix-Loop-Helix Transcription Regulators. (2012)

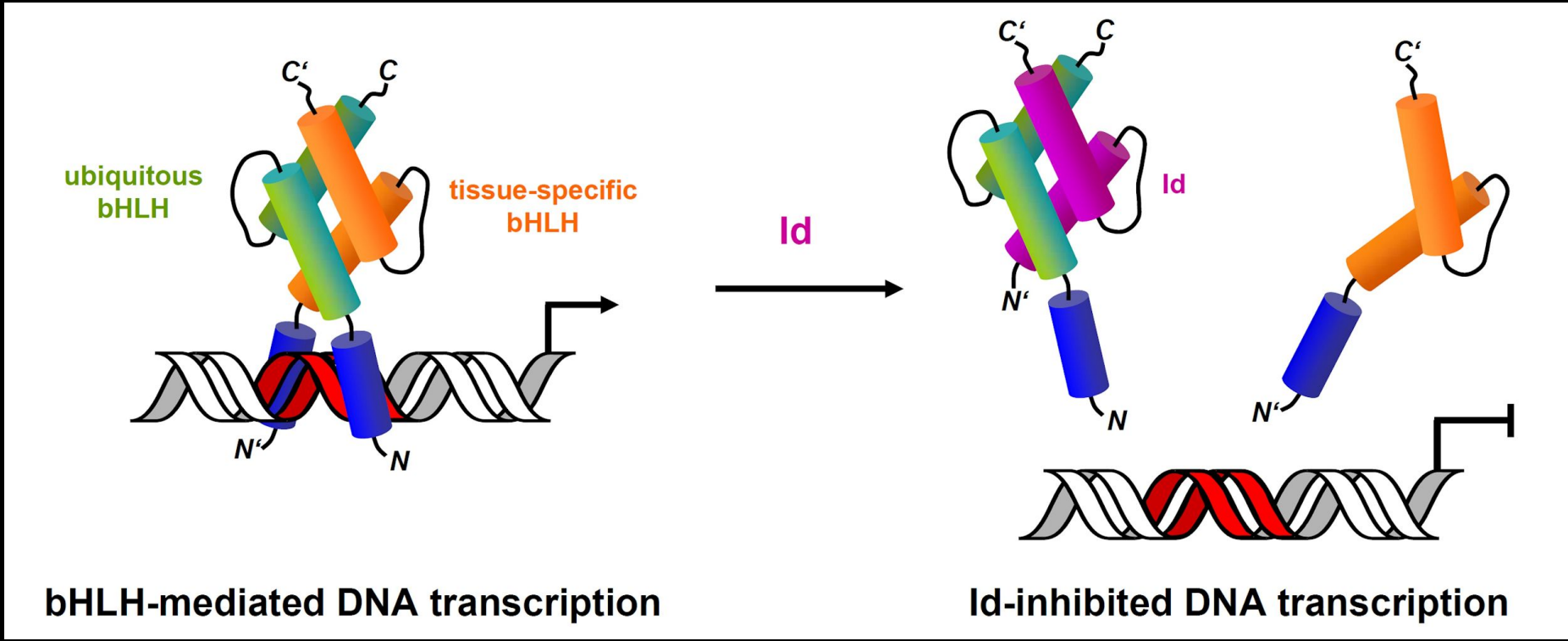
Interactions

- Interacts with basic-helix-loop-helix factors
  - Two types
    - Class I - Ubiquitous
    - Class II - Tissue specific
  - Class I can homodimerize or heterodimerize with Class II
  - Each contain a basic region of residues which interact with DNA
    - Unique E box sequences, the series of nucleotides onto which the factor binds
    - Different combinations of bHLH factors allow for binding to a variety of regions

Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. (2008)

Interactions cont.d

- Also able to dimerize with Class I factors and Class II myogenic regulatory factors
- However, because it lacks the basic region and thus does not allow the dimer to bind to DNA
- Thus, is an antagonist of transcription factor function
- The end result is a decrease in the expression of proteins involved in differentiation


ubiquitous
bHLH
tissue-specific
bHLH
C'
C
N'
N
Id
Id
C'
C
N'
N
C'
N'
bHLH-mediated DNA transcription
Id-inhibited DNA transcription

Why design?

- Medical relevance
  - Research being done into ID gene methylation as a possible cause of prostate cancer, breast cancer, leukemia among other cancers
  - Therapeutics (either in the form of the protein or gene therapy) with the optimized structure could be critical in fighting these ailments
- Size
  - At 134 residues, EvoDesign and other programs should be able to develop possible new designs for similar proteins
- Known structural features
  - The exact residues that engage in dimerization are known and thus can be held constant to ensure that a redesign binds in the same way
  - The structure has also been experimentally solved, allowing for use of backbone
- Function
  - ID2 has the greatest binding affinity with transcription factors of ID proteins with solved structures

Design rationale

- In order to act in cases in which naturally produced ID proteins are either less expressed or lose function, an ID-like protein could be an effective therapeutic
- ID2 is already relatively simple and small
- Using ID2 as a backbone in EvoDesign while keeping the residues that are involved in dimerization constant seems to be a good first step
- Important to remember that goal is not complete competitive inhibition of transcription factors, but rather a similar level of activity to human ID2

EvoDesign results

Design Number	Sequence Identity (%)[?]	Normalized Relative Error [?]			
		Secondary Structure	Solvent Accessibility	Torsion Angle ϕ	Torsion Angle ψ
Design 1	37.3	0.50	-0.25	-0.01	-0.09
Design 2	39.0	-0.50	-0.28	-0.00	0.10
Design 3	32.2	4.50	-0.08	0.05	0.47
Design 4	42.4	-0.25	-0.02	-0.01	-0.02
Design 5	35.6	-0.50	-0.10	0.01	-0.00
Design 6	33.9	3.50	-0.08	0.10	0.76
Design 7	39.0	-0.25	-0.19	0.04	-0.11
Design 8	25.4	3.75	0.17	0.09	0.68
Design 9	30.5	2.00	0.36	0.15	0.87
Design 10	30.5	4.50	0.16	0.26	1.22

EvoDesign results: design 1

Cluster 1 Center

Design Number	Energy	Iden. (%)
1	-106.83	37.3

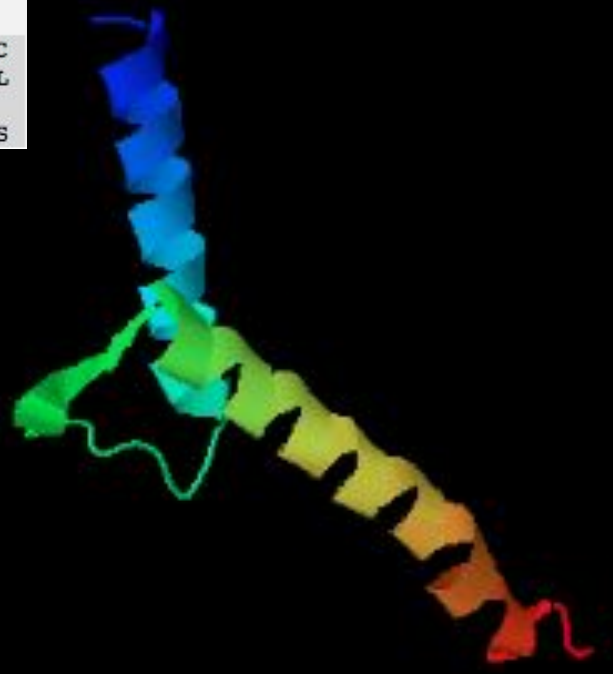
```
DSSP_SS  CHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHCCHHHHCC
Scaffold DDPMSLLYNMNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQIALDSHLKPSFL
                |  |||  ||  ||     |      ||  |||  ||    |  ||   ||   |
Design1  EAEKRLMNNMNICYQKLSKYLPMVNHKKKMSKMMILRMMIAYIHKLQHMLHHEMEEQYS
```

Predicted Secondary Structure

```
                          20                  40
                          |                   |
Sequence    EAEKRLMNNMNICYQKLSKYLPMVNHKKKMSKMMILRMMIAYIHKLQHMLHHEMEEQYS
Prediction  CHHHHHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHCC
Conf.Score  979999987999999999778899987866699999999999999999998601454239
            H:Helix; S:Strand; C:Coil
```

Predicted Solvent Accessibility

```
                          20                  40
                          |                   |
Sequence    EAEKRLMNNMNICYQKLSKYLPMVNHKKKMSKMMILRMMIAYIHKLQHMLHHEMEEQYS
Prediction  846443364242024303521421434542331102311132034134224453576 38
            Values range from 0 (buried residue) to 9 (highly exposed residue)
```



EvoDesign results: design 2

Cluster 1 Lowest Energy Sequences

Design Number	Energy	Iden. (%)
2	-123.29	39.0

```
DSSP_SS   CHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHCCHHHHCC
Scaffold  DDPMSLLYNMNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQIALDSHLKPSFL
                 ||| || ||| |||       | ||| ||    | ||  ||  |
Design2   EIEKMMMNNMNMCYDKLKMLVPMLNPSEKMSKMMILRIMIKYIKKLQEMLVREMEEIYS
```

Predicted Secondary Structure

```
                      20                  40
                      |                   |
Sequence    EIEKMMMNNMNMCYDKLKMLVPMLNPSEKMSKMMILRIMIKYIKKLQEMLVREMEEIYS
Prediction  CHHHHHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHCCCCCCC
Conf.Score  969889999999999999977799998787679999999999999999999854513369
            H:Helix; S:Strand; C:Coil
```

Predicted Solvent Accessibility

```
                      20                  40
                      |                   |
Sequence    EIEKMMMNNMNMCYDKLKMLVPMLNPSEKMSKMMILRIMIKYIKKLQEMLVREMEEIYS
Prediction  853353243214014403621343467440323301420340042024105646664448
            Values range from 0 (buried residue) to 9 (highly exposed residue)
```

EvoDesign results: design 3

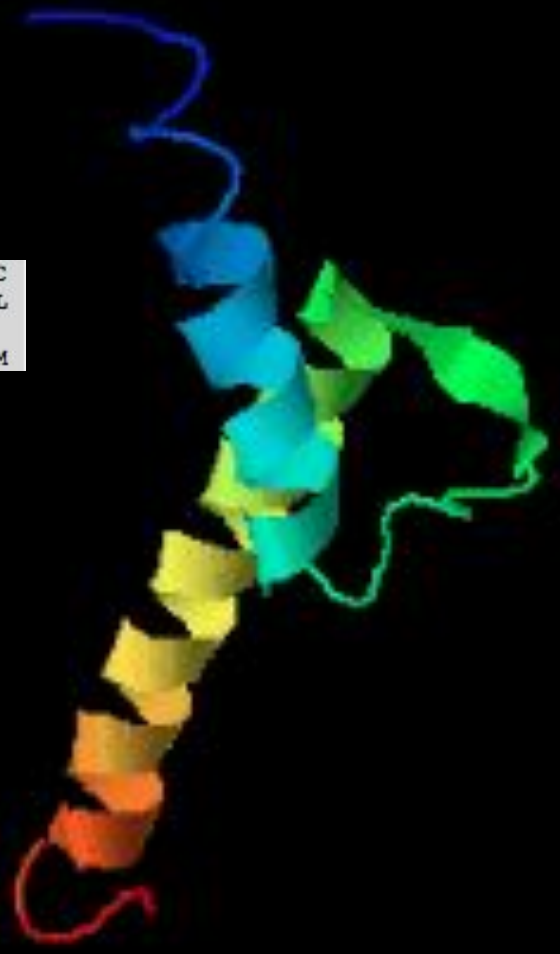
3	-122.96	32.2	DSSP_SS CHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHCCHHHHCC Scaffold DDPMSLLYNMNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQIALDSHLKPSFL Design3 NSPRRRKDNMNYCYNKLKKMIPMVMENEKMDKMMVLMLAIKHIMEIQDMLRRSASMMYM

Predicted Secondary Structure

```
                        20                  40
                        |                   |
Sequence    NSPRRRKDNMNYCYNKLKKMIPMVMENEKMDKMMVLMLAIKHIMEIQDMLRRSASMMYM
Prediction  CCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
Conf.Score  98220111267899999988169983378999999999999999999999987766419
            H:Helix; S:Strand; C:Coil
```

Predicted Solvent Accessibility

```
                        20                  40
                        |                   |
Sequence    NSPRRRKDNMNYCYNKLKKMIPMVMENEKMDKMMVLMLAIKHIMEIQDMLRRSASMMYM
Prediction  86564466424302530341033025474143110011114313403411444243347
            Values range from 0 (buried residue) to 9 (highly exposed residue)
```



Future plans

- Experimentally verify the efficacy of the generated sequences
- Develop alternative inhibition methods
  - Binding to only one site on the transcription factor instead of two
  - Inducing a conformational change in the transcription factor
  - Competitively bind to same region of DNA
  - Design of a peptide based off of a binding region using QUARK
    - Similar to work being done by Cabrele group