Design of intrinsically disordered region binding proteins

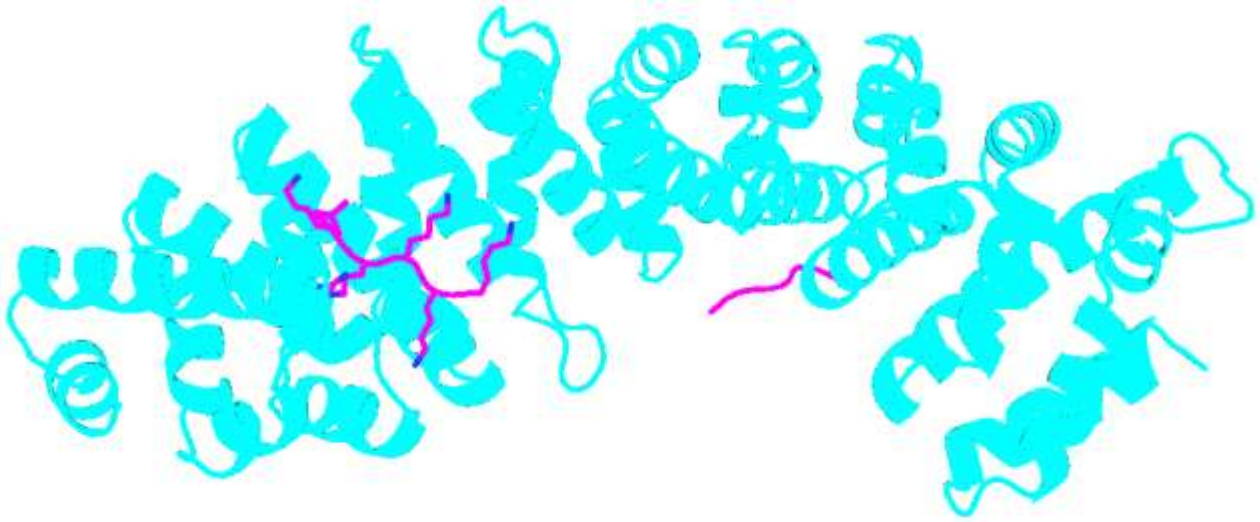
KEJIA WU, HANLUN JIANG, DERRICK R. HICKS, CAIXUAN LIU, [...], AND DAVID BAKER +24 authors Authors Info & Affiliations

Wenyue (Eva) Dai, 202550803
Study purpose

Background

Natural proteins to target IDR: Antibodies, Armadillo repeat protein etc.

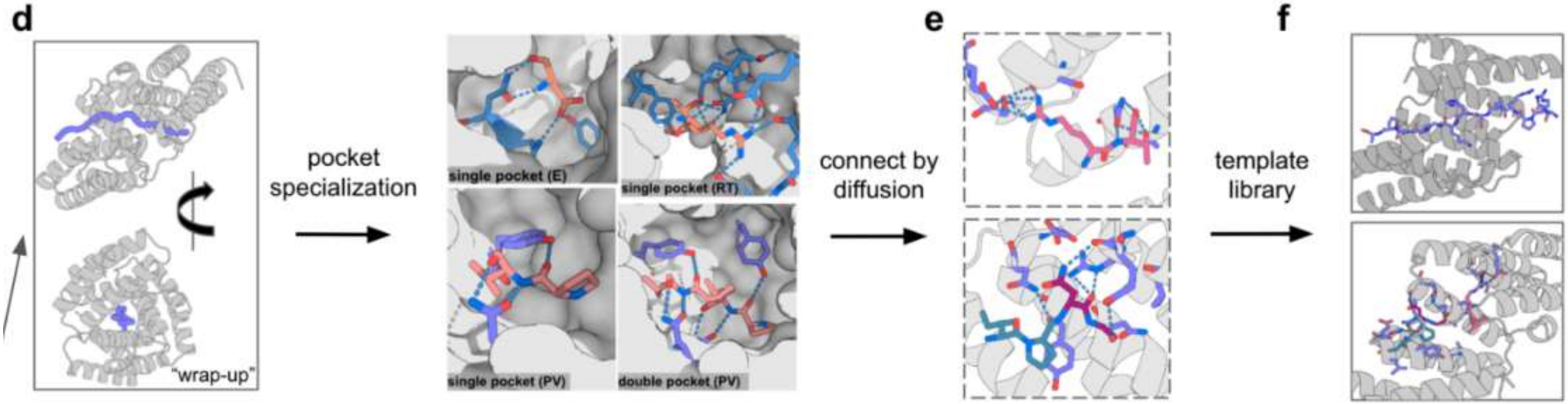
Repeat protein: DARPins, tetra-trico-peptide TRP



Armadillo repeat protein,
42 residues per repeat (pr)

Pdb: 4k5b
DARPins, 33 residues pr

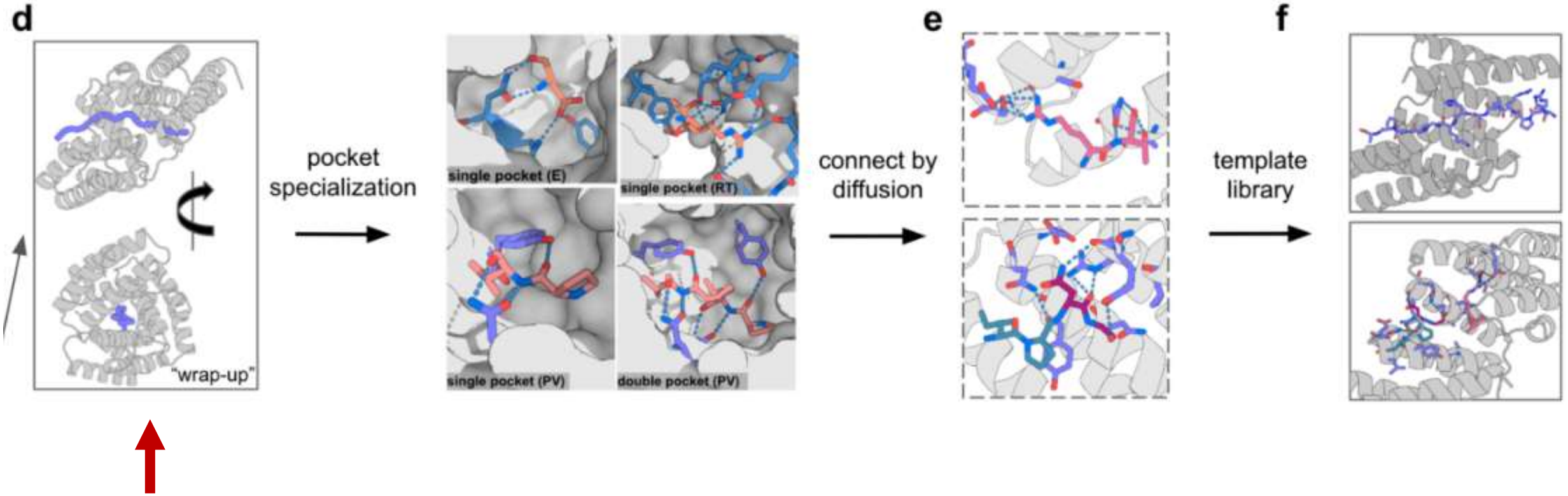
Strategy – step 1, Generation of template library



The template library should have two properties:

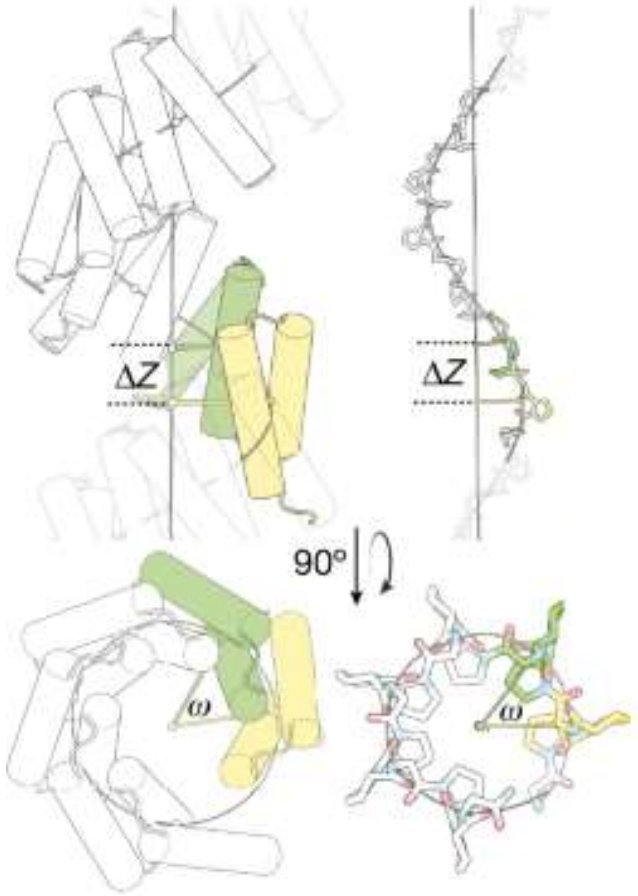
1. Each template structure should wrap around extended peptide conformation, with exten
2. The template library should be diverse.

Strategy – step 1.1, Scaffold generation



Design repeat protein that wrap around peptide in different repeating conformation

Strategy – step 1.1.1, Scaffold generation – backbone generation



- The translation (rise) along the helical axis per repeat unit;
- The rotation (twist) around this axis;
- The distance (radius) of the repeat unit centroid from the axis

Generated large sets of repeating-protein backbones that sampled a wide range of superhelical geometries

The helices range from 18 to 30 residues and the loops from 3 to 4 residues.

- Rise between 0 and 10 Å
- Twist (omega) between 0.6 and 1.0 radians
- Radius of 0 to 13 Å

https://www.nature.com/articles/s41586-023-05909-9

The backbone is usually ‘too perfect’ solely by parametric design:
Helix are perfect straight, loops are linear connection.
Missing backbone, N, C, O atoms, sidechain etc

Strategy – step 1.1.2, Scaffold generation – fragment asembly

- Fragment library: Pre-computed 3-9 residue fragments from high-resolution PDB structures
- For each window, calculate psi phi angle, and search in fragment library for matches
- Replace parametric segment with best matching fragment

But there are too many fragment from the library to try on, how to select the best one?

Monte Carlo & Metropolis-Hastings acceptance: Use score function to evaluate each replacement,
for example, total_score = w1*steric_score + w2*rama_score + w3*hbond_score + w4*rpx_score

Steric_score: Atom-atom clashes (soft repulsive lennard-Jones)
Rama_score: Ramachandra likelihood
Hbond_score: hydrogen bonding satisfaction
Rpx_score: hydrophobic packing prior to side chain assignment

https://www.nature.com/articles/s41586-023-05909-9

Strategy – step 1.1.3, Final filter for backbone genera

Final check the backbone topology

Loops must have been within 0.4 Å of a naturally occurring loop or be rebuilt.
Structures with helices deviating more than 0.14 Å were considered bent or kinked and were discarded.
Structures with fewer than eight helices in contact were also filtered out.
The distance between the first and the last helix was calculated and discarded long distance ones.

Strategy – step 1.1.4, ProteinMPNN & AF2 Filter

Sequence design was performed on each filtered backbone using ProteinMPNN, with a customized weight on certain AAs: {"A": -0.15, "G": -0.15, "M": -0.35, "P": 0.15, "E": 0.1}.

Negative – penalize
Positive – favor

Penelize A, G to force better packing with sidechain & avoid too flexible
Penelize M for steric clash and oxidation risk
Favor P: to allow loop rigidity, helix capping, enrtopic bonus (fold faster and more thermostable)
Favor E: solubility, specificity
Why not favor K/R?

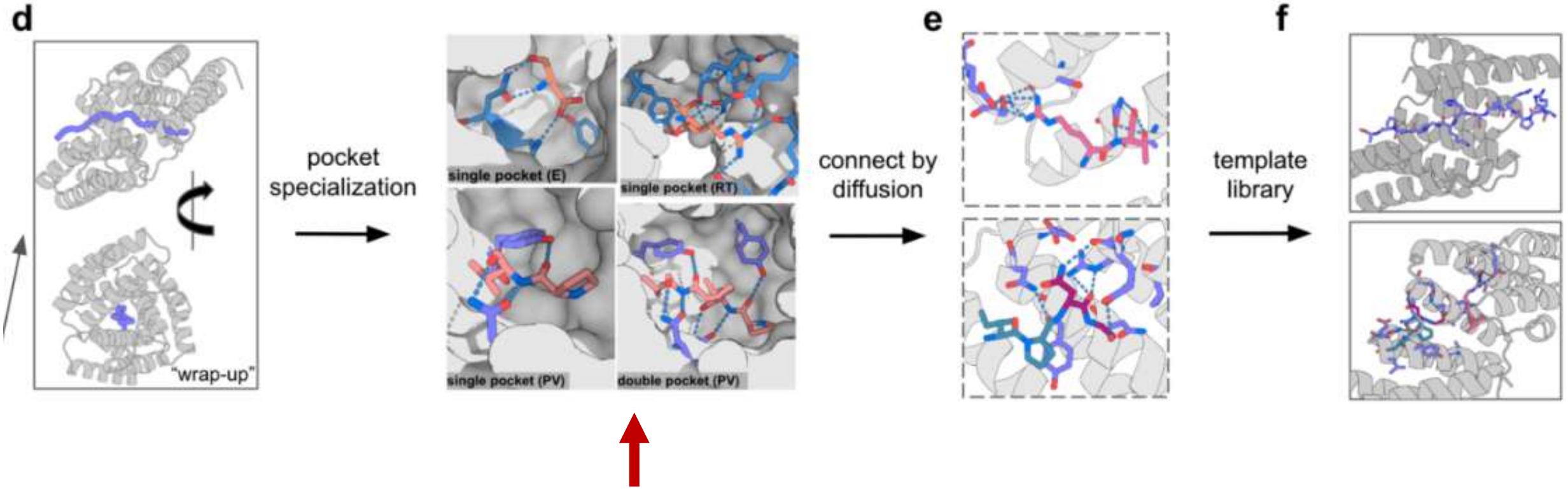
- IDR enriched in KR, less in E
- E can form bidentate with R (only consider side chain O, not backbone)
- K/R is long flexible chain, E is more compact (entropic penalty)
- K/R: higher non-specific risk. cation-phi interaction
- Favor expression: ribosome exiting tunnel

AF2 filter

Structural prediction was conducted by AF2(46) with PLDDT > 90, Cα RMSD < 2 Å.

https://pubs.acs.org/doi/full/10.1021/acs.jctc.9b00742

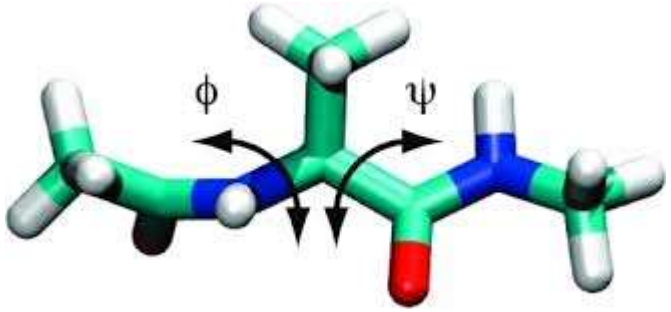
Strategy – step 2.1, Pocket specialization



Convert generic helical repeat scaffolds into amino acid-specific binding pocket capable of recognizing individual residues from IDR

Strategy – step 2.1.1, Di-peptide binder design

1. Generate peptide conformation: Sample phi/psi angle from extended Ramachandran and filter with Rosetta Ramachandran score to ensure no clashes. (Remember, side

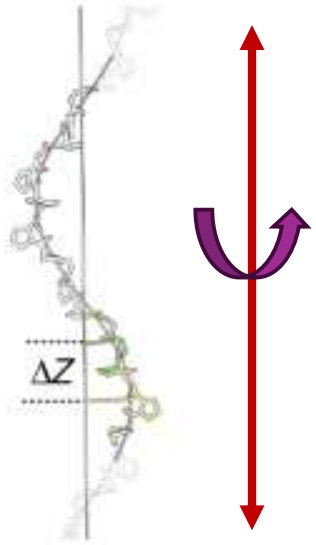


2. z-axis alignment and grid search:
Rigid dock, side chain retain initial conformatoin

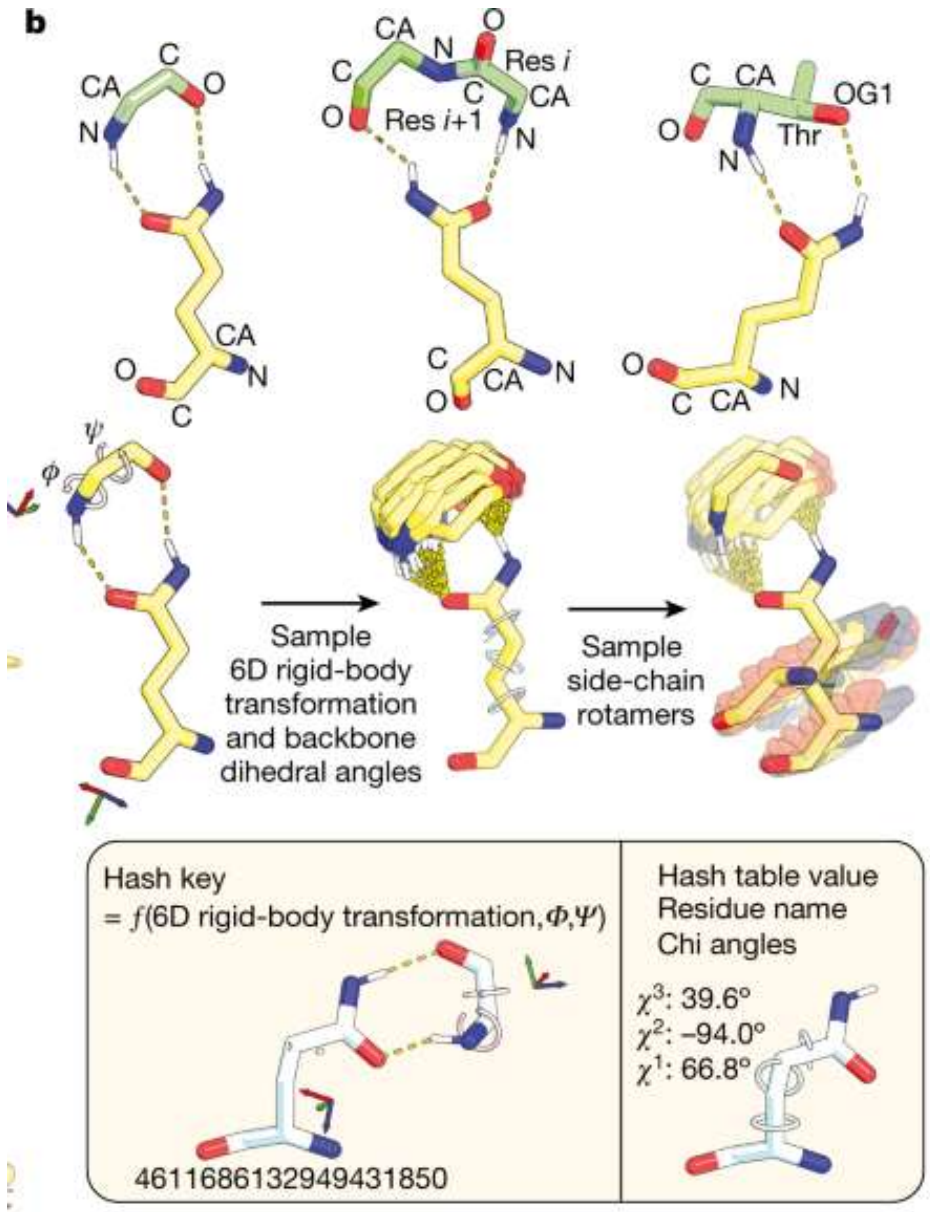
Align peptide & protein along superhelical axis (z-axis)

Perform 2D grid search (rotation around z-axis + translation along z-axis)

Accept only clash-free conformations



Strategy – step 2.1.2, Hash table look-up



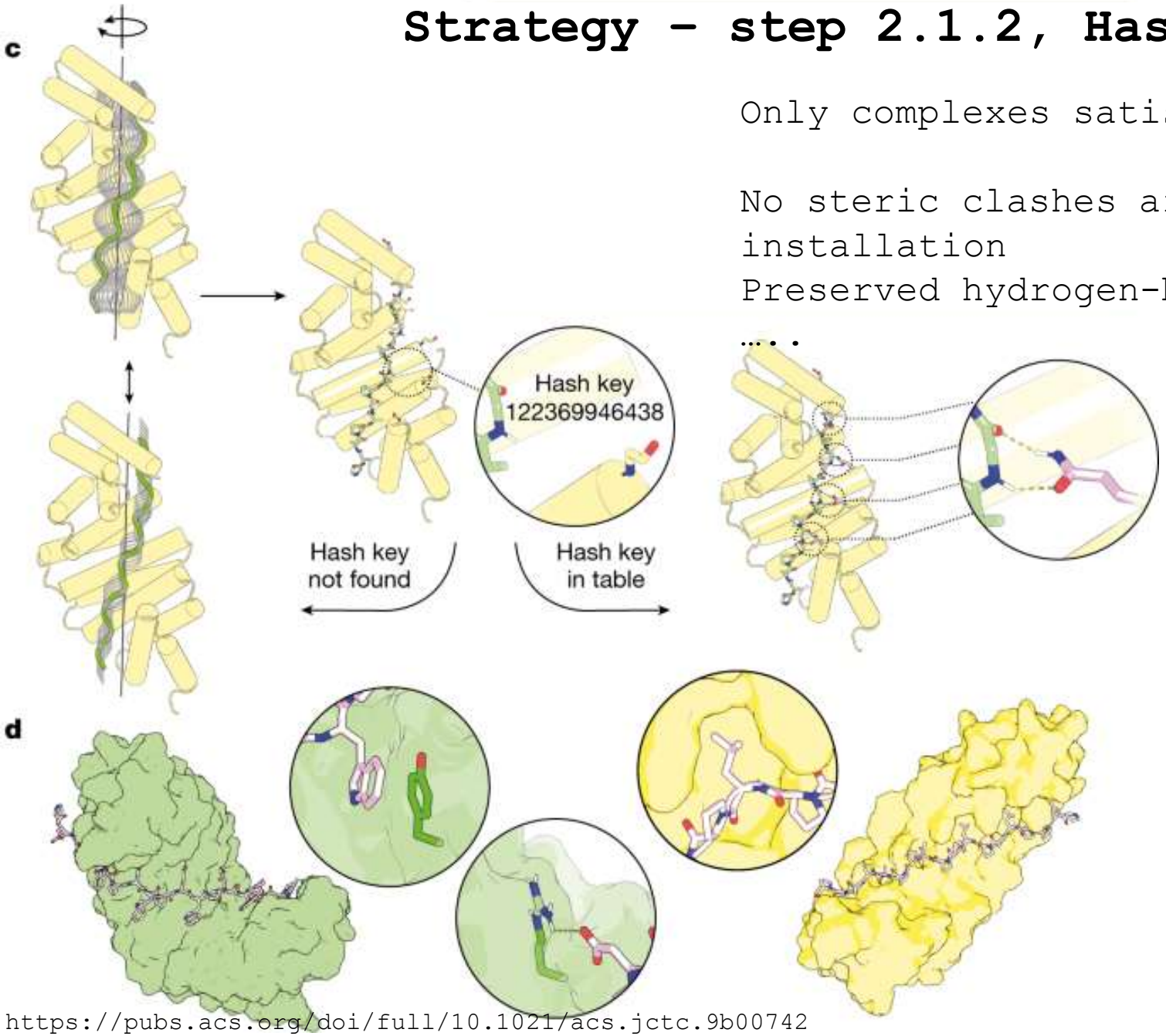
1. For clarity, let's call it donor and acceptor
2. For donor, 6D (3*translation, 3*rotation), psi, ph
3. For acceptor, the rotamer can be calculated

Then the donor's 6D+phi+psi are converted to hash key
The side chain rotamer are corresponding hash value.
This process can allow fast search.

Strategy – step 2.1.2, Hash table look-up

Only complexes satisfying:

No steric clashes after rotamer installation
Preserved hydrogen-bond interactions
…..



https://pubs.acs.org/doi/full/10.1021/acs.jctc.9b00742

Strategy – step 2.1.3, ProteinMPNN & AF2 validation

After anchoring the key interacting residues from hash table

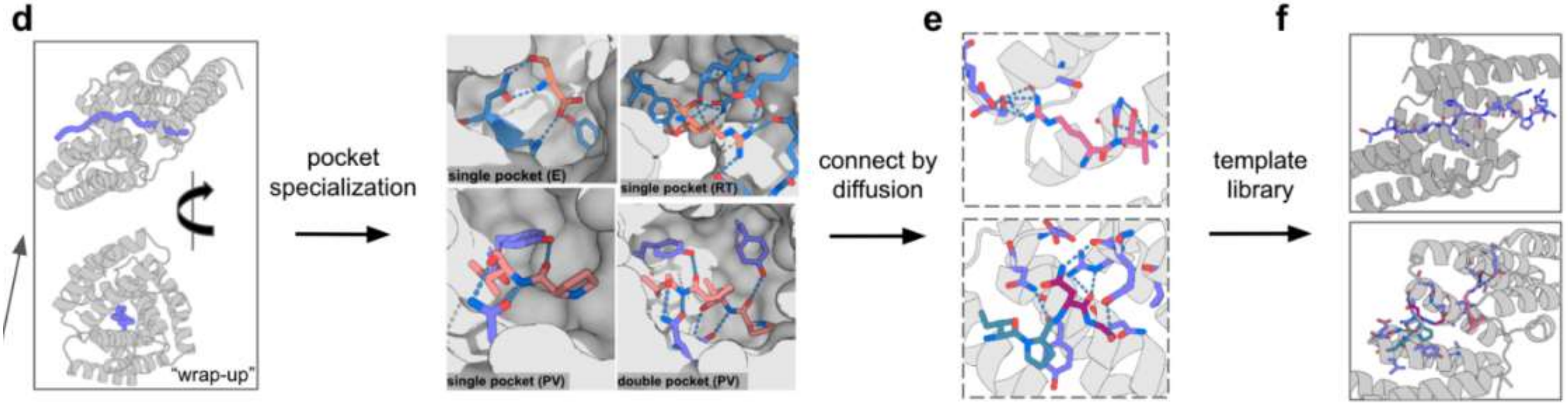
Interface design optimization:

1. Optimize non-anchor residues (ProteinMPNN?)
2. Validate hydrogen bond network

AF2 & Rosetta filter

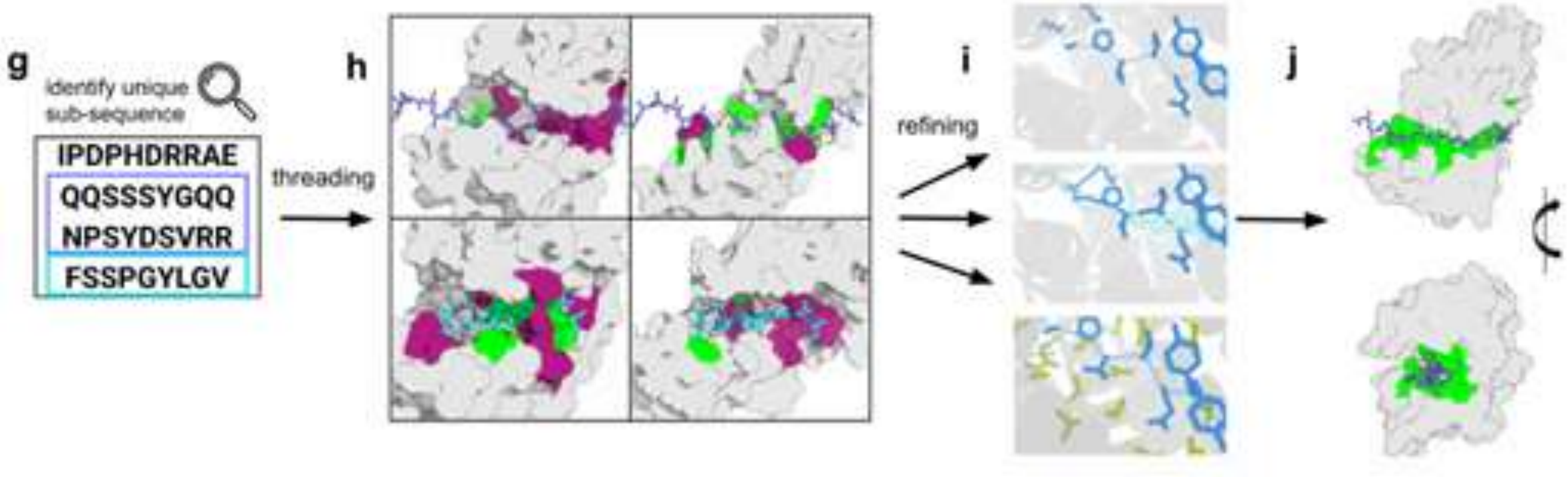
PAE_interaction ≤10,
PLDDT > 92 for the complex,
Cα RMSD < -50,
contact_molecular_surface (CMS) > 500
BUNS (buried_unsatisfied_penalty) < 1.

Strategy – step 2.3, Pocket assembly to generate template



For each dipeptide, collect 10AA binding pocket,
Assemble combination of 2-6 binding pockets using di
By different combination of binding pocket, end up w
Each template contains protein and peptide docking c

Strategy – step 2.4, Thread IDR to template library



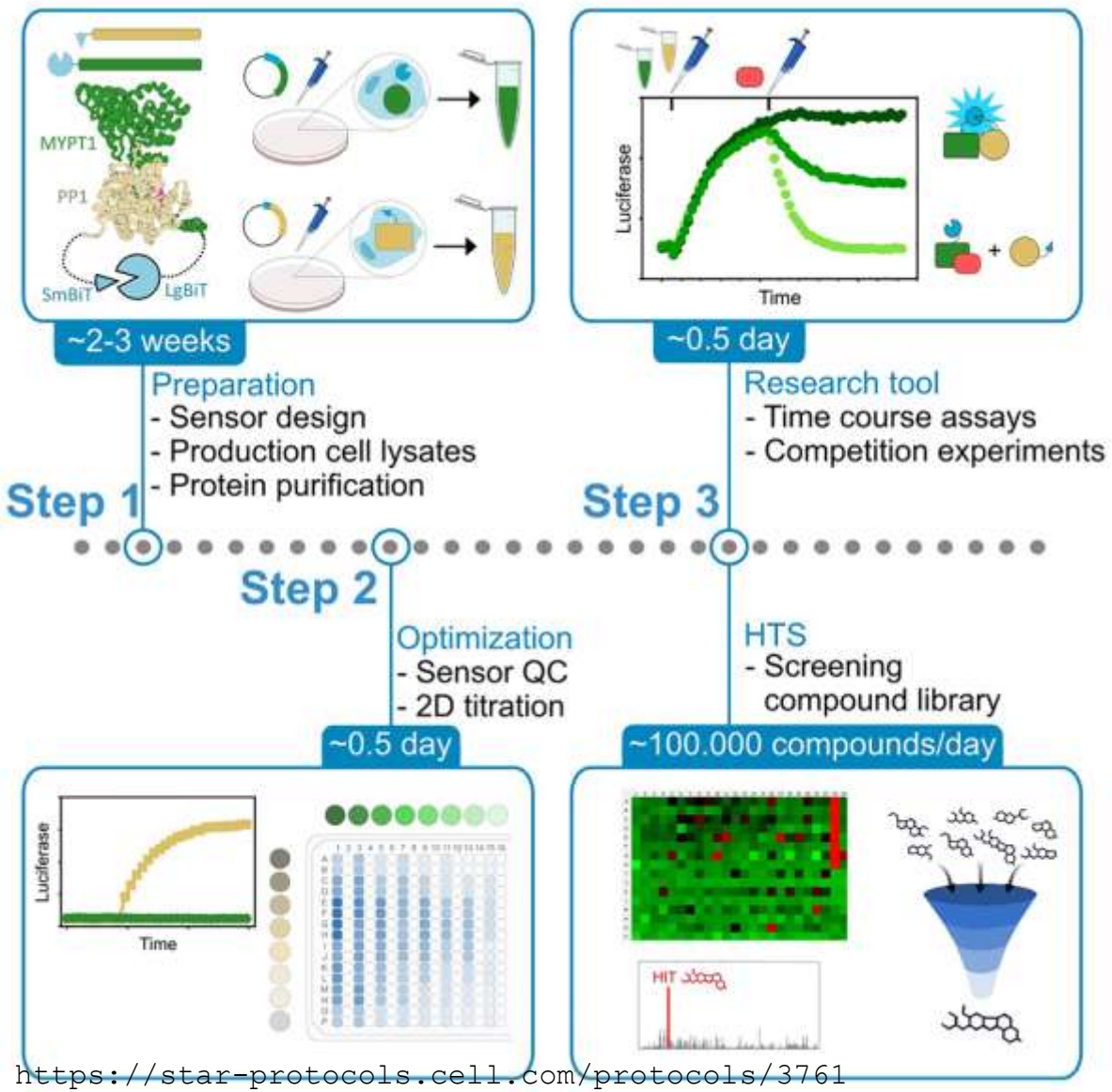
Thread IDR sequence (8-40aa) to the peptide in the template library.
Refine with one-side / two-side/motif Rfdiffusion & ProteinMPNN

- One side: only optimize scaffold
- Two side: both scaffold & IDR
- Motif: only keep essential binding site, rebuild the rest

Score with geometry & Rosetta score

Experimental results

nanoBiT split luciferase

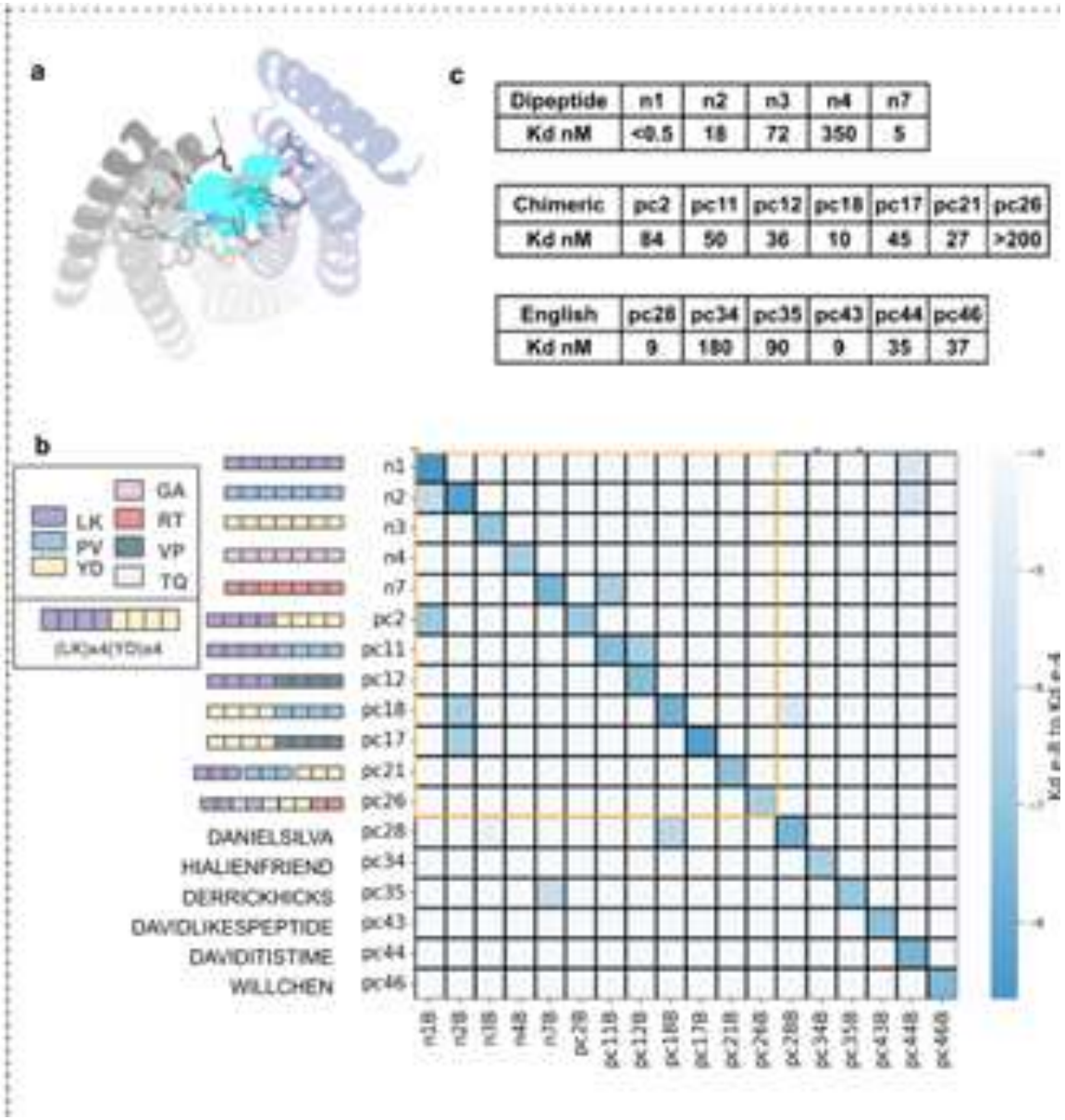


1. Fuse split luciferase in IDR and designed p
2. Collect cell lysate that express these two
3. Mix and detect

https://star-protocols.cell.com/protocols/3761

Experimental results: to synthesized peptide

nanoBiT split luciferase



1. High success rate: among 18 IDR, most designed binder kd < 100 nM

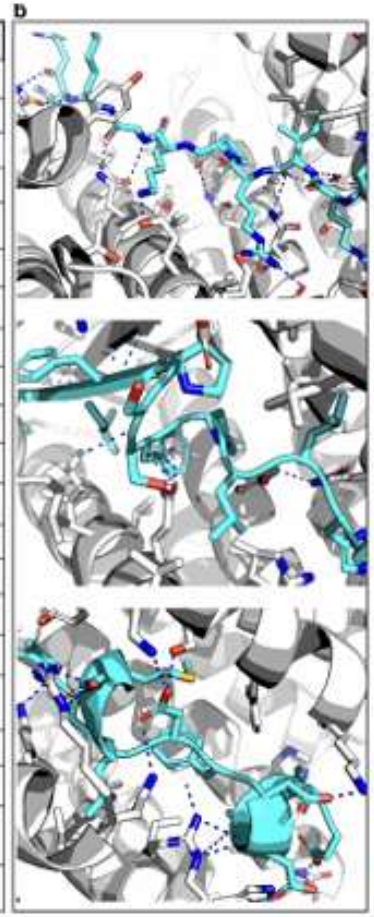
2. High specificity: 18x18 cross exam, most only show binding to its own

Experimental results: to native IDR

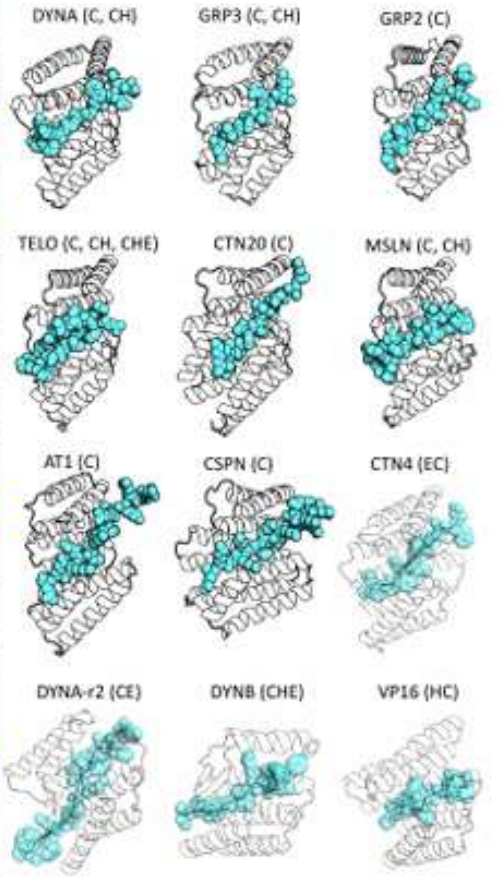
a

Abbrev.	Full Name	Sequence	Hits %	Kd, nM (v1, v2)
DYNAr2	Dynorphin A-r2	YGGF**LRRIRPKLKWDNQ**	90*	<0.06*
DYNA	Dynorphin A	YGGF**LRRIRPKLK(WDNQ)**	50	<1, <1
DYNB	Dynorphin B	YGGF**LRRQFKVVT**	34	<1
MSLN	Mesothelin	NGY**LVLDLSMQEALS**	21	<1
SP	Substance P	**RPKPQQFFGLM**	13	7
CTN20	Cystinosin20-33	C**ESSVSLTVPPVVK**	40	9
GRP3	Glycine-rich extracellular	VLW**NSRWPTLQAWGAGLK PGY**	9	10
CTN4	Cystinosin4-15	**NWLTIFILFPLK**	35	25
CPS	C-peptide-short	**SLQPLALEGSLQ**	8	33
FAM21	Wash complex	**SSDDDLFQSAKPKPAKKTN PFPLLEDE**	67	35
GRP2	Glycine-rich extracellular	HENGPWPGPC**NARVAPMLL PRLPTPGVPSD**	20	42
VP16	Viral VP16	**DALDDFDLDMLPA**	35	45, 19
INS9	Insulin9-23	**SHLVEALYLVCGERG**	19	59
PER2	Period 2	**AVPFPAPVPAAY**	42	67
GRP1	Glycine-rich extracellular	**RRPWVPHLLPFSSPGYL**GV M	10	72
AT1	Angiotensin I	**DRVYIHPFHL**	25	75
TELO	Telomerase	**SWAHPGRTRGPSDRGFCVV SPARPAEEATSLEGALSGT**	42	91
APE	Apelin-13	**QRPRLSHKGPMPA**	9	92
EA1	EML4/ALI fusion protein	**PTPGKGPKVYHHKHQE**	7	100
CSPN	Circumsporozoite	**DNEKLRKPKHKKLKQ**	15	195
EF1	EWS/FLI fusion protein	**SQGSSSYGQQNPSYDSVRR**	2	>500

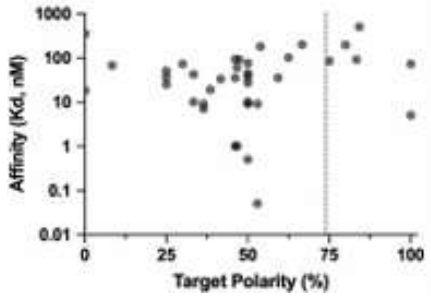
b



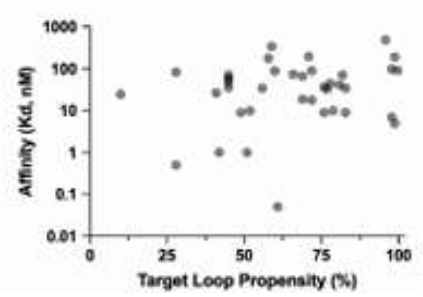
c



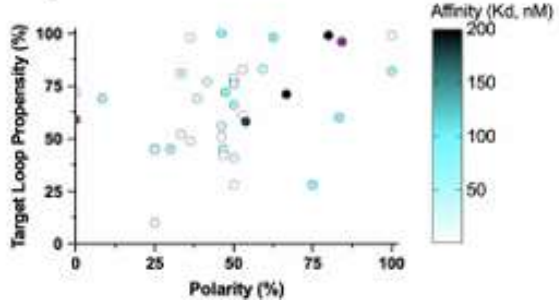
d



e



f



Among 21 target, 21 all successful
Kd range a lot
EF1 Kd > 500
DYNAr2 Kd < 0.06

Conclusion

A combination of physical properties selection (hash table), different diffusion stra
rounds of design with different hightlight are all useful and inspiring.

However, the pipeline seems to be very tedious, and some tech details are not entirel