Protein modelling using dynamic programming and constraints

Maryana Wånggren[1] Martin Billeter[2] Graham J.L. Kemp[1]

[1]Department of Computer Science and Engineering,
Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

[2]Department of Chemistry and Molecular Biology, University of Gothenburg,
SE-405 30 Gothenburg, Sweden

22 May 2017

Overview

- Aims of the project
- Protein structure basics
- Types of constraints
- Zipping and assembly method (dynamic programming) for generating solutions
- Results
- Conclusions and to-do list

Overview

- **Aims of the project**
- Protein structure basics
- Types of constraints
- Zipping and assembly method (dynamic programming) for generating solutions
- Results
- Conclusions and to-do list

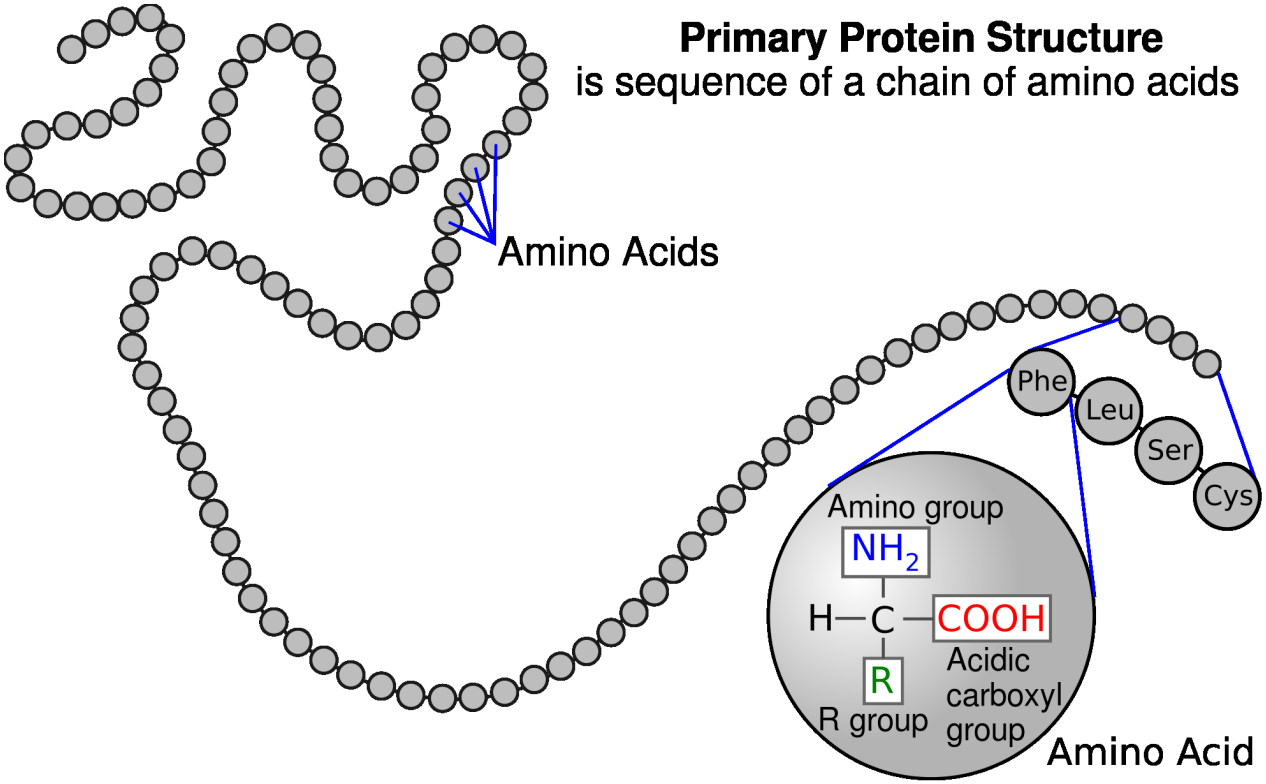
Aims of the project

- **Practical side:**
  - protein modelling using a limited set of constraints for reducing the cost of expensive and time-consuming experiments
- **Technical side:**
  - combination of zipping and assembly method with easy-to-obtain distance and angle constraints
  - use Prolog for generating distance and angle constraints

Overview

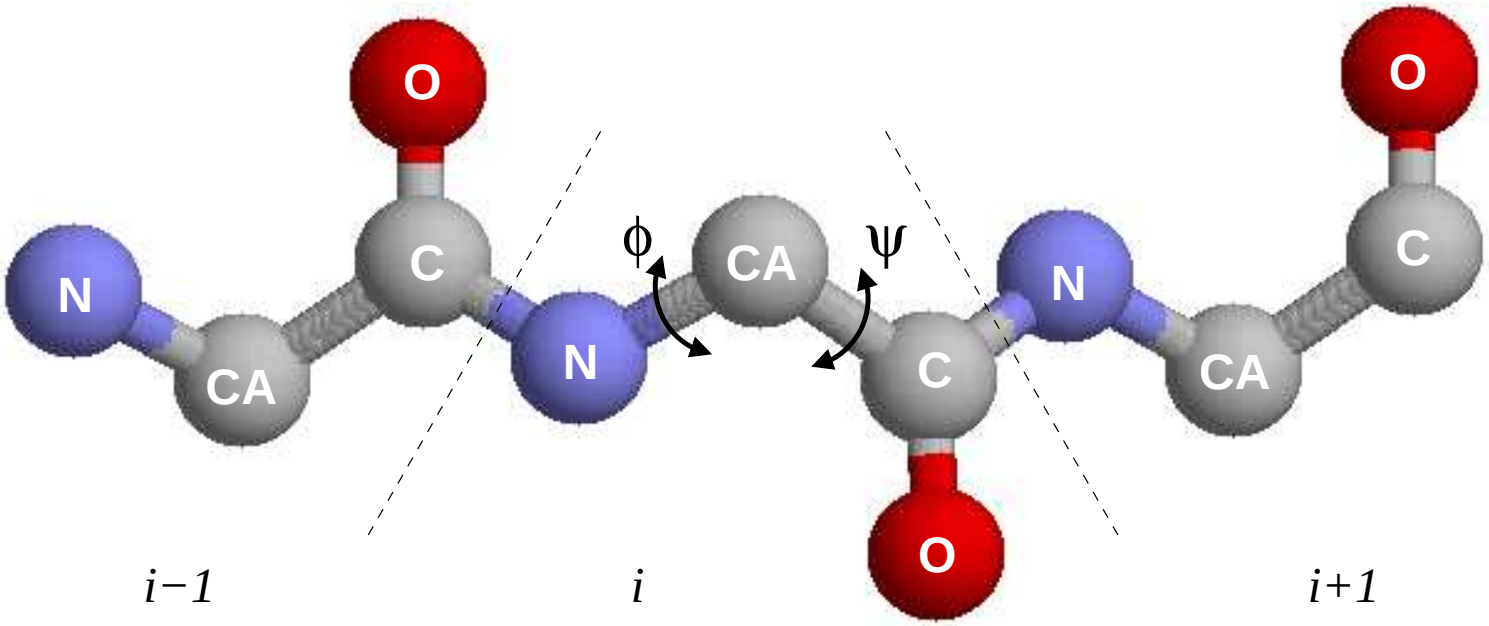
- Aims of the project
- **Protein structure basics**
- Types of constraints
- Zipping and assembly method (dynamic programming) for generating solutions
- Results
- Conclusions and to-do list

Protein primary structure

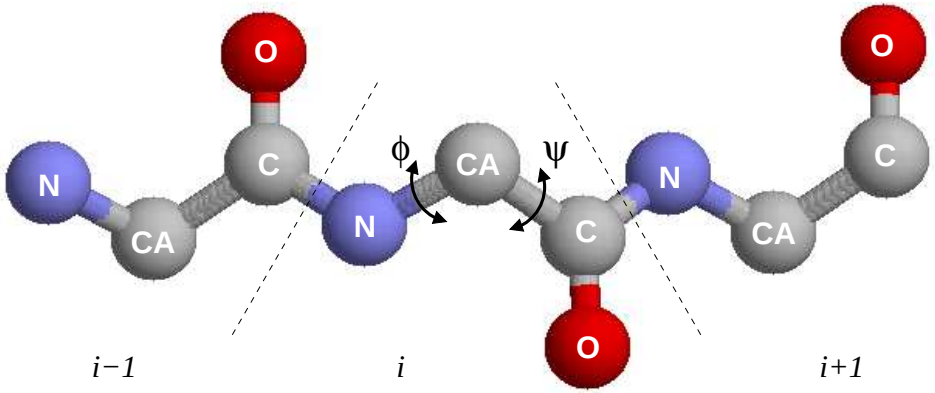


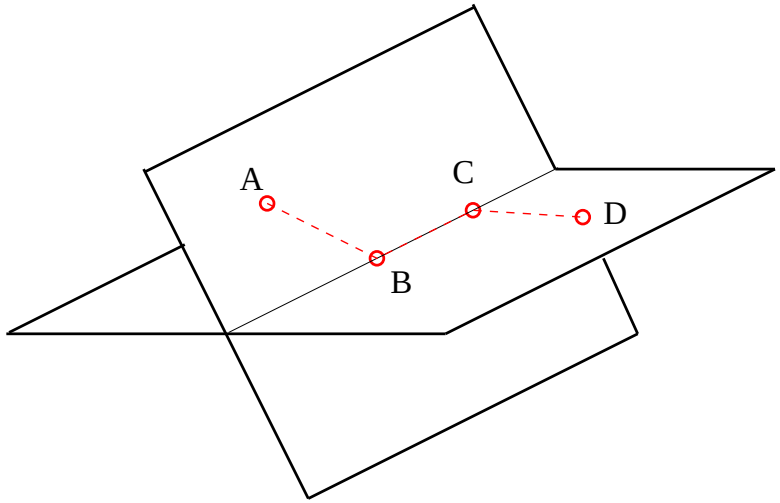
[By National Human Genome Research Institute [Public domain], via Wikimedia Commons]

Protein main chain model

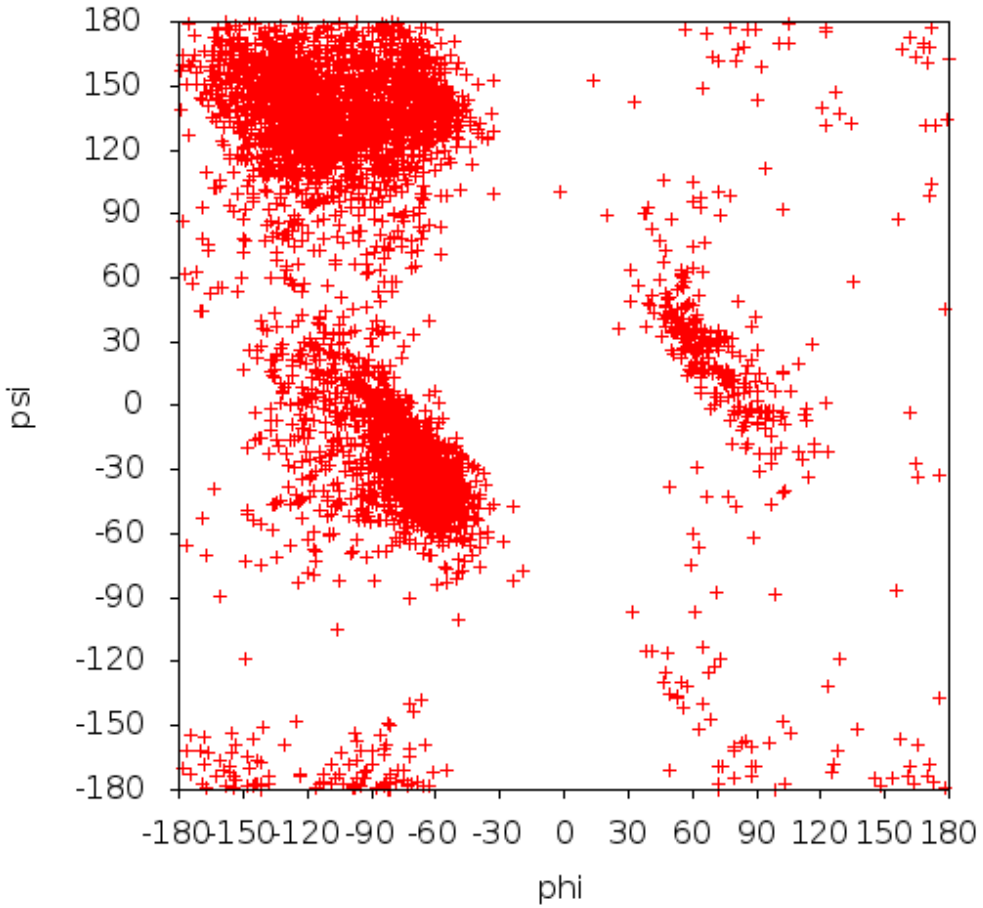


Torsion angles





ϕ and ψ torsion angles from a library of proteins from the Protein Data Bank



Overview

- Aims of the project
- Protein structure basics
- **Types of constraints**
- Zipping and assembly method (dynamic programming) for generating solutions
- Results
- Conclusions and to-do list

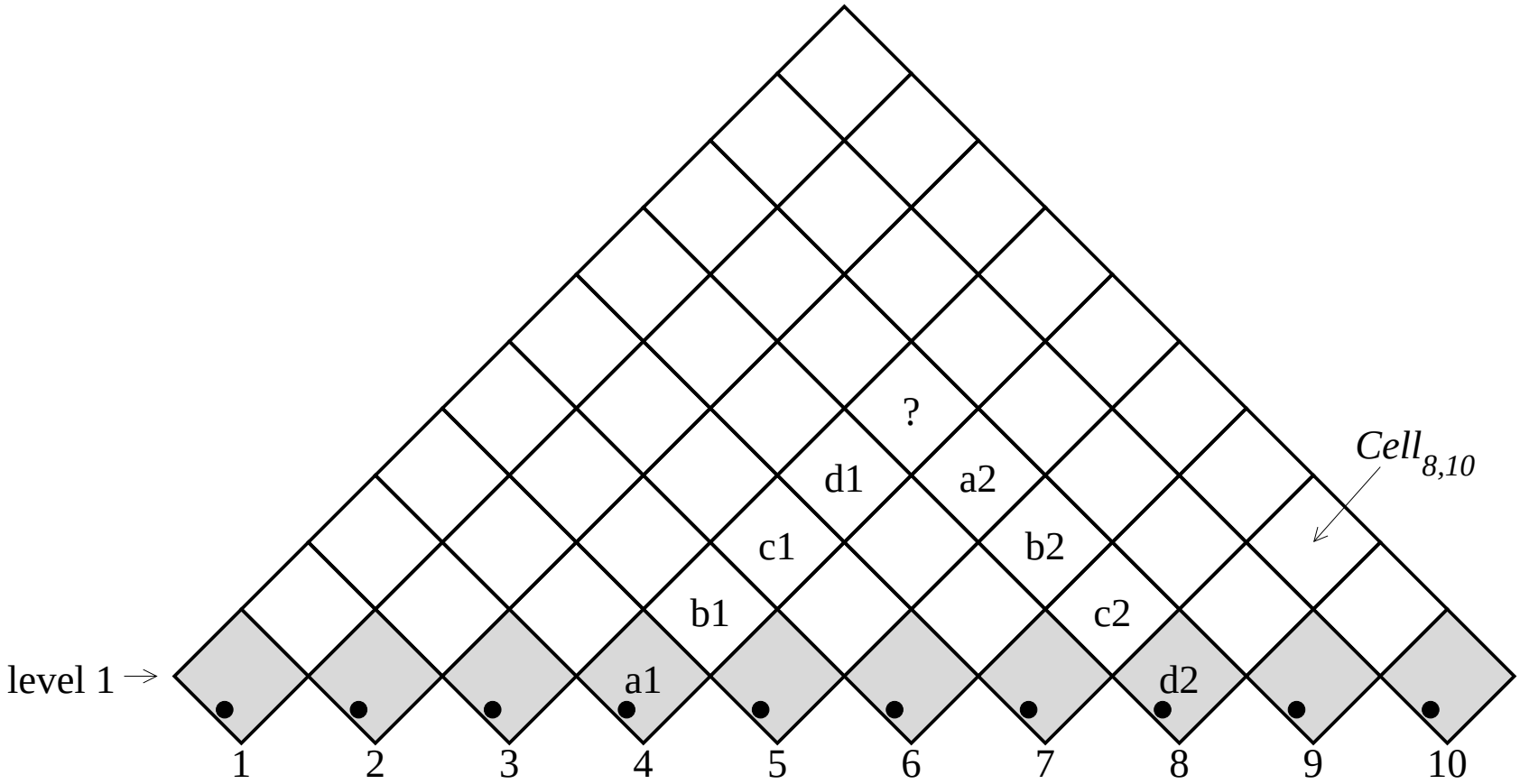
Protein modelling: what information is used?

- Angle constraints
  - range limits on ϕ and ψ
- Distance constraints
  - upper and lower bounds on the distances between pairs of atoms

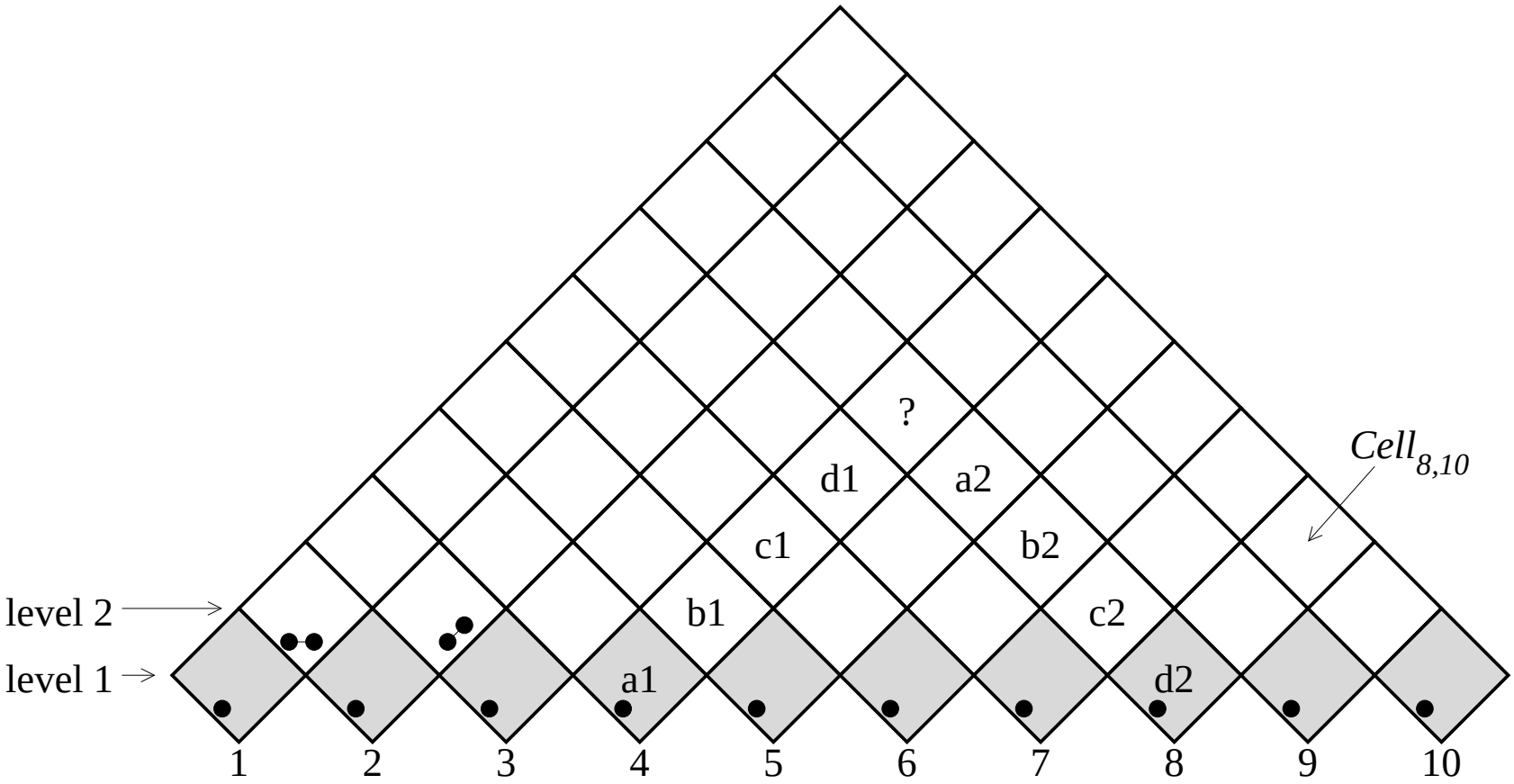
Overview

- Aims of the project
- Protein structure basics
- Types of constraints
- **Zipping and assembly method (dynamic programming) for generating solutions**
- Results
- Conclusions and to-do list

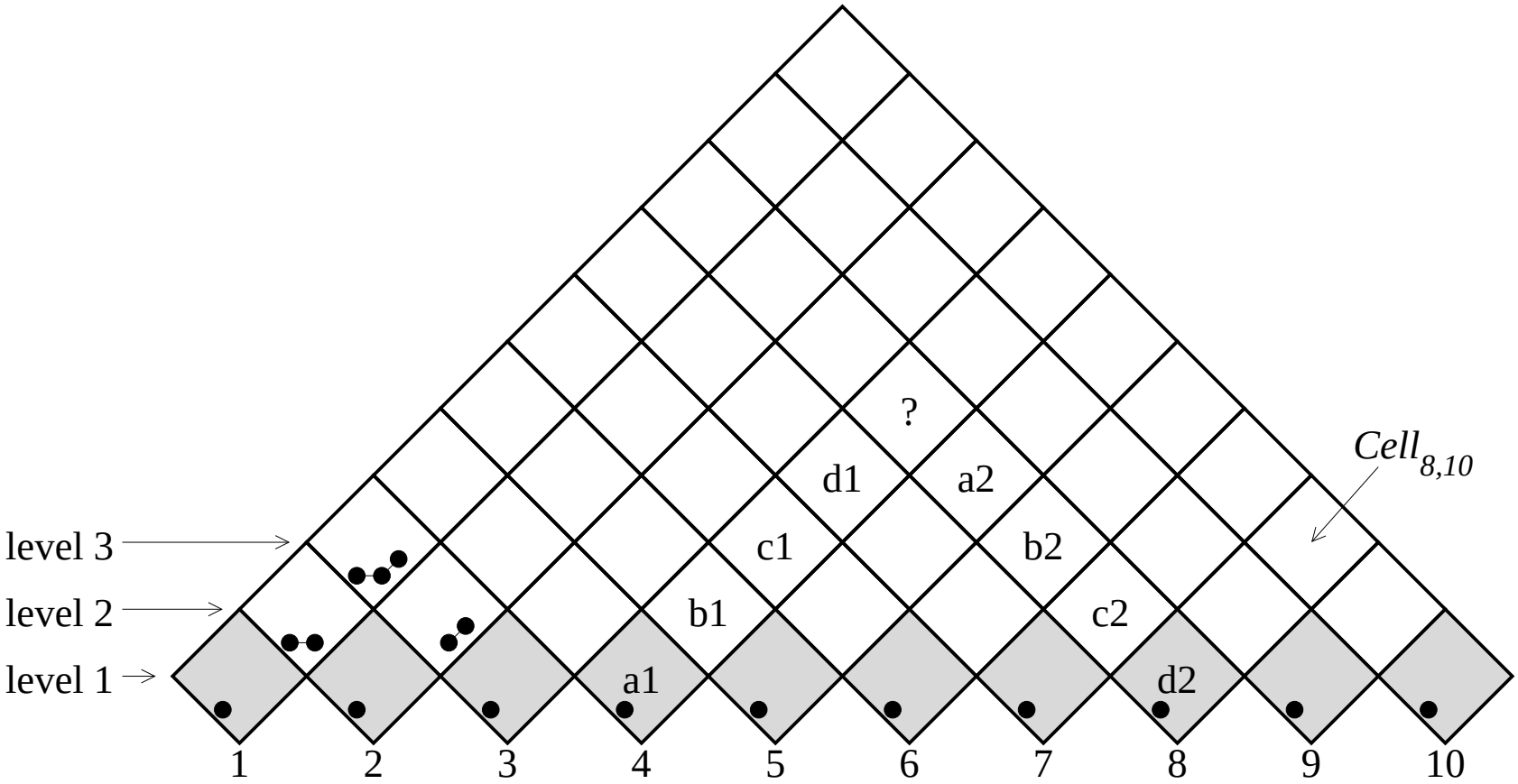
Zipping and assembly of a protein with 10 residues (single residues)



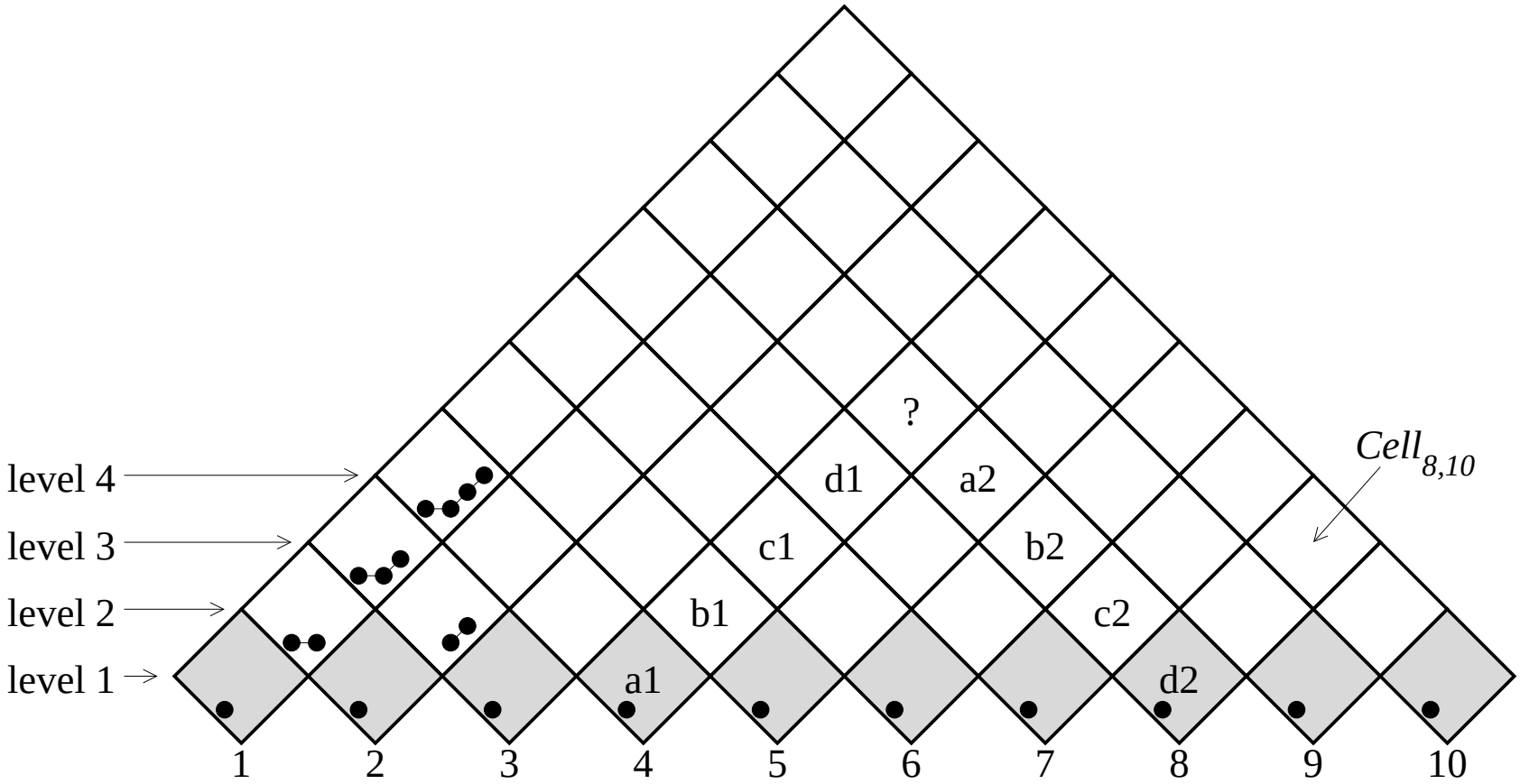
Second level of ZAM



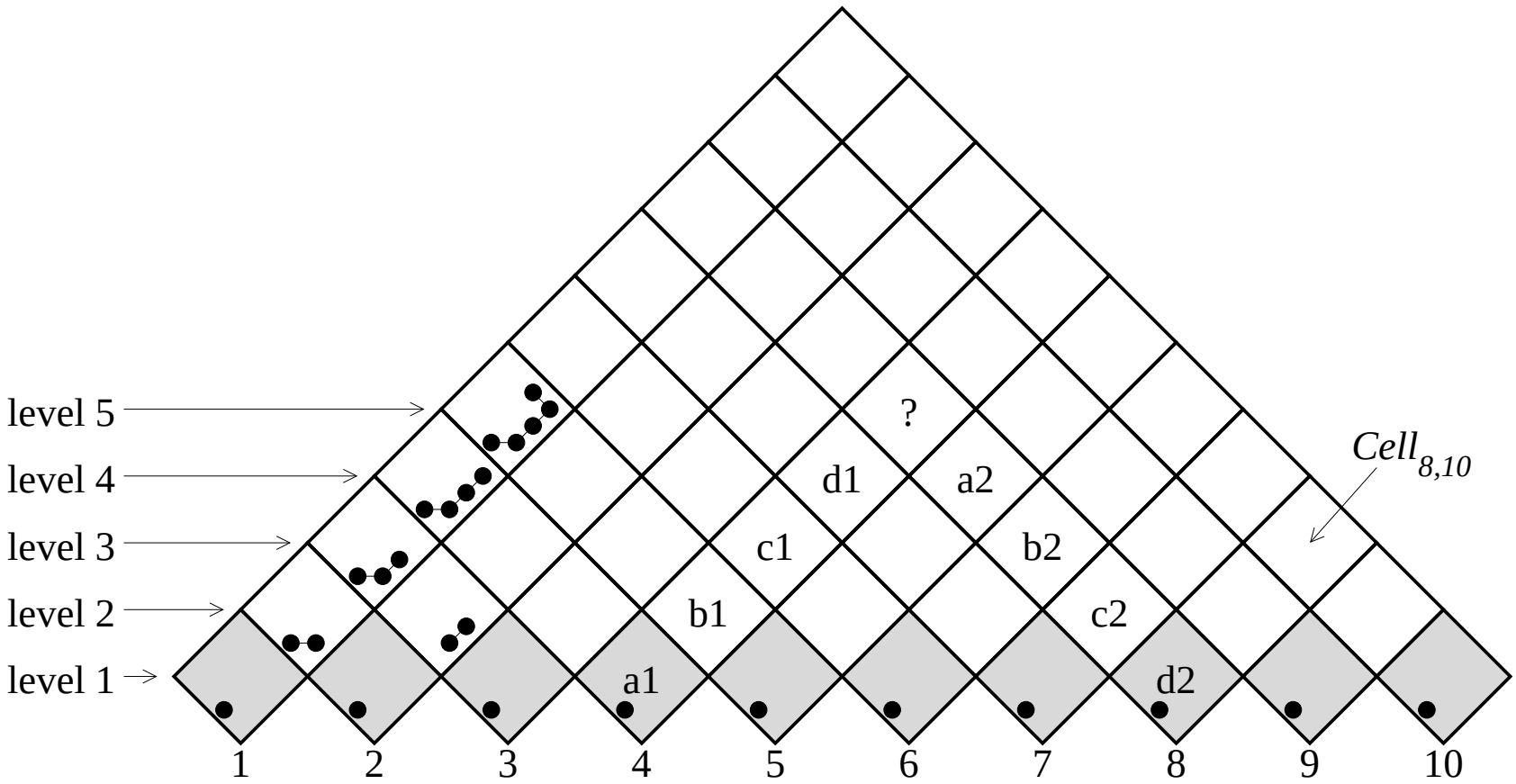
Zipping and assembly of a protein with 10 residues (3 residues)



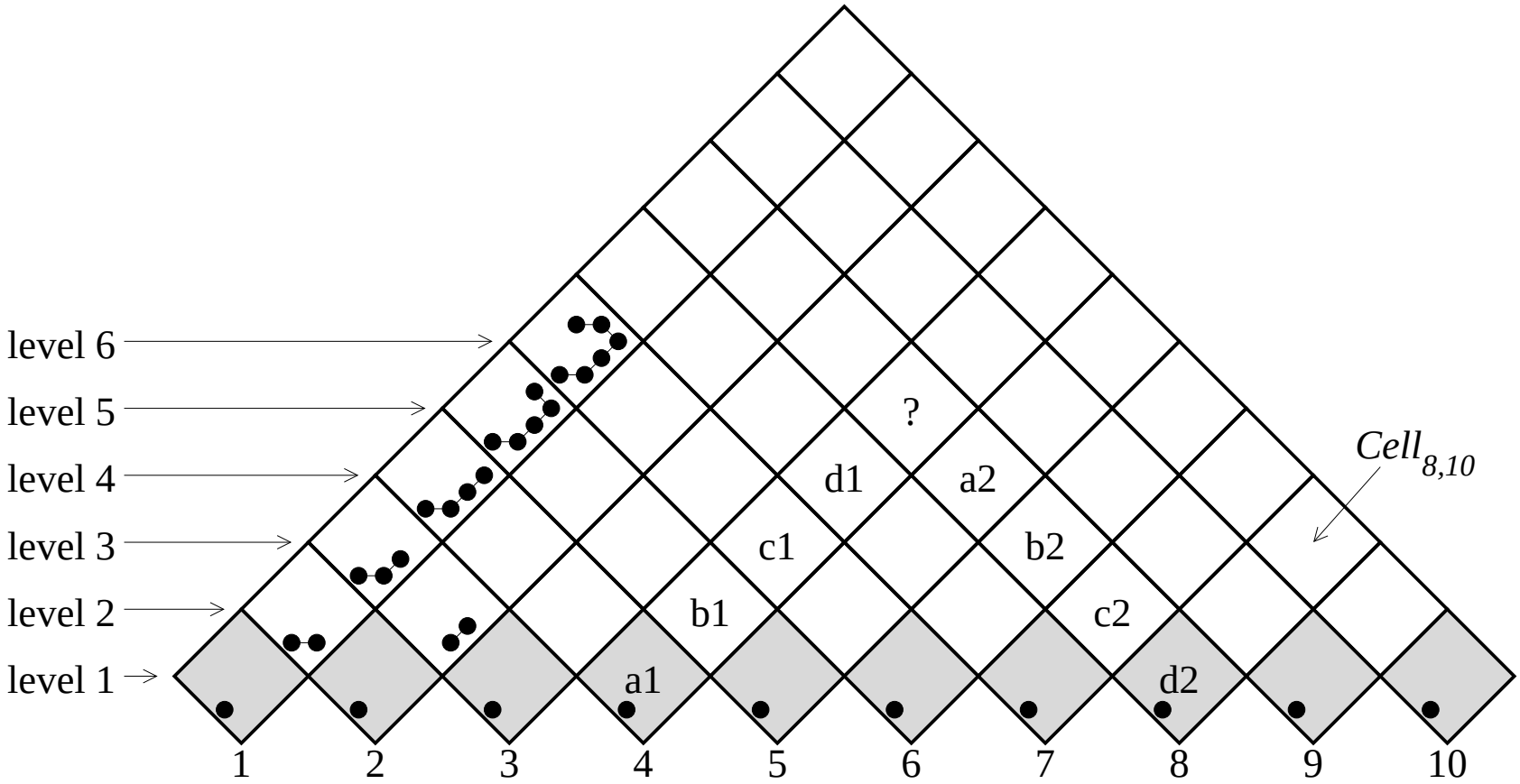
Zipping and assembly of a protein with 10 residues (4 residues)



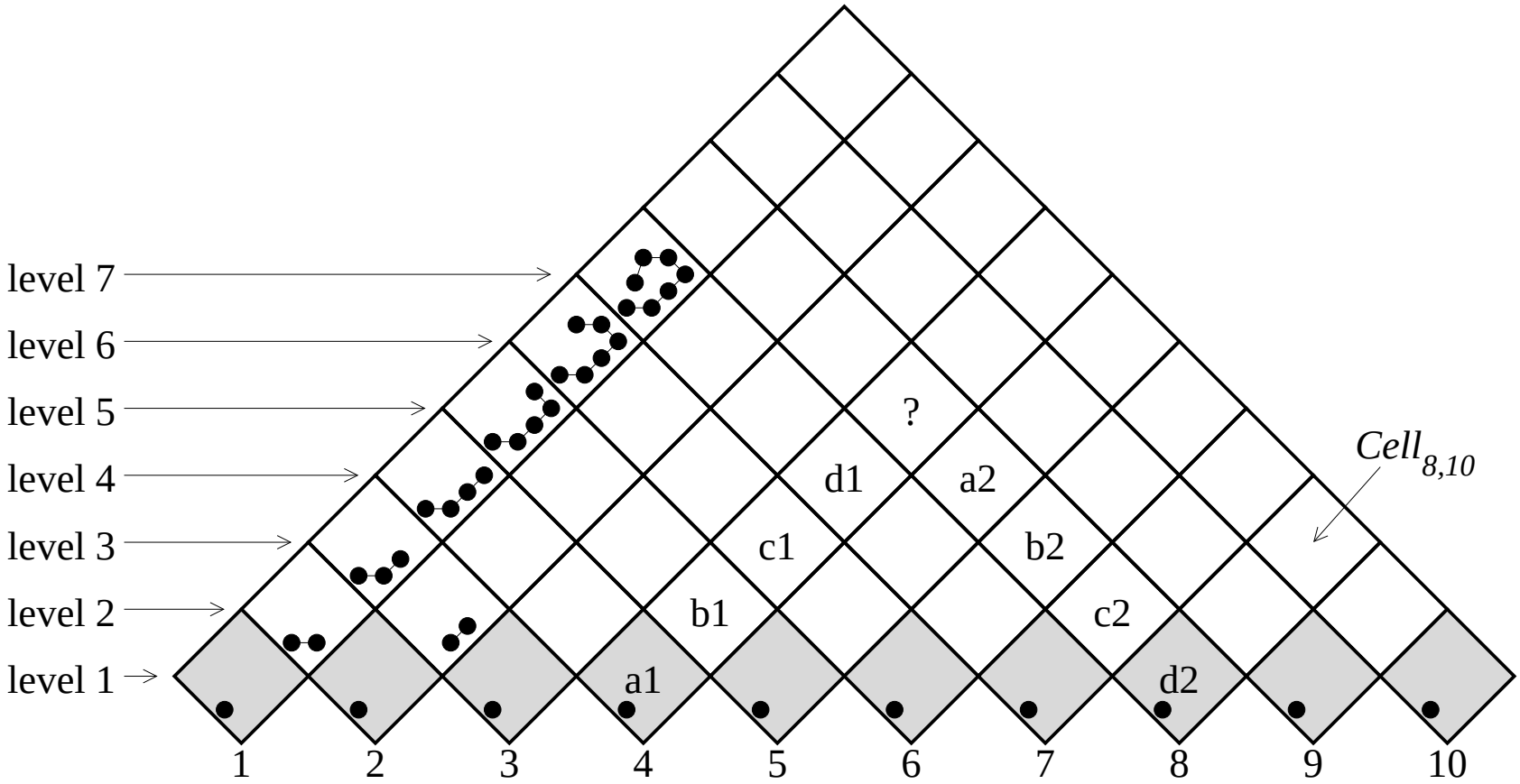
Zipping and assembly of a protein with 10 residues (5 residues)



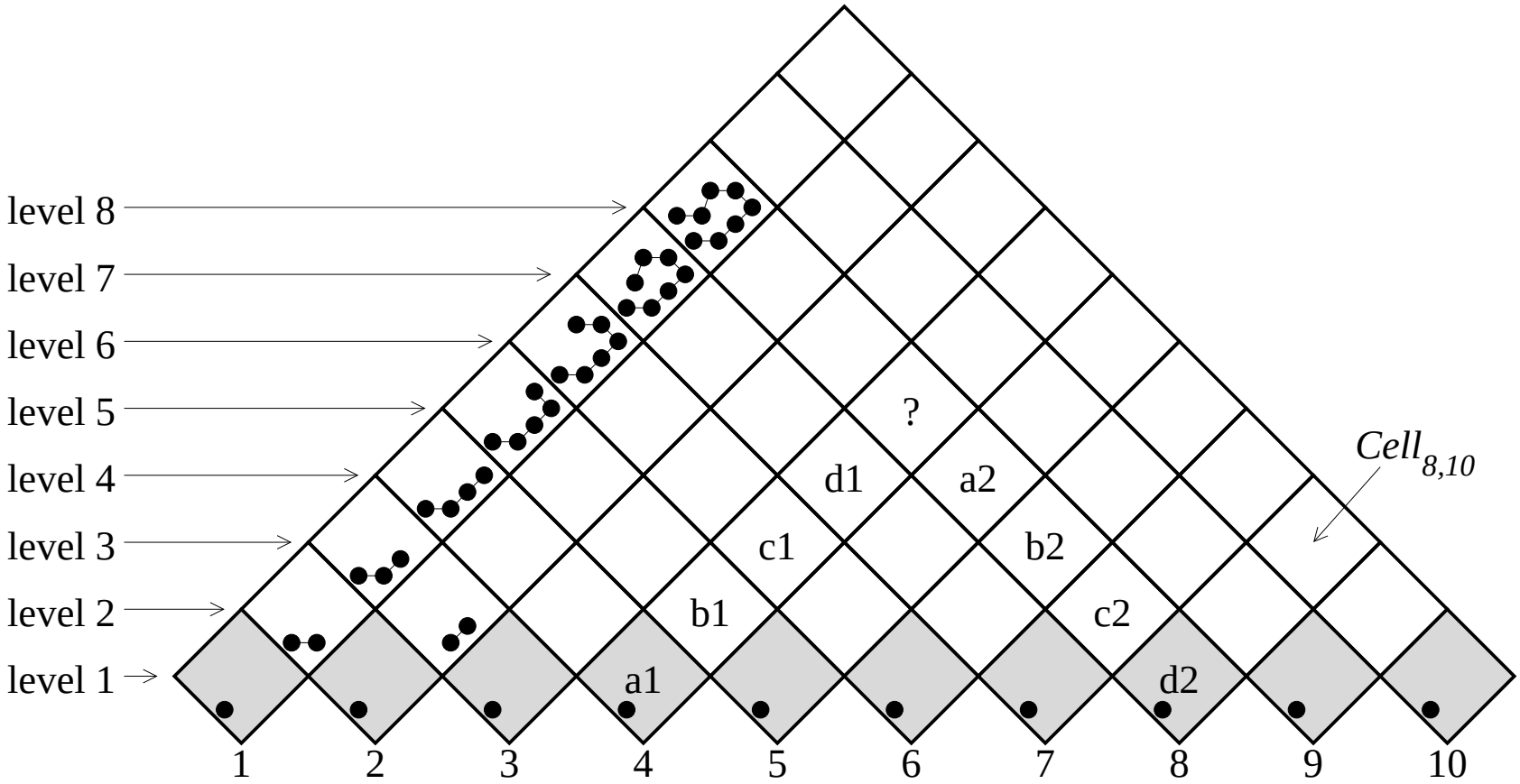
Zipping and assembly of a protein with 10 residues (6 residues)



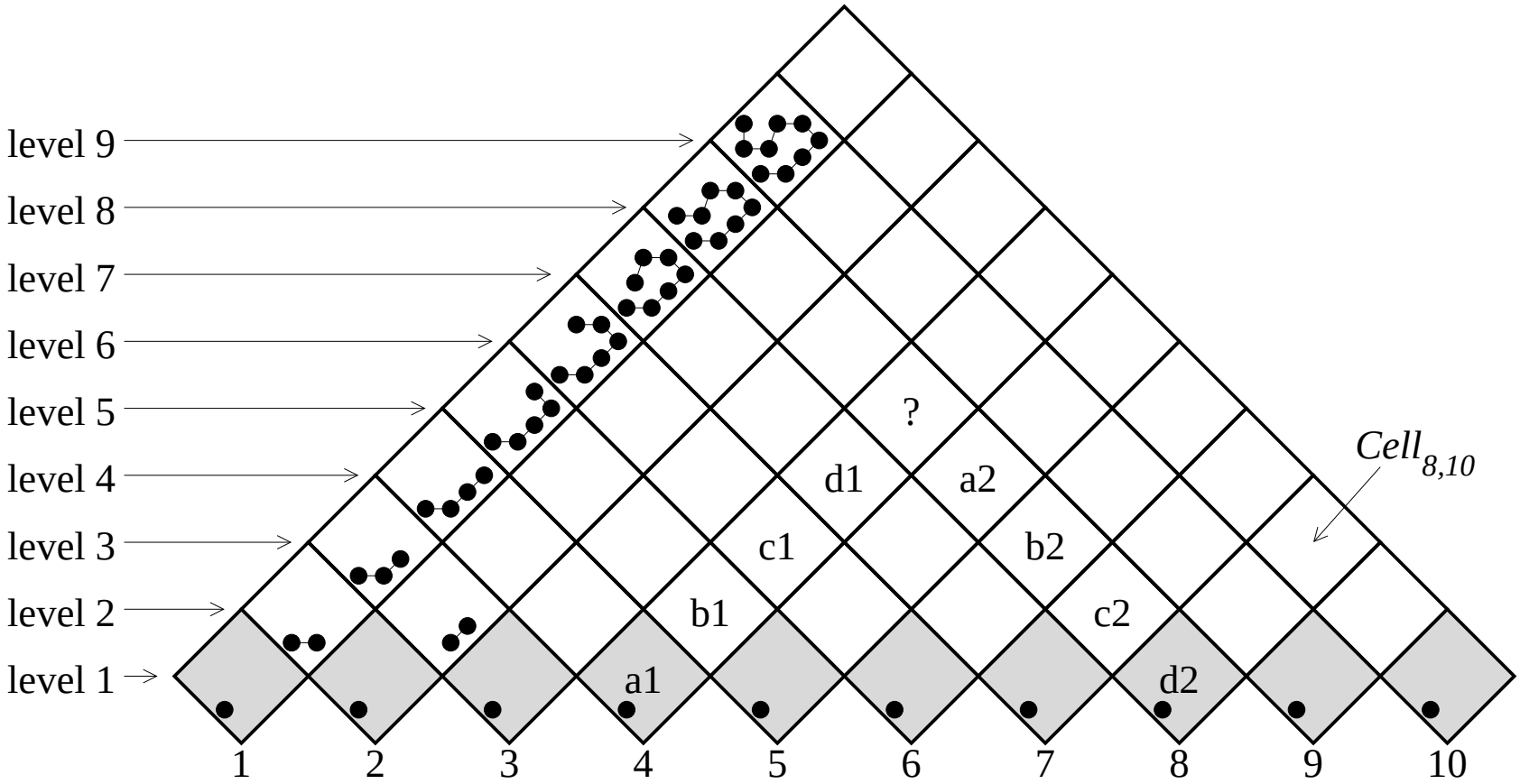
Zipping and assembly of a protein with 10 residues (7 residues)



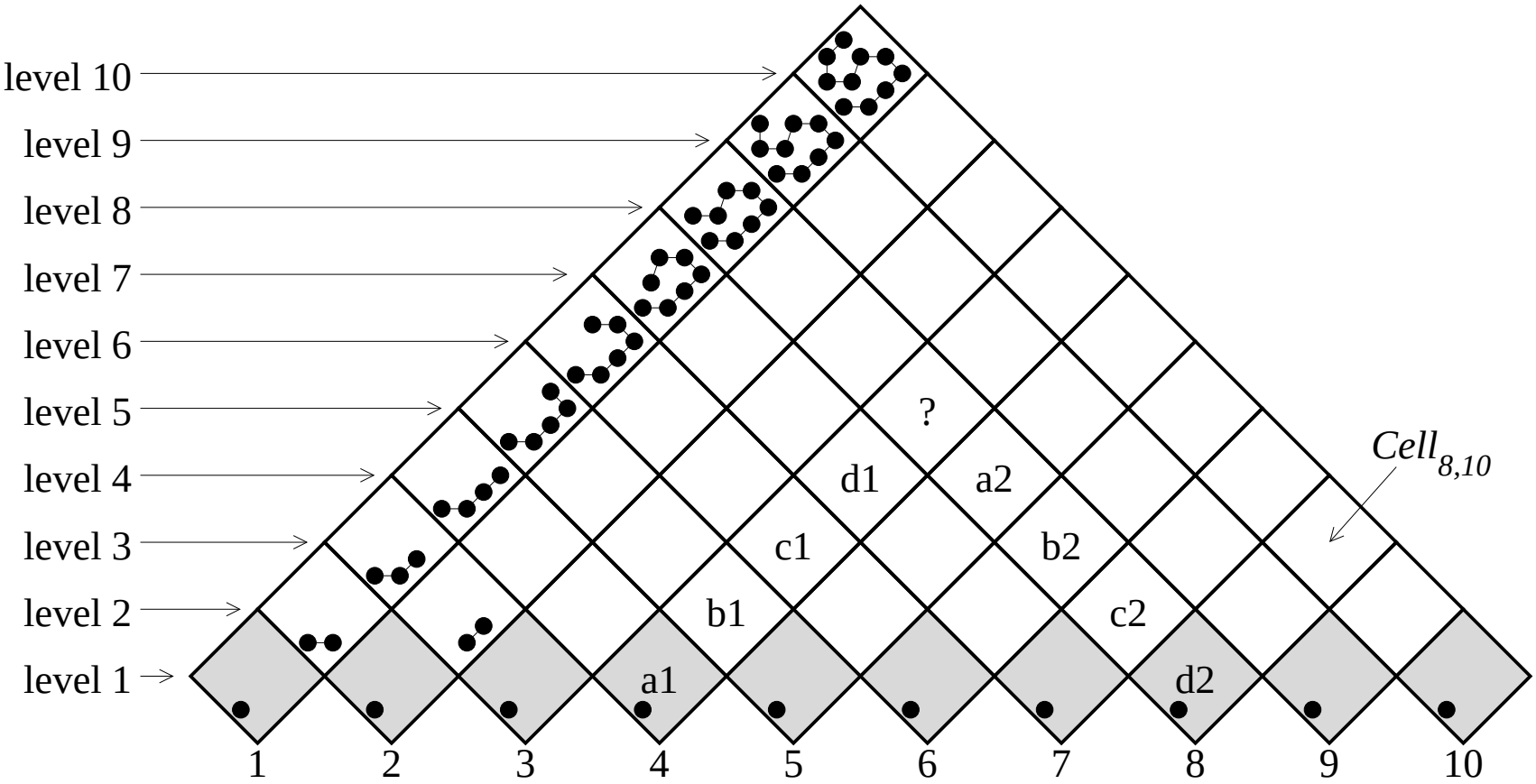
Zipping and assembly of a protein with 10 residues (8 residues)



Zipping and assembly of a protein with 10 residues (9 residues)



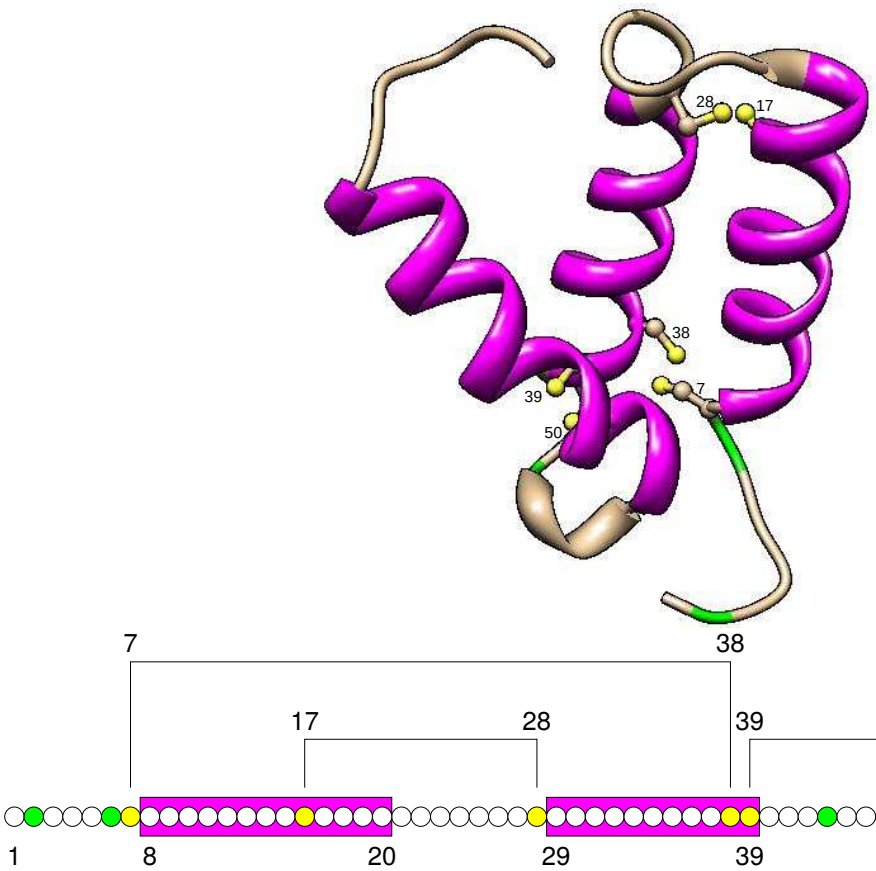
Zipping and assembly of a protein with 10 residues (10 residues)



Overview

- Aims of the project
- Protein structure basics
- Types of constraints
- Zipping and assembly method (dynamic programming) for generating solutions
- **Examples**
- Conclusions and to-do list

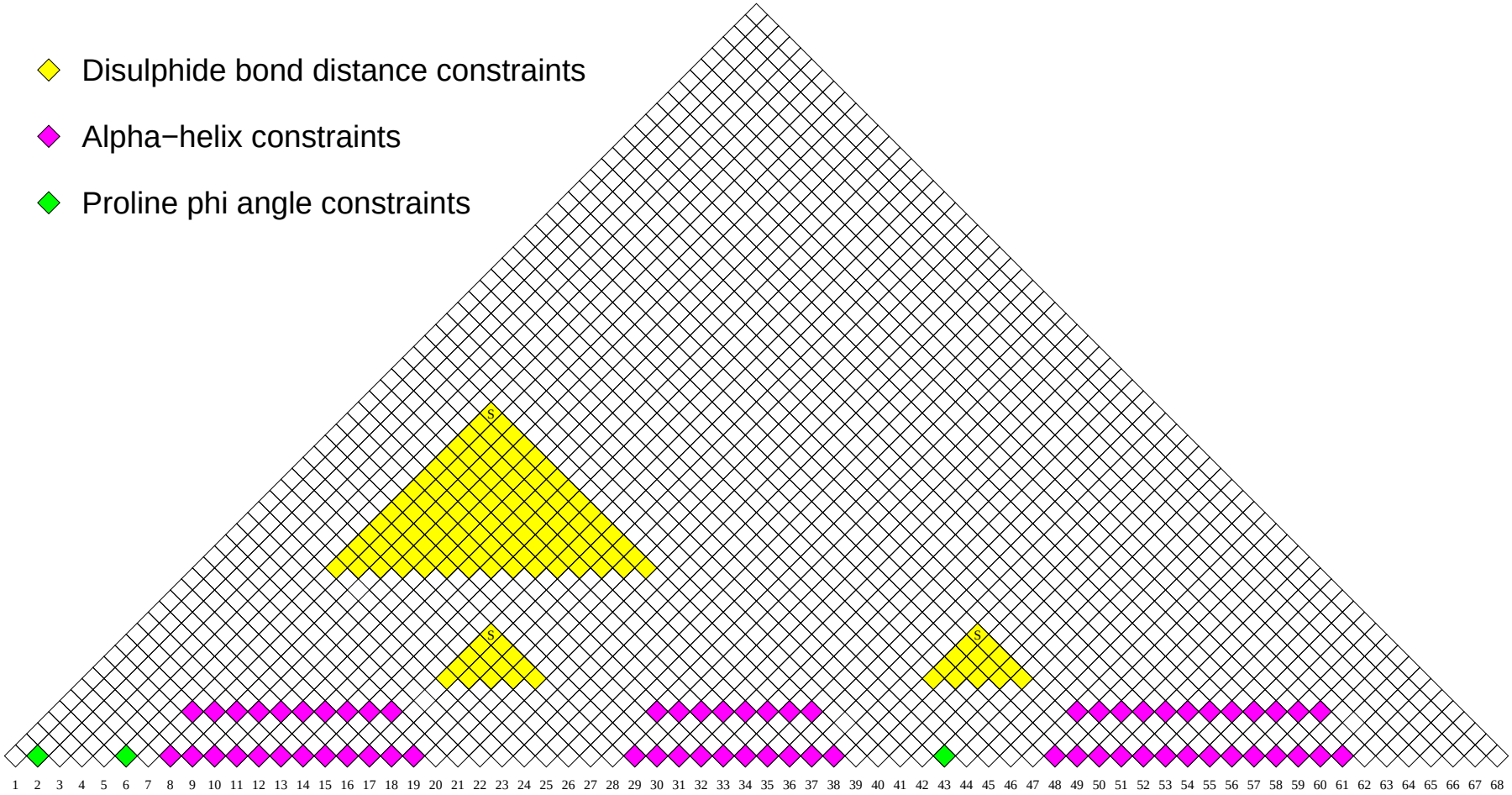
human p8MTCP1 [PDB entry: 2HP8]



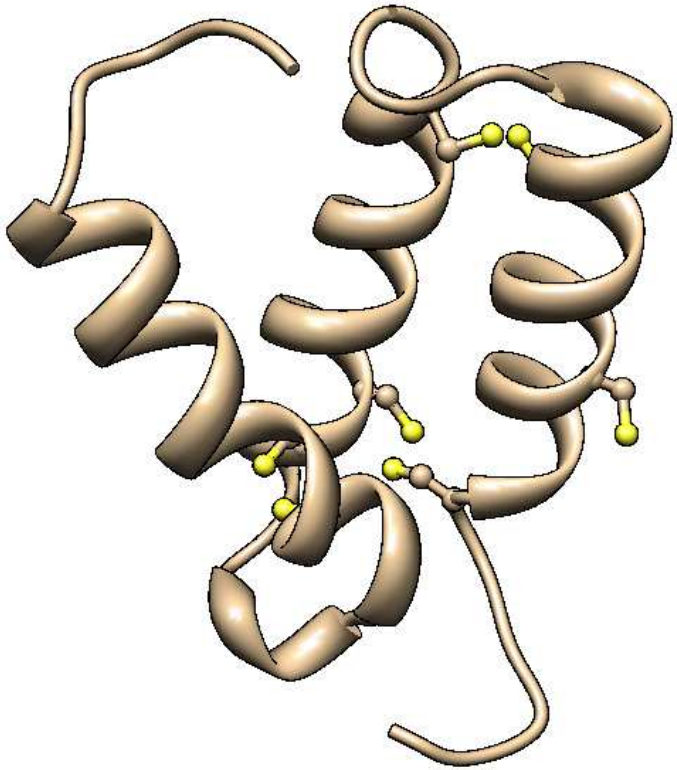

```
residue(1,'MET').
residue(2,'PRO').
residue(3,'GLN').
residue(4,'LYS').
residue(5,'ASP').
residue(6,'PRO'). % etc.

disulphide_bond(7,38).
disulphide_bond(17,28).
disulphide_bond(39,50).

alpha_helix(8,20).
alpha_helix(29,39).
alpha_helix(48,62).
```

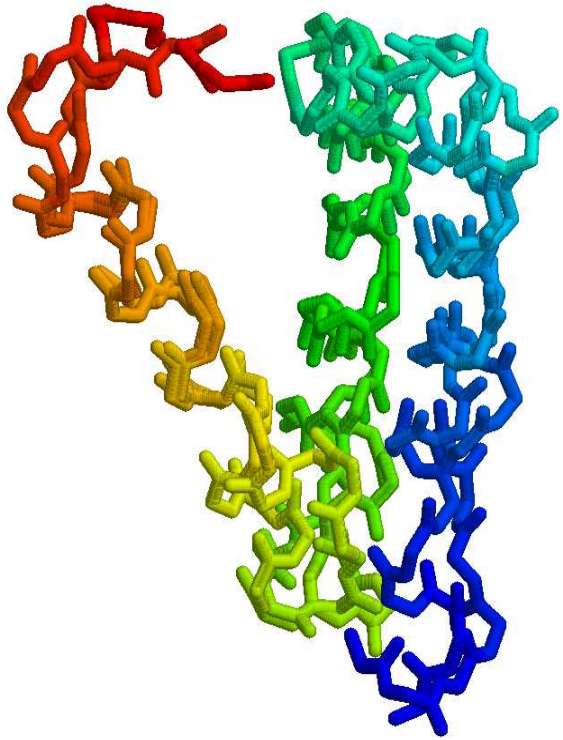
Constraints used in modelling human p8MTCP

- Disulphide bond distance constraints
- Alpha−helix constraints
- Proline phi angle constraints

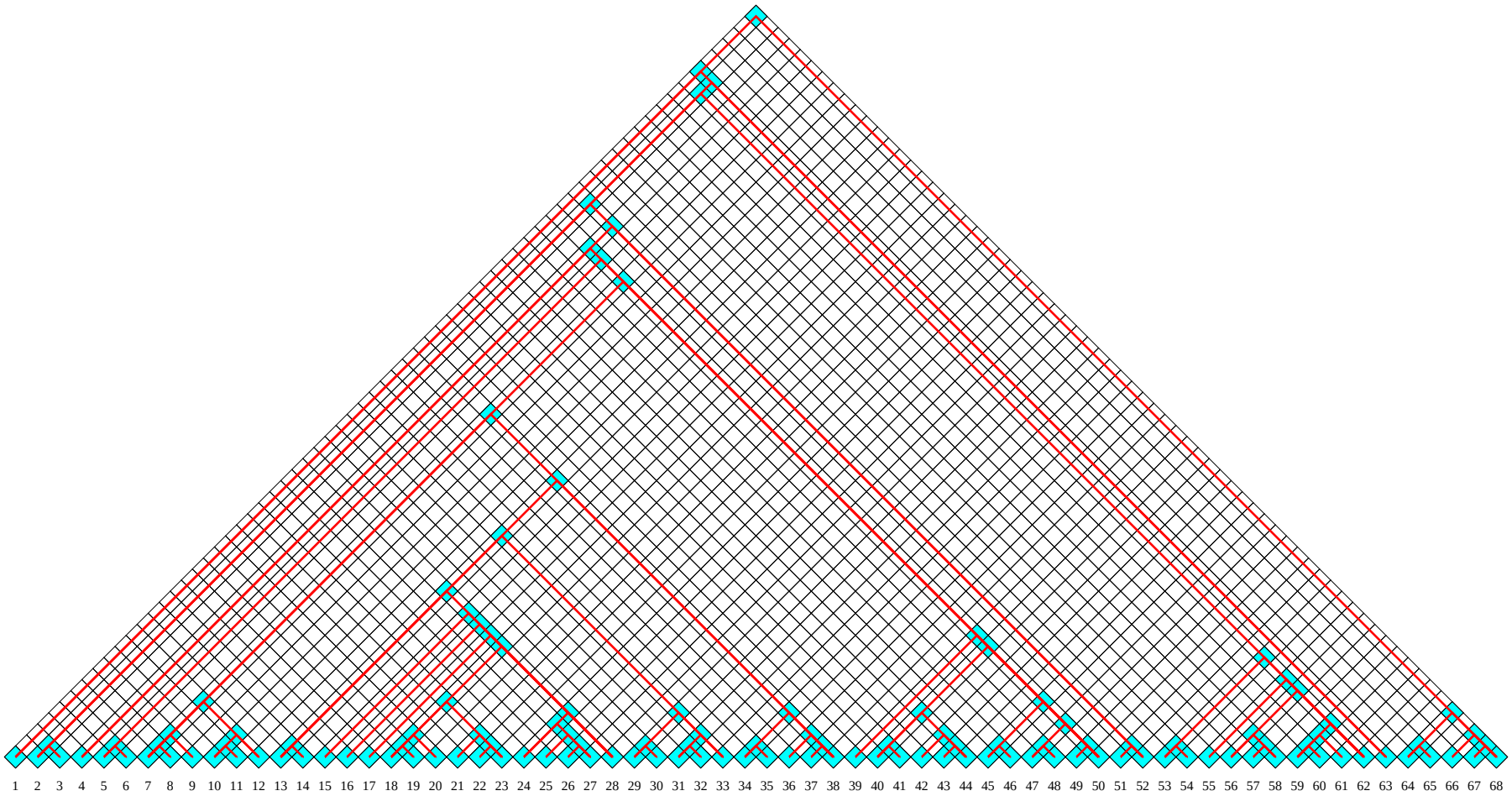


Best model and experimental structure (Cα RMSD: 2.6 Å)

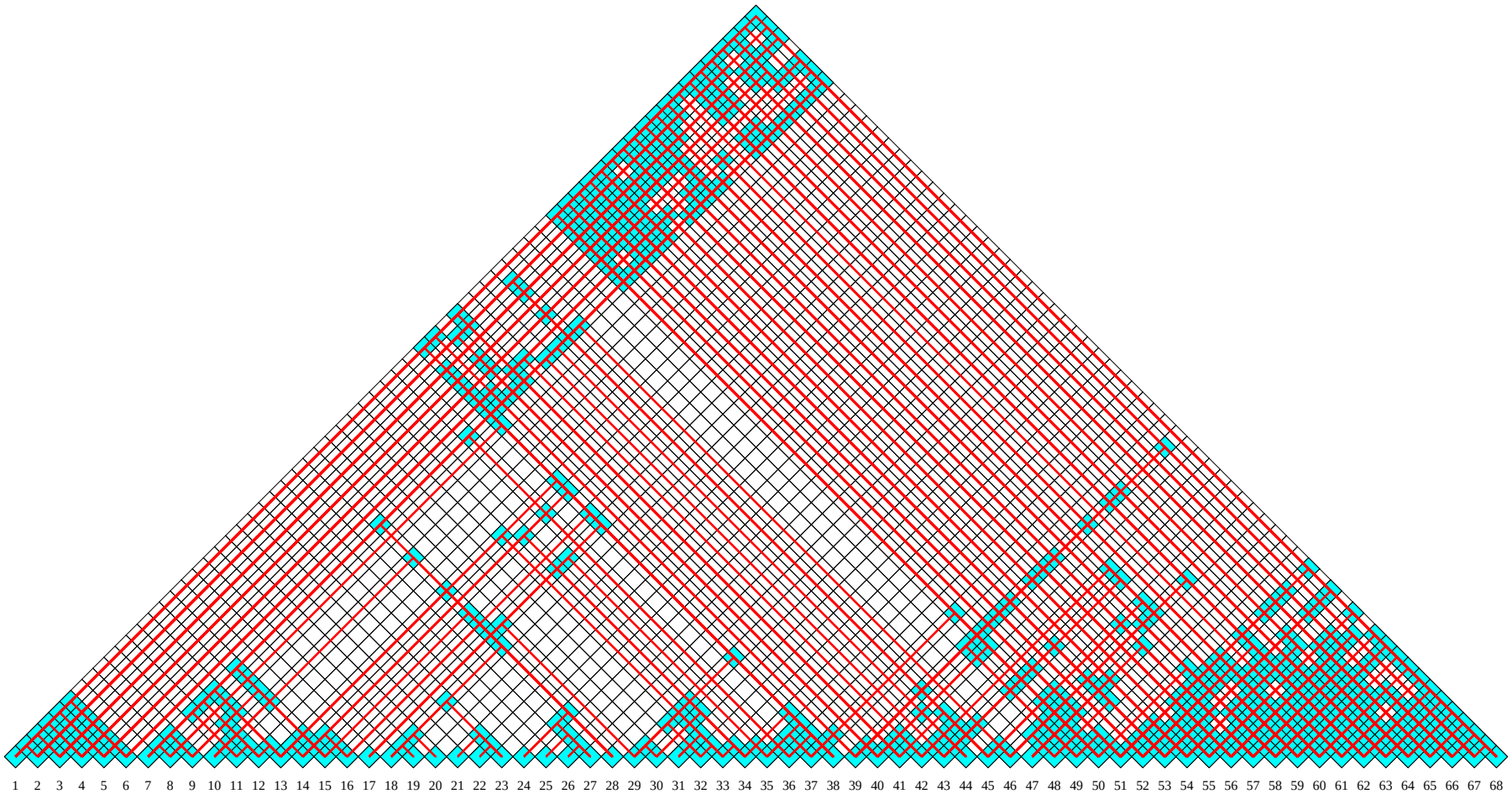




Actual cells used for constructing 1 protein model

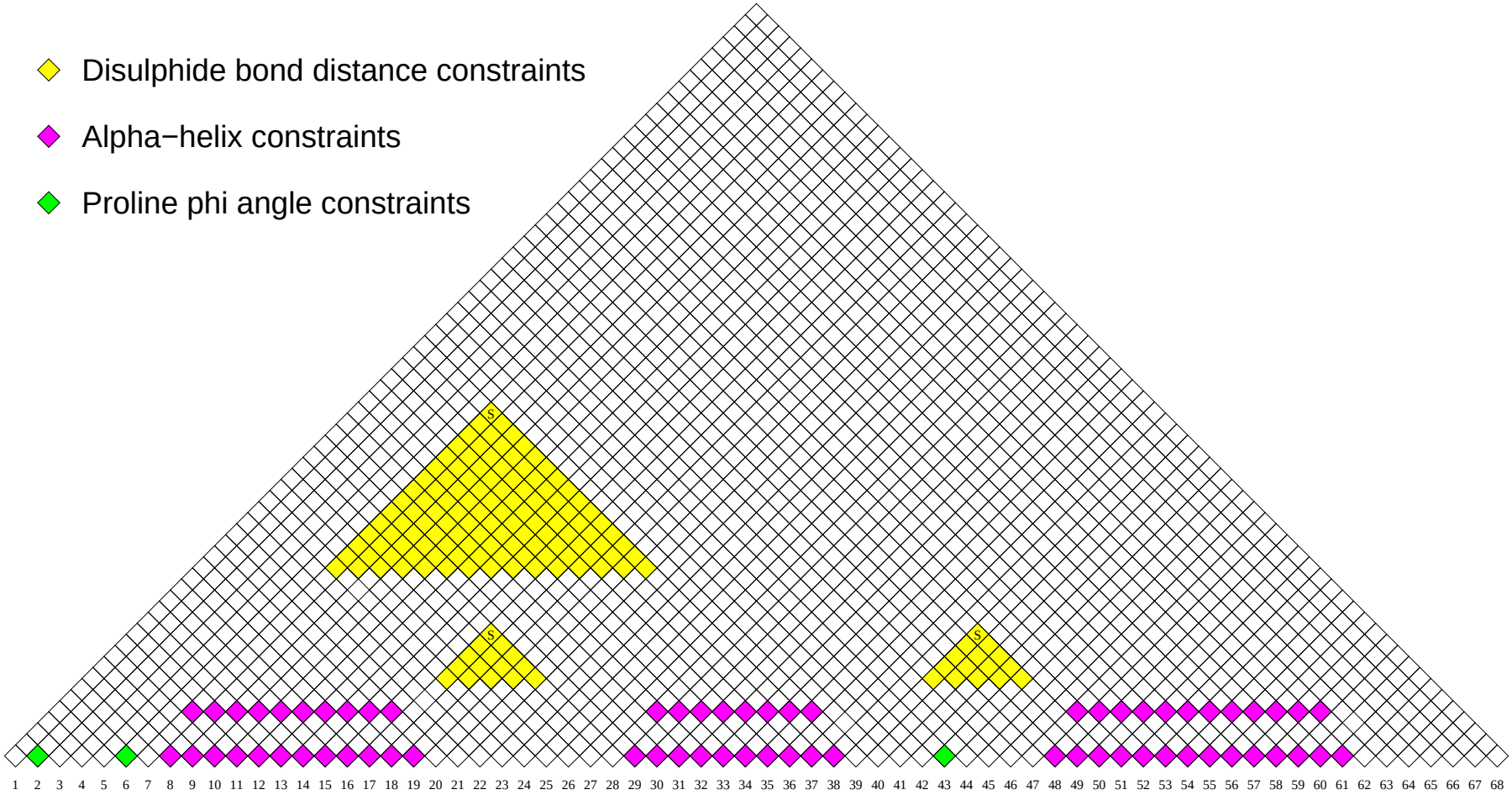


Actual cells used for constructing 50 protein model

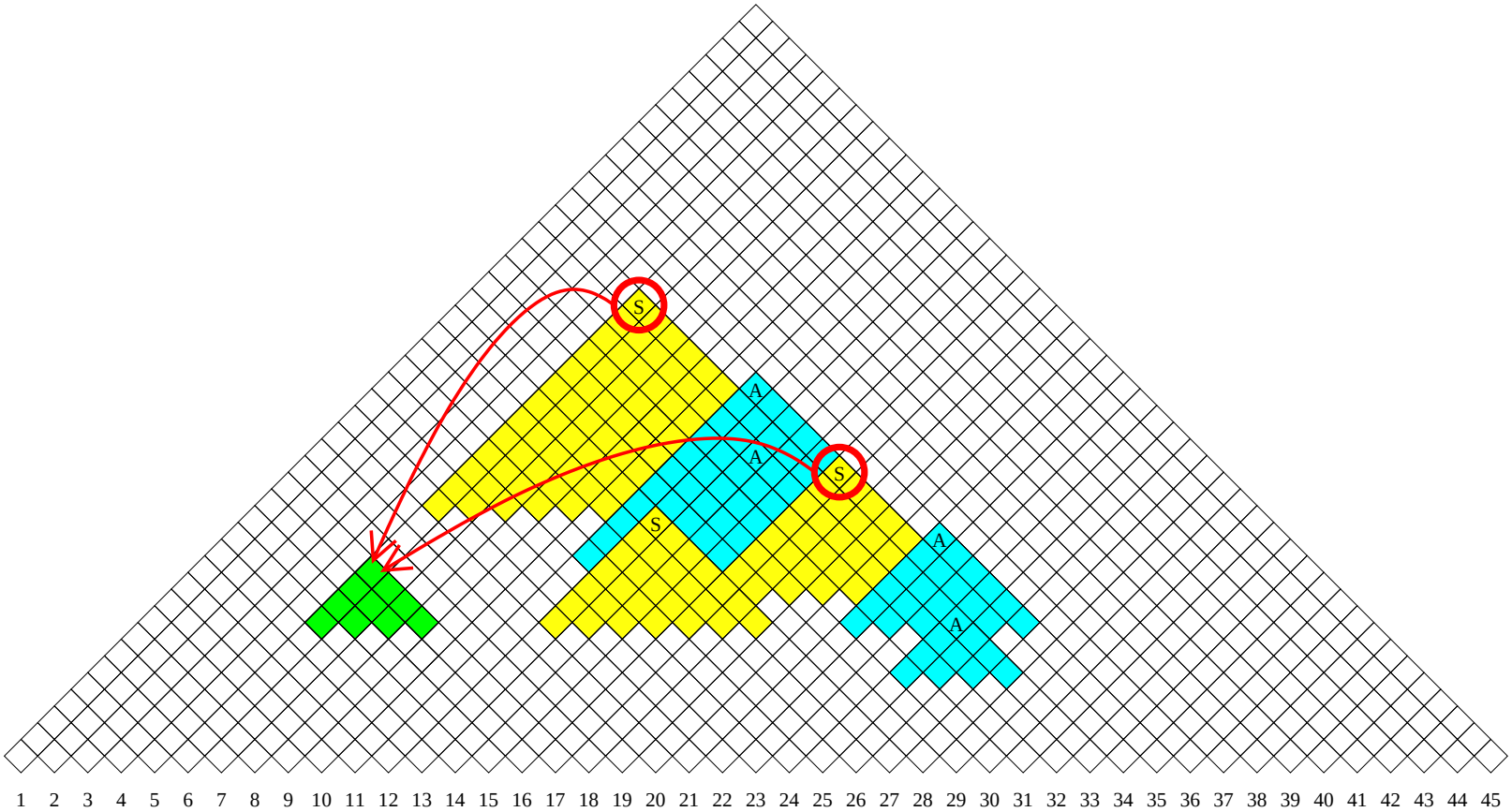


Constraints used in modelling human p8MTCP

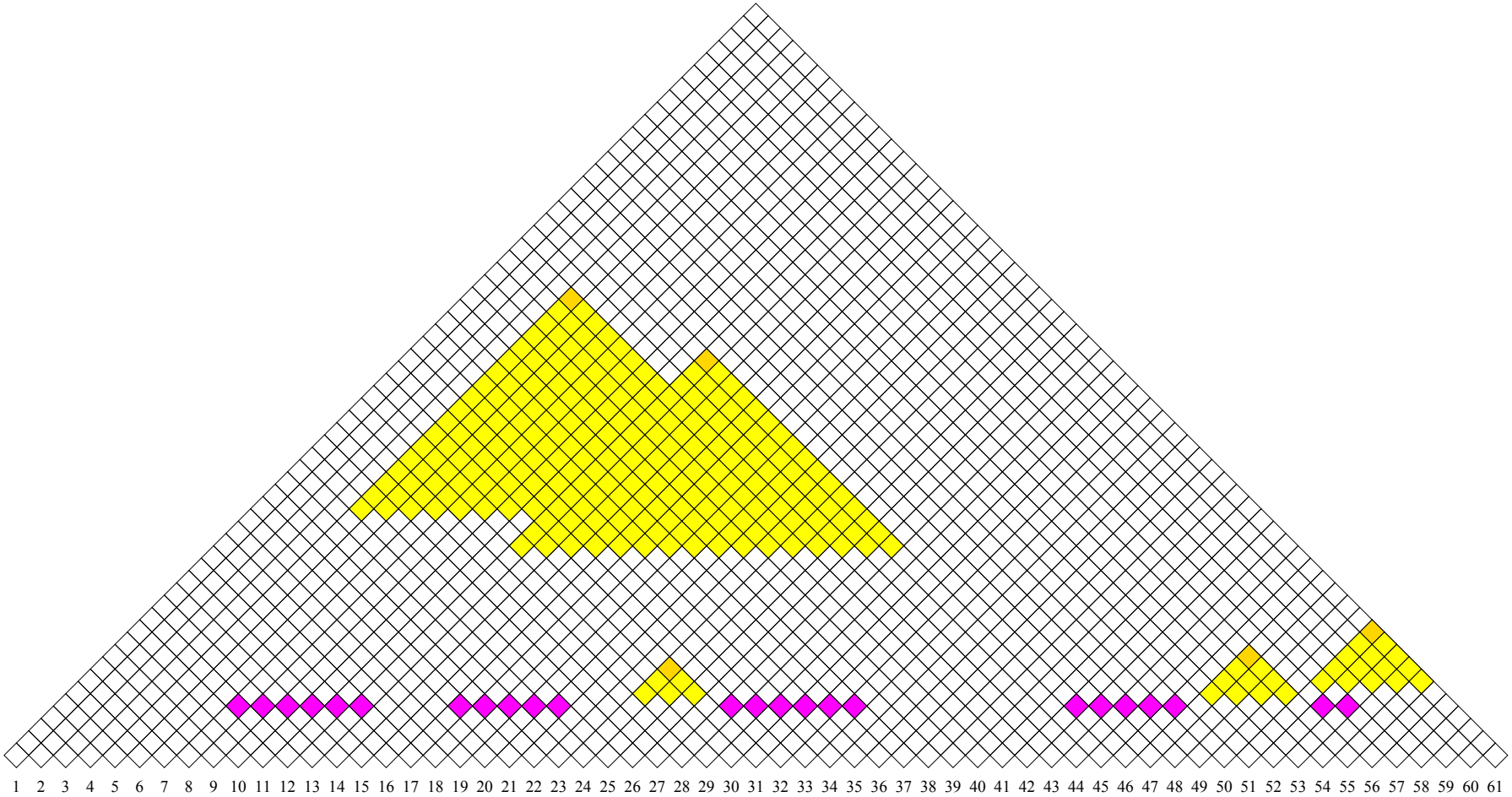
- Disulphide bond distance constraints
- Alpha−helix constraints
- Proline phi angle constraints



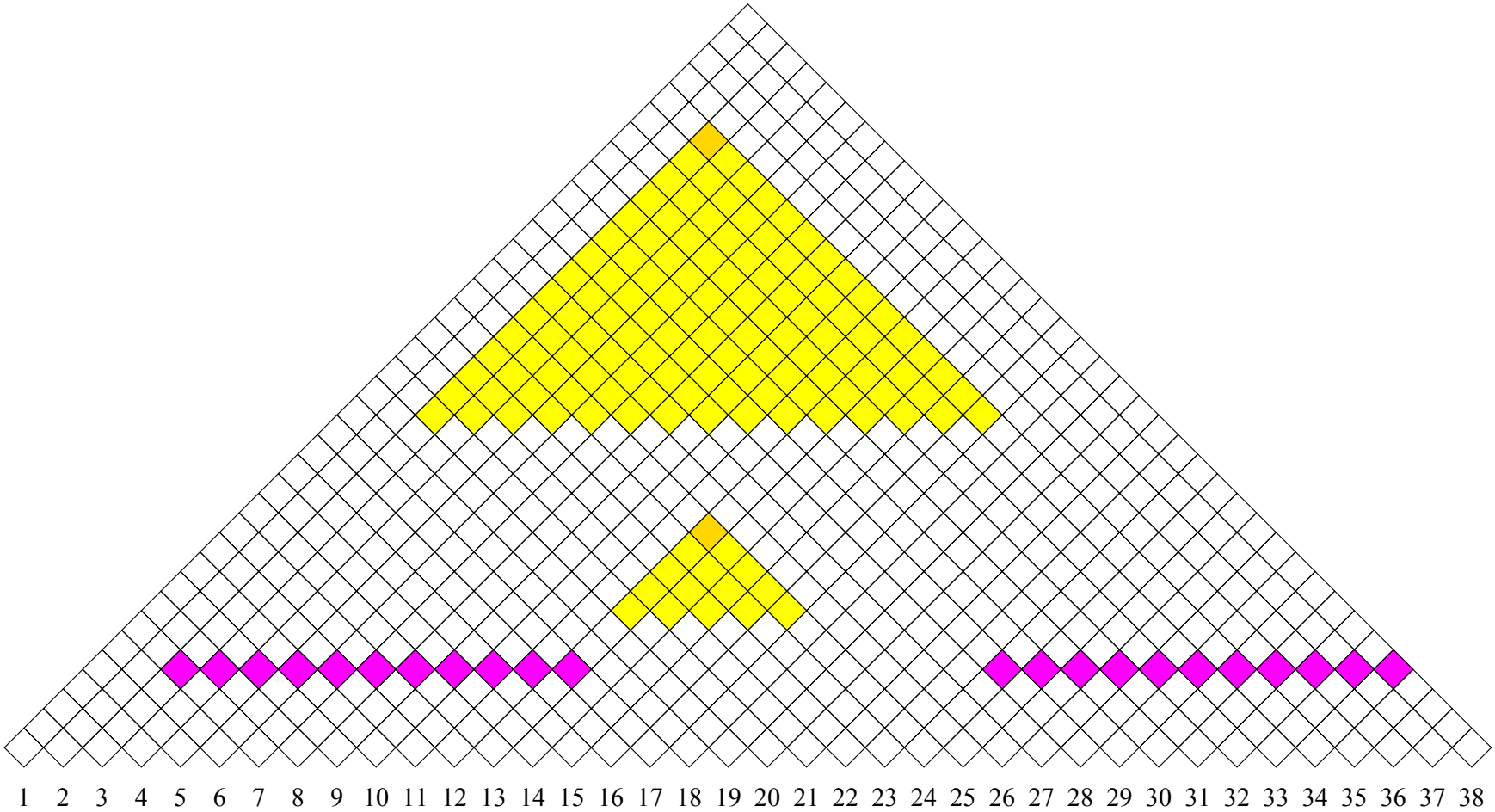
Constraints used in modelling 2LWL



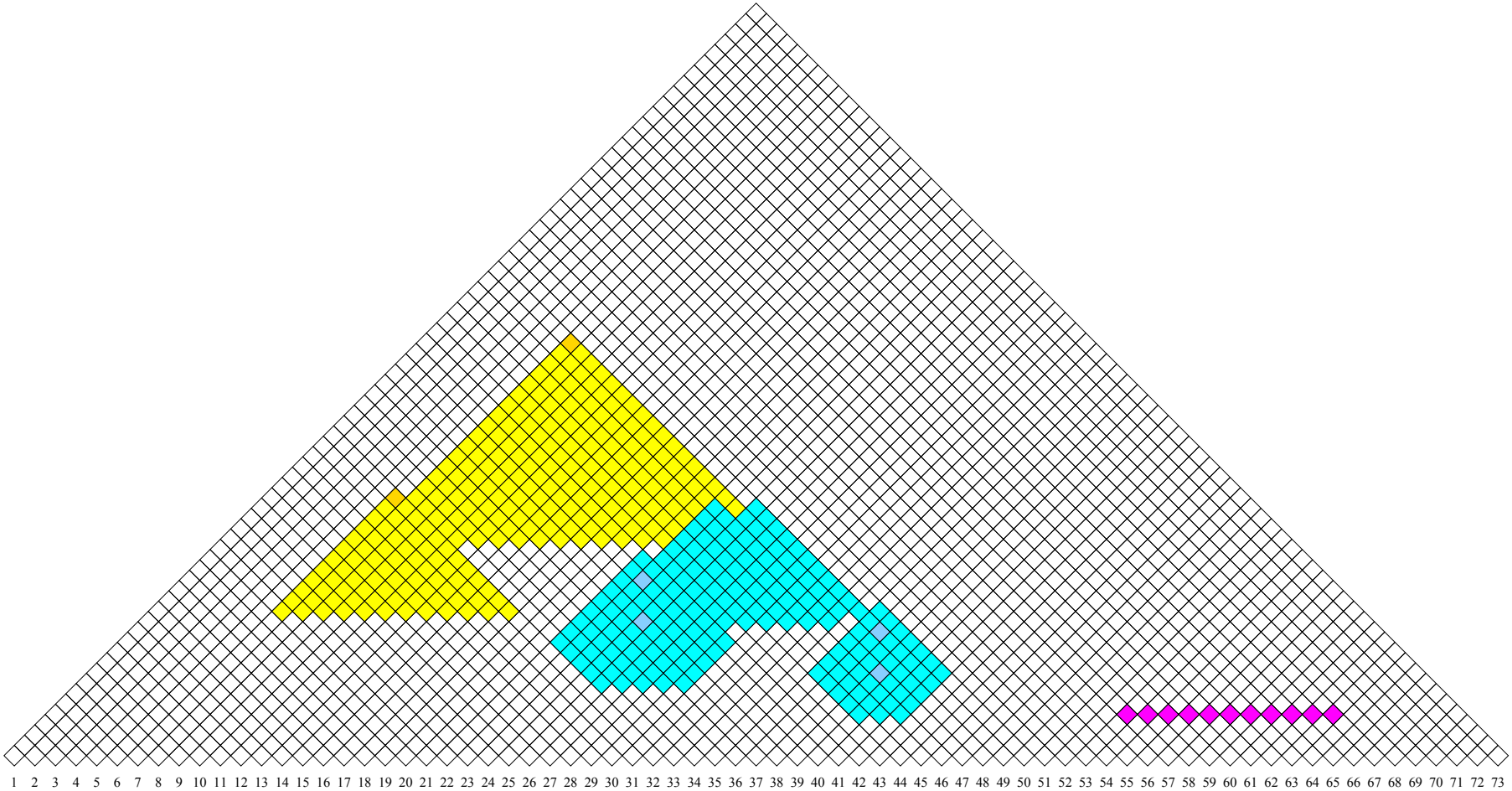
Constraints used in modelling 2LRD



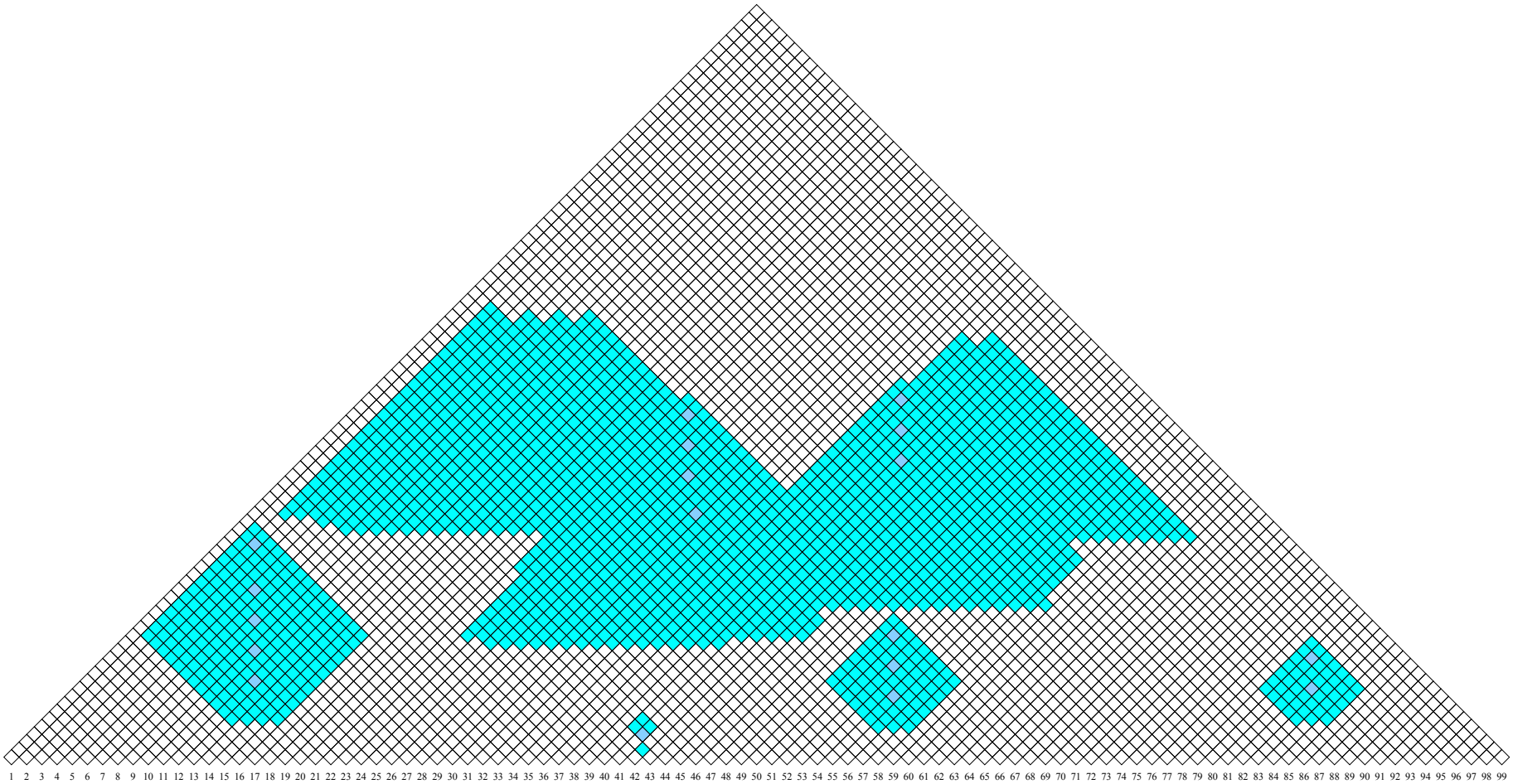
Constraints used in modelling 1EI0



Constraints used in modelling 1EIG



Constraints used in modelling 5CKA



Overview

- Aims of the project
- Protein structure basics
- Types of constraints
- Zipping and assembly method (dynamic programming) for generating solutions
- Examples
- **Conclusions and to-do list**

Conclusion and to-do list

- **Conclusion:**
  - Our program is able to produce protein models
  - Not all the cells contribute to the final structures
- **To-do list:**
  - Use longer fragments
  - Make better use of memory
  - Use scoring function for filtering and ranking

Acknowledgements

We are grateful for support from Vetenskapsrådet