

Combinatorial Optimization for Life Sciences

Simon de Givry
SaAB team, MIAT – INRA
Toulouse, France

November 7th 2017

Unité Mathématiques et Informatique Appliquées - Toulouse

Directeur d'unité : Sylvain Jasson

Équipe administrative

Fabienne Aygnac (TR)
Nathalie Julliard (TR)
Alain Pérault (TR)

Équipe collective (ID & AU)



Équipe informatique

Damien Berry (AI)
Mikaël Grialou (TR)



Plateforme

Titulaires

Estelle Ancelet (IE)
Eric Casellas (IE)
Patrick Chabrier (IE)
Hélène Raynal (IR)
Nathalie Rousse (IR)
Ronan Trepos (IR)
Marie-Stéphane Trotard (IE)

CDD

Mohamed El Hadramy (IE)
Guillaume Robaldo (IE)



Équipe MAD
Modélisation des
Agro-écosystèmes et Décision

Titulaires

Stéphane Couture (CR)
Marie-Josée Cros (IE)
Céline Delmas (CR)
Nathalie Dubois-Peyrard (DR)
Robert Faivre (DR)
Sylvain Jasson (IR)
Roger Martin-Clouaire (DR)
Victor Picheny (CR)
Gauthier Quesnel (CR)
Régis Sabbadin (DR)
Patrick Taillandier (CR)

Doctorants

Etienne Auclair
Mathieu Bourgaïs
David Gaudrie
Sebastian Le Coz
Léonard Torossian
Hui Xiao

Post doctorants

Yann Dujardin



Équipe SaAB
Statistiques et Algorithmique
pour la Biologie

Titulaires

David Allouche (IR)
Hélène Chiapello (IR)
Christine Gaspin (DR)
Simon de Givry (CR)
Georges Katsirelos (CR)
Damien Leroux (IE)
Annick Moisan (IE)
Thomas Schiex (DR)
Nathalie Villa-Vialaneix (CR)
Matthias Zytnicki (CR)

CDD

Ignacio Gonzalez Fuentes (IR)
Gaëlle Lefort (IE)
Ludovic Mallet (IR)

Doctorants

Walid Ben Saoud Benjerri
Franck Cerutti
Alyssa Imbert
Sara Maqrot
Jérôme Mariette
Manon Ruffini
Clément Viricel
Jelena Vučinić

Post-doctorants

Lise Pomiès



Plateforme

Titulaires

Christine Gaspin (DR)
Claire Hoede (IR)
Christophe Klopp (IR)
Didier Laborie (IE)
Jérôme Mariette (IE)
Céline Noirot (IE)
Marie-Stéphane Trotard (IE)

CDD

Floréal Cabanettes (IE)
Léo Lamy (IE)



Plateforme

Titulaires

Christophe Klopp (IR)

PLAN

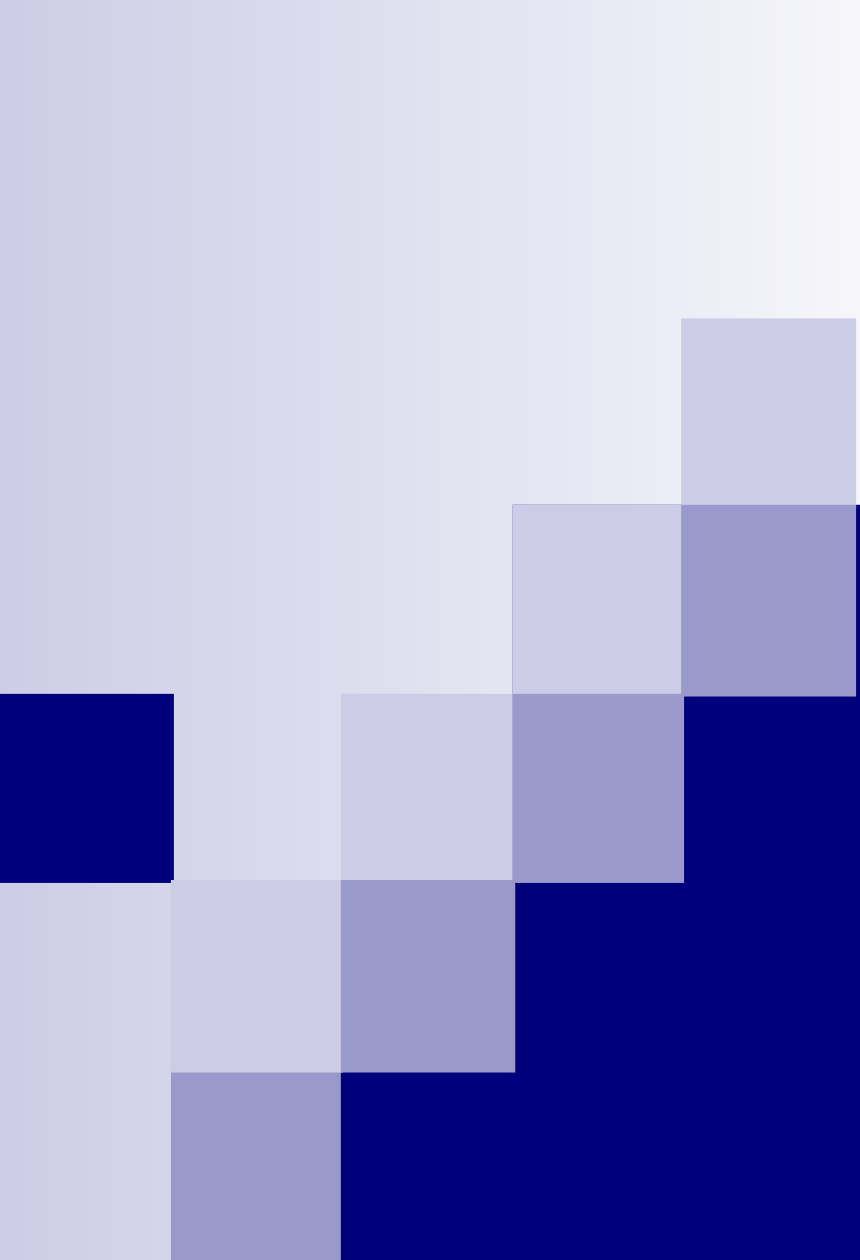
Constraint Optimization Problems expressed as
Cost Function Networks
aka Weighted Constraint Satisfaction Problem (WCSP)

► Applications in **Genetics**

- Mendelian error detection in complex pedigrees
- *Haplotype reconstruction in half-sib families*
- *TagSNP selection*
- *Optimizing the reference population in a genomic selection design*

► Applications in **Bioinformatics**

- *Searching RNA motifs and their intermolecular contacts*
 - Computational Protein Design
 - *Genome scaffolding*
 - *Gene regulatory network reconstruction using bayesian networks*
- 



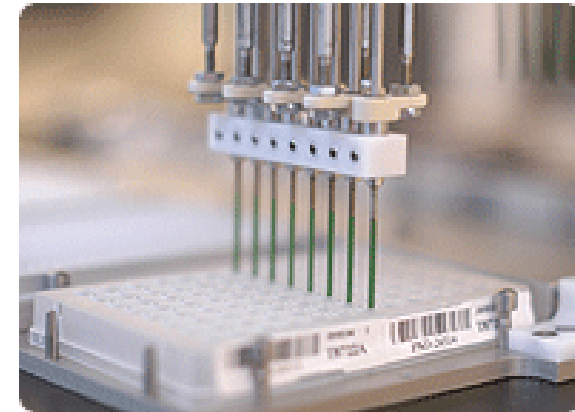
Mendelian error detection in complex pedigree

Simon de Givry, Marti Sanchez, Isabelle Palhière, Zulma Vitezica, and Thomas Schiex
INRA, Toulouse, France

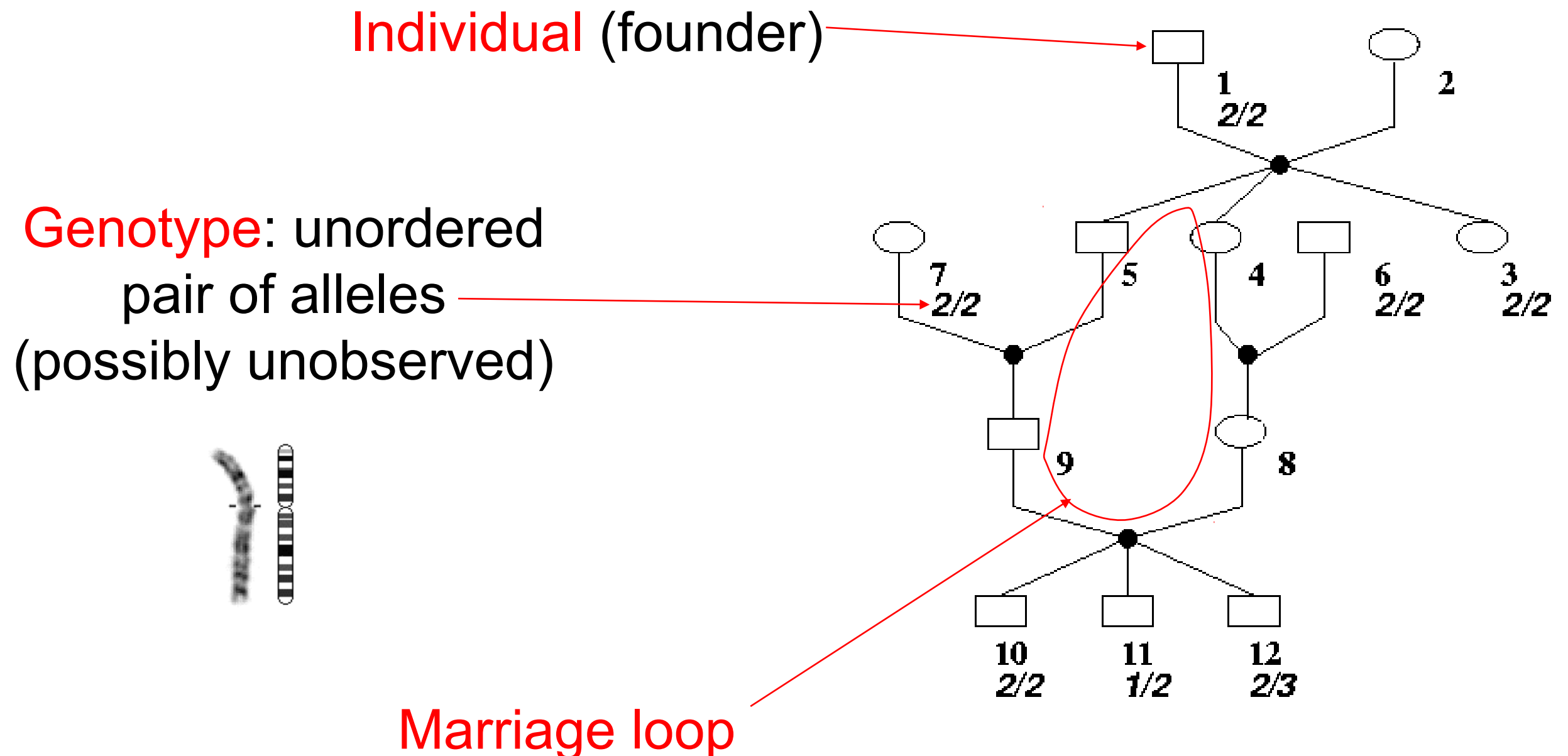
WCB 2005, WCGALP 2006, Constraints 2008

Cleaning the data after genotyping

Today, about 1% errors remain after SNP genotyping



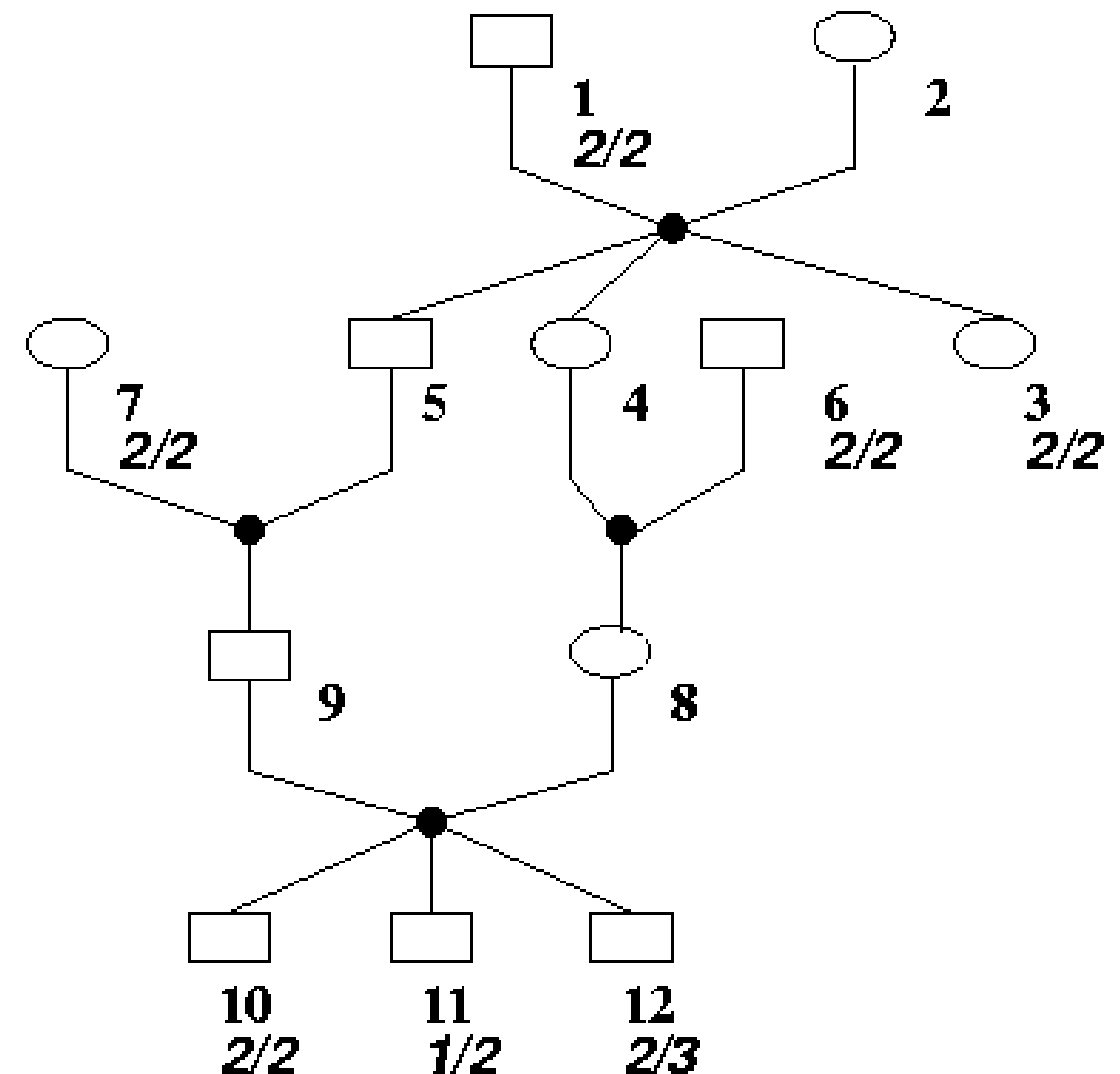
✉ For each marker, detects Mendelian inheritance errors



Task 1: Consistency Checking

- Assuming the pedigree is correct, checks if it exists a **complete genotype assignment** consistent with the observed genotypes and with the Mendelian laws of inheritance

- Complexity results (Aceto et al., 2003)
 - NP-complete for a pedigree with loops and more than three alleles
 - Polynomial if no loops or just two alleles

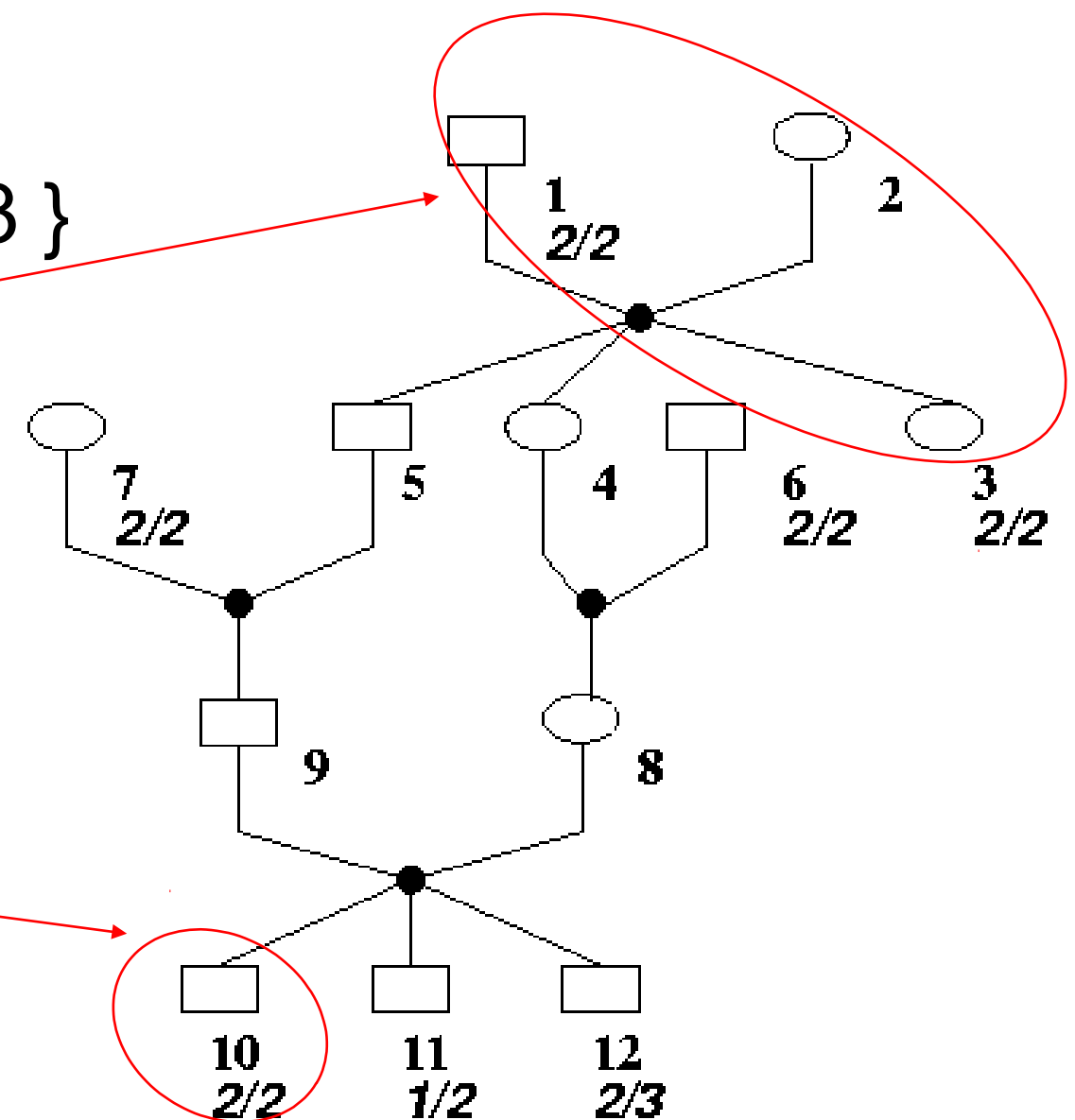


Constraint Satisfaction Problem

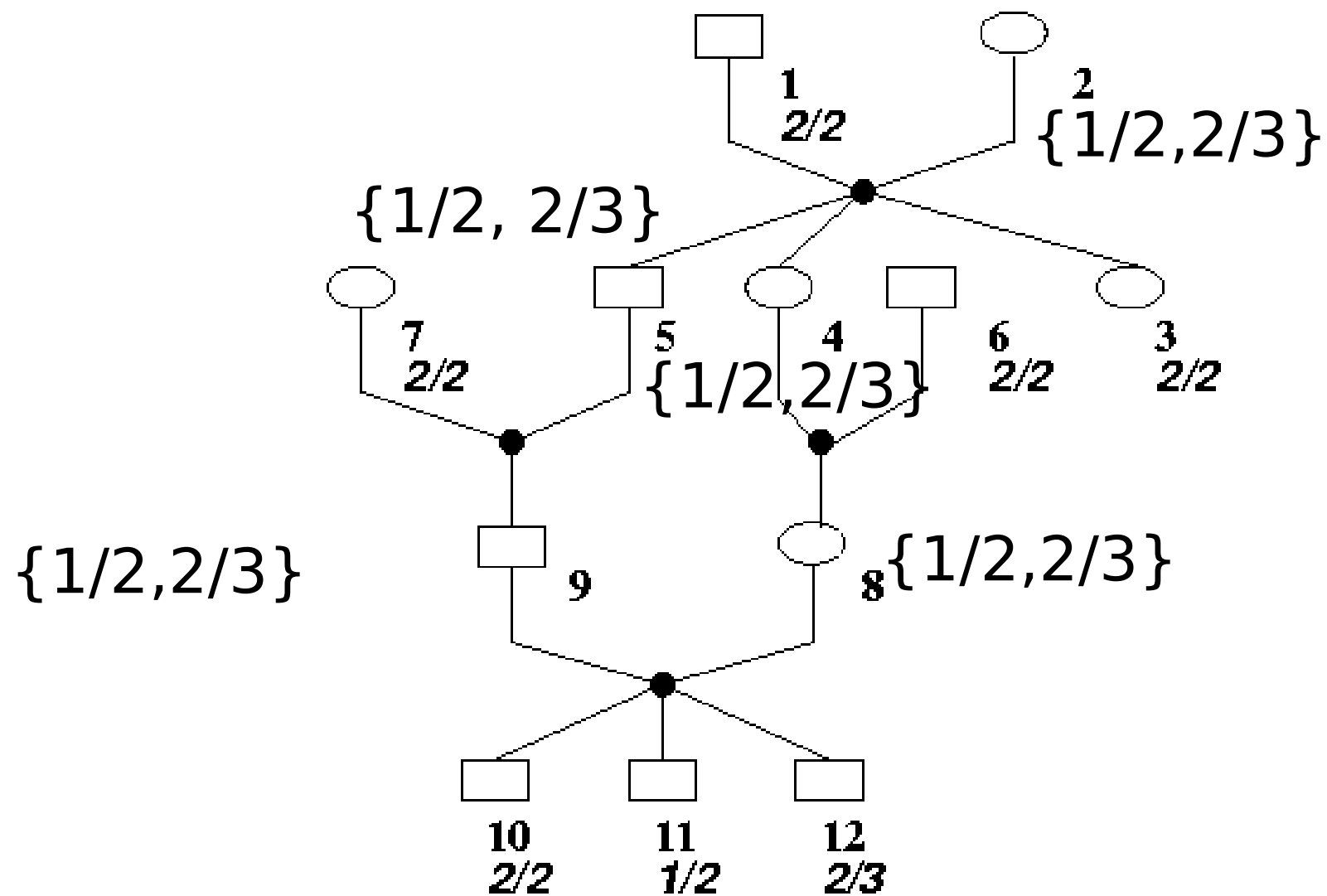
- **X**: one variable per individual
- **D**: domain of every variable is defined as the set of all possible genotypes

Here: $\{ 1/1, 1/2, 1/3, 2/2, 2/3, 3/3 \}$

- **C**:
 - Ternary constraints to encode Mendelian laws for any non founder
 - Unary constraints to encode genotyping data



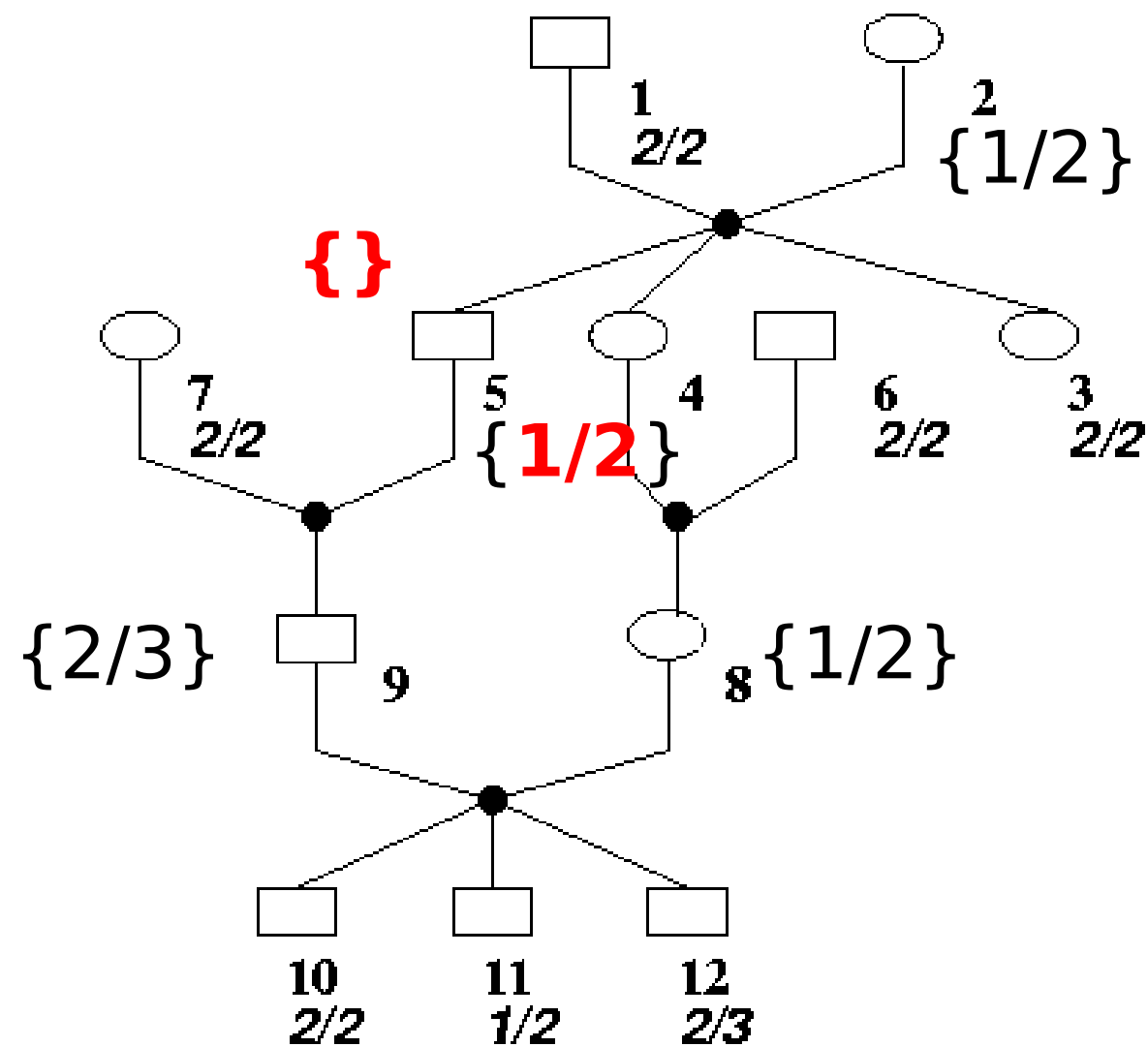
Generalized Arc Consistency on nuclear family



(Mackworth, AIJ 1977)

(Lange, Goradia, Am J Hum Genet 1987)

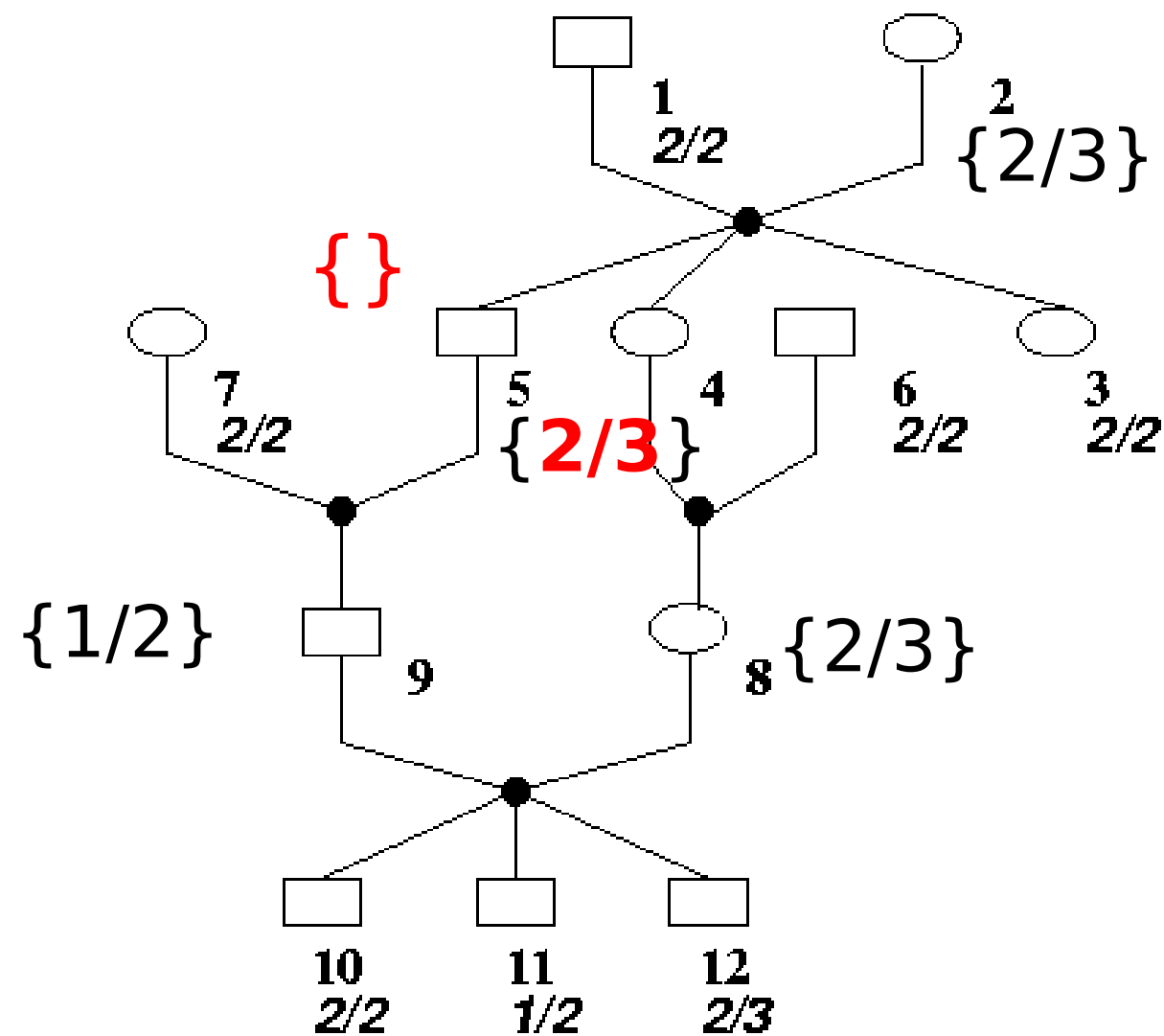
Backtrack search on loop-breaker individuals



(Dechter, AIJ 1990)

(O'Connell, Weeks, Am J Hum Genet 1997)

Backtrack search on loop-breaker individuals

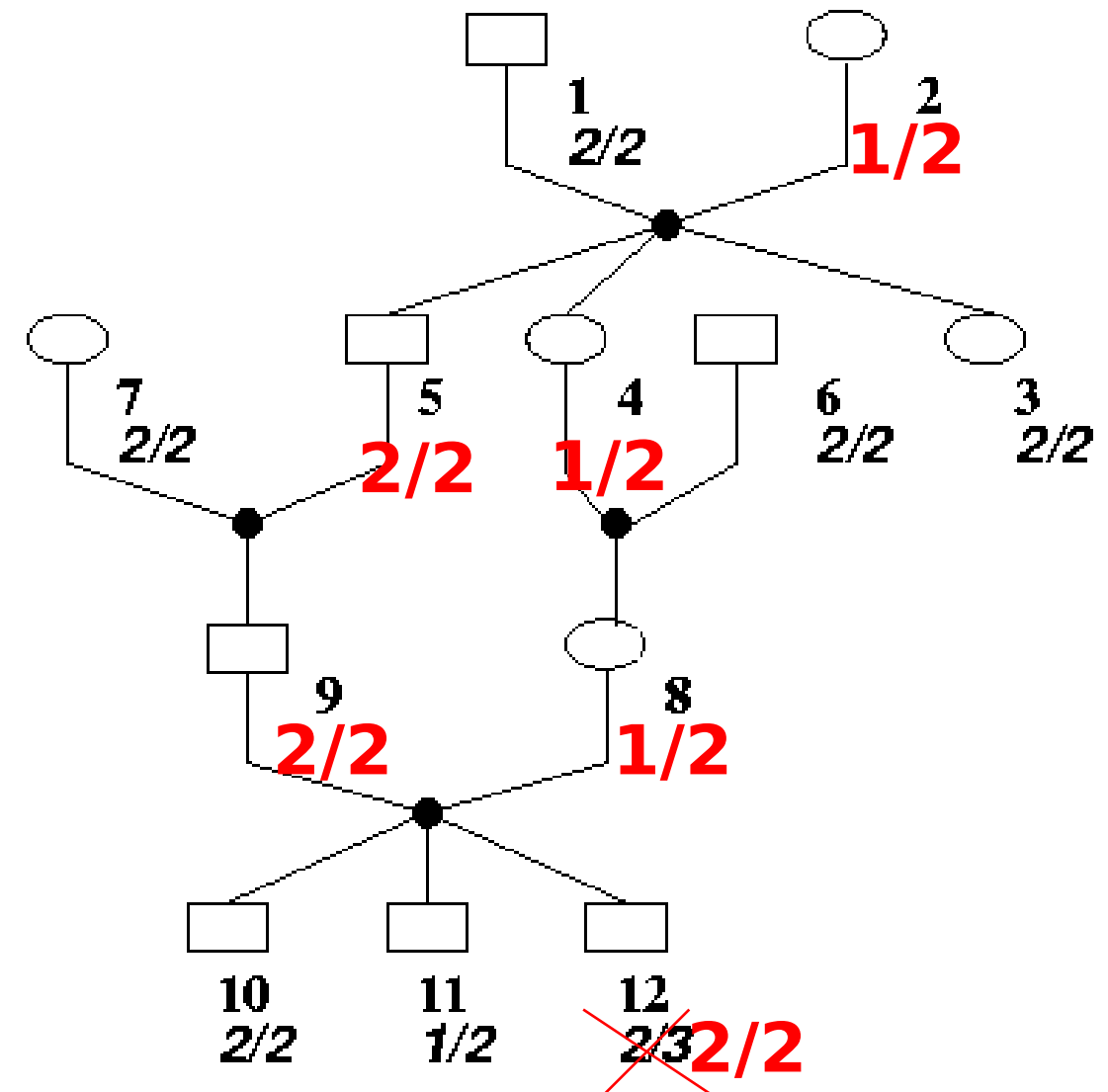


(Dechter, AIJ 1990)

(O'Connell, Weeks, Am J Hum Genet 1997)

Task 2: Error Detection

- Finds a complete assignment with the minimum number of errors
- ✉ Follows the parsimony principle



Weighted CSP (WCSP)

► (X, D, F)

◦ $X = \{X_1, \dots, X_n\}$ n variables

◦ $D = \{D_1, \dots, D_n\}$ n finite domains of maximum size d

◦ $F = \{f_{S_1}, \dots, f_{S_e}\}$, e cost functions

f_{S_i} : associates a finite or infinite (k) positive integer to every tuple in $I(S_i)$

► Goal: find a complete assignment A minimizing

$$\sum_{f_{S \in F}} f_S (A[S])$$

NP-hard problem

(Shapiro, Haralick, IEEE PAMI 81)

(Freuder, Wallace, AIJ 92)

(Schiex, Fargier, Verfaillie, IJCAI 95)

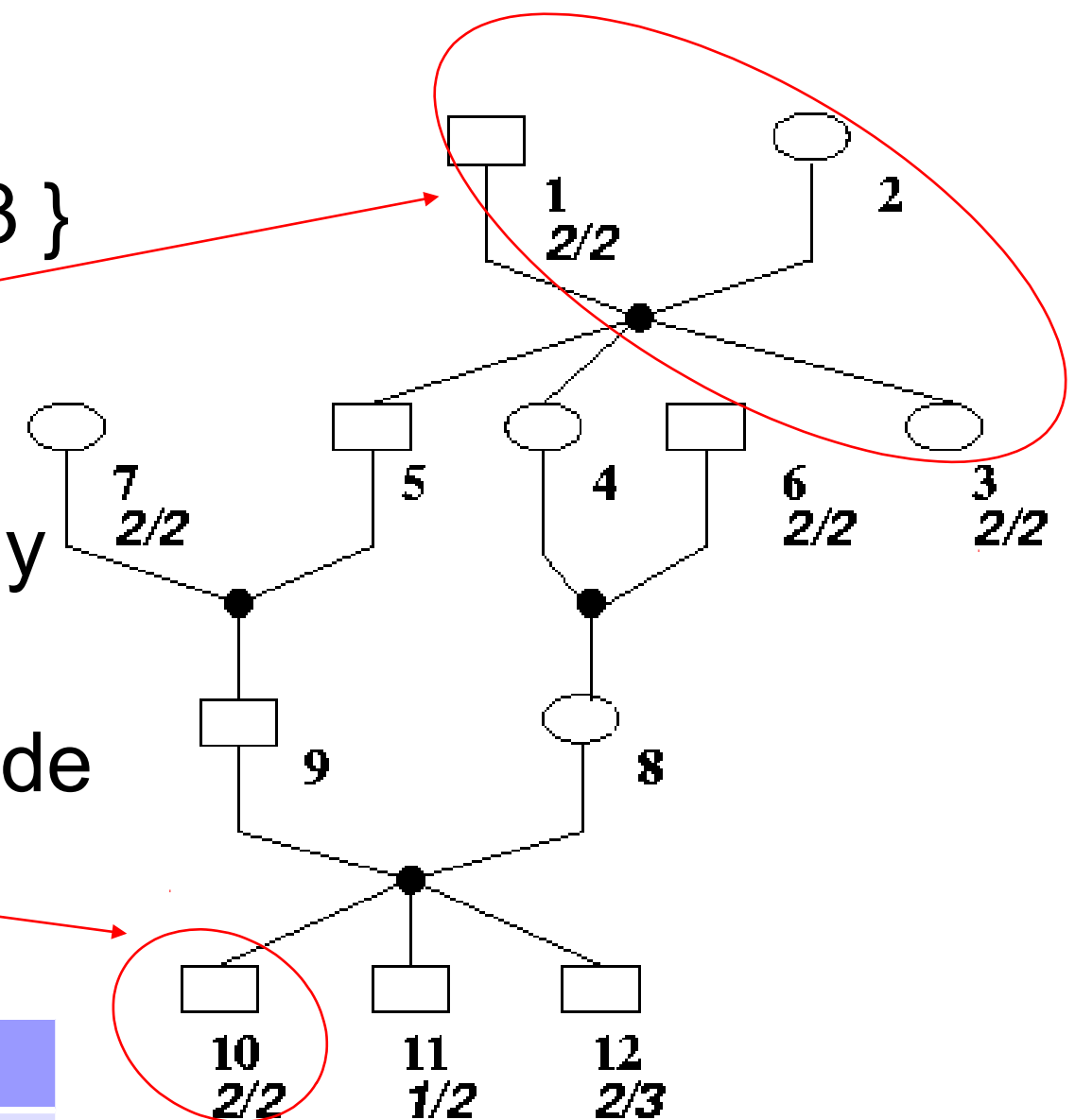
Weighted Constraint Satisfaction Problem

- **X**: one variable per individual
- **D**: domain of every variable is defined as the set of all possible genotypes

Here: $\{ 1/1, 1/2, 1/3, 2/2, 2/3, 3/3 \}$

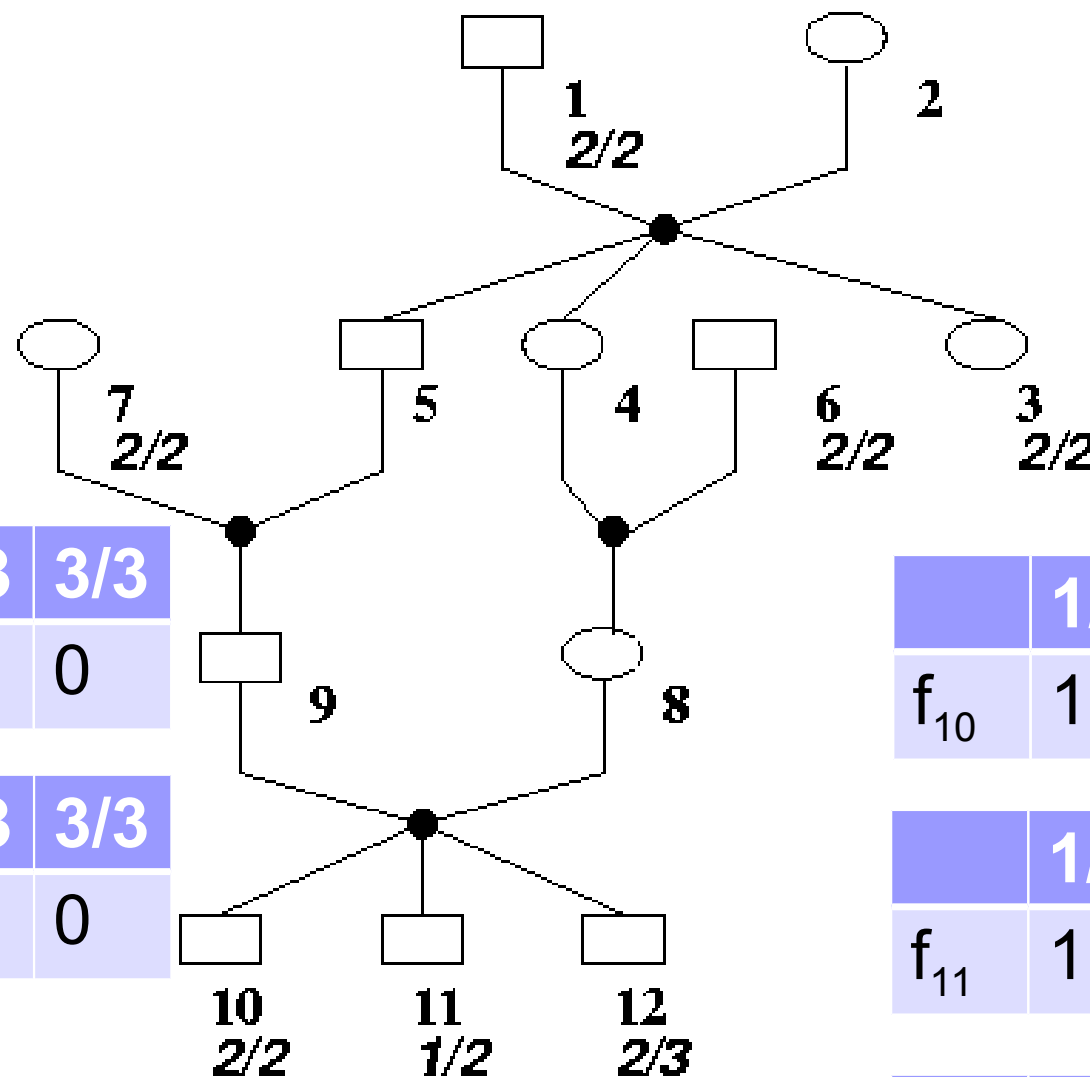
- **F**:

- Ternary **hard** constraints to encode Mendelian laws for any non founder
- Unary **soft** constraints to encode genotyping data



	1/1	1/2	1/3	2/2	2/3	3/3
f_{10}	1	1	1	0	1	1

Generalized **Soft** Arc Consistency



	1/1	1/2	1/3	2/2	2/3	3/3
f_8	0	0	0	0	0	0

	1/1	1/2	1/3	2/2	2/3	3/3
f_9	0	0	0	0	0	0

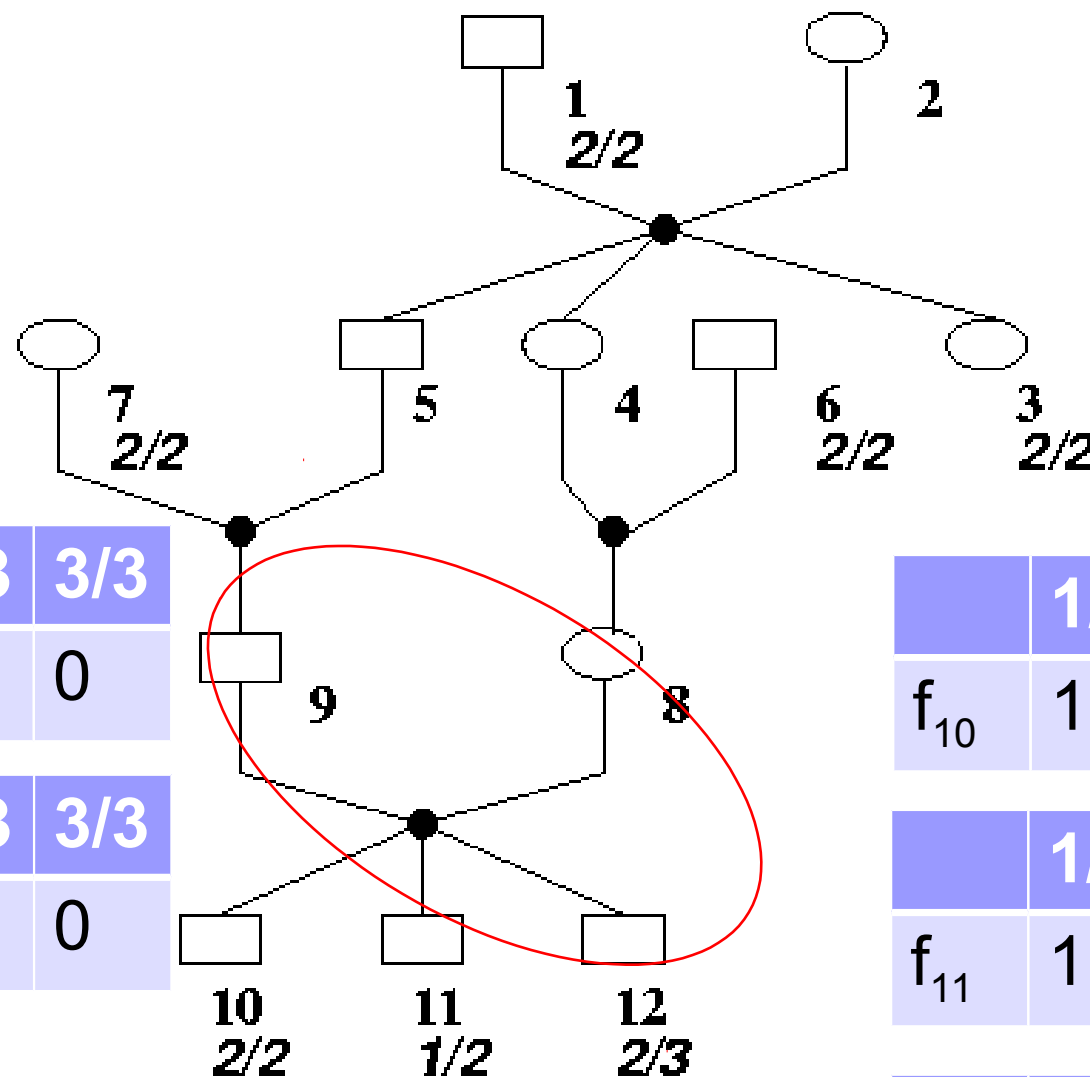
	1/1	1/2	1/3	2/2	2/3	3/3
f_{10}	1	1	1	0	1	1

	1/1	1/2	1/3	2/2	2/3	3/3
f_{11}	1	0	1	1	1	1

	1/1	1/2	1/3	2/2	2/3	3/3
f_{12}	1	1	1	1	0	1

(Schiex, CP 2000), (Cooper et al, AIJ 2010), (Allouche et al, AIJ 2016),...

Generalized **Soft** Arc Consistency



	1/1	1/2	1/3	2/2	2/3	3/3
f_8	1	0	0	0	0	0

	1/1	1/2	1/3	2/2	2/3	3/3
f_9	0	0	0	0	0	0

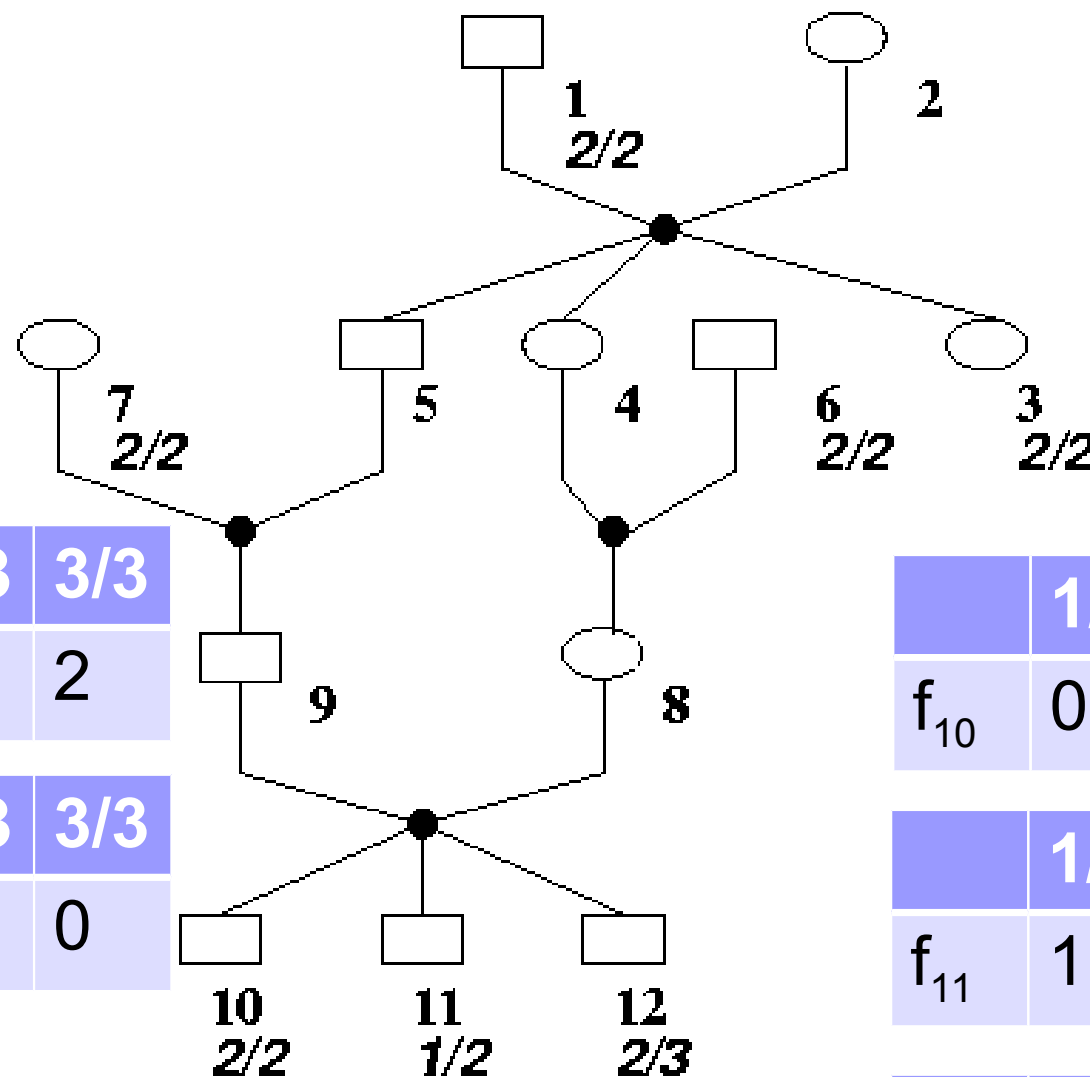
	1/1	1/2	1/3	2/2	2/3	3/3
f_{10}	1	1	1	0	1	1

	1/1	1/2	1/3	2/2	2/3	3/3
f_{11}	1	0	1	1	1	1

	1/1	1/2	1/3	2/2	2/3	3/3
f_{12}	0	0	0	1	0	1

(Schiex, CP 2000), (Cooper et al, AIJ 2010), (Allouche et al, AIJ 2016),...

Generalized **Soft** Arc Consistency



	1/1	1/2	1/3	2/2	2/3	3/3
f_8	2	0	1	0	0	2

	1/1	1/2	1/3	2/2	2/3	3/3
f_9	0	0	0	0	0	0

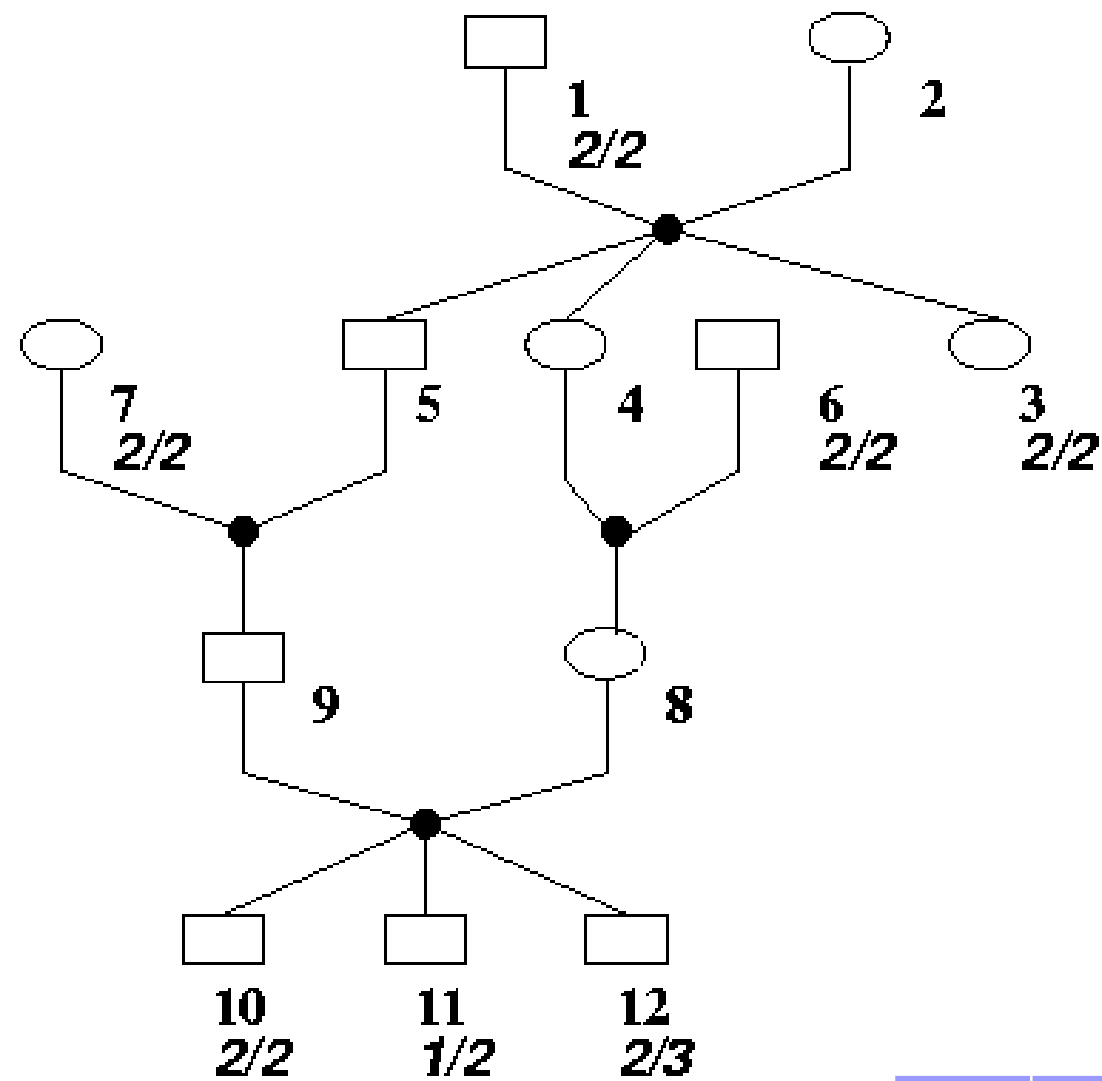
	1/1	1/2	1/3	2/2	2/3	3/3
f_{10}	0	0	0	0	0	0

	1/1	1/2	1/3	2/2	2/3	3/3
f_{11}	1	0	0	1	0	0

	1/1	1/2	1/3	2/2	2/3	3/3
f_{12}	0	0	0	1	0	1

(Schiex, CP 2000), (Cooper et al, AIJ 2010), (Allouche et al, AIJ 2016),...

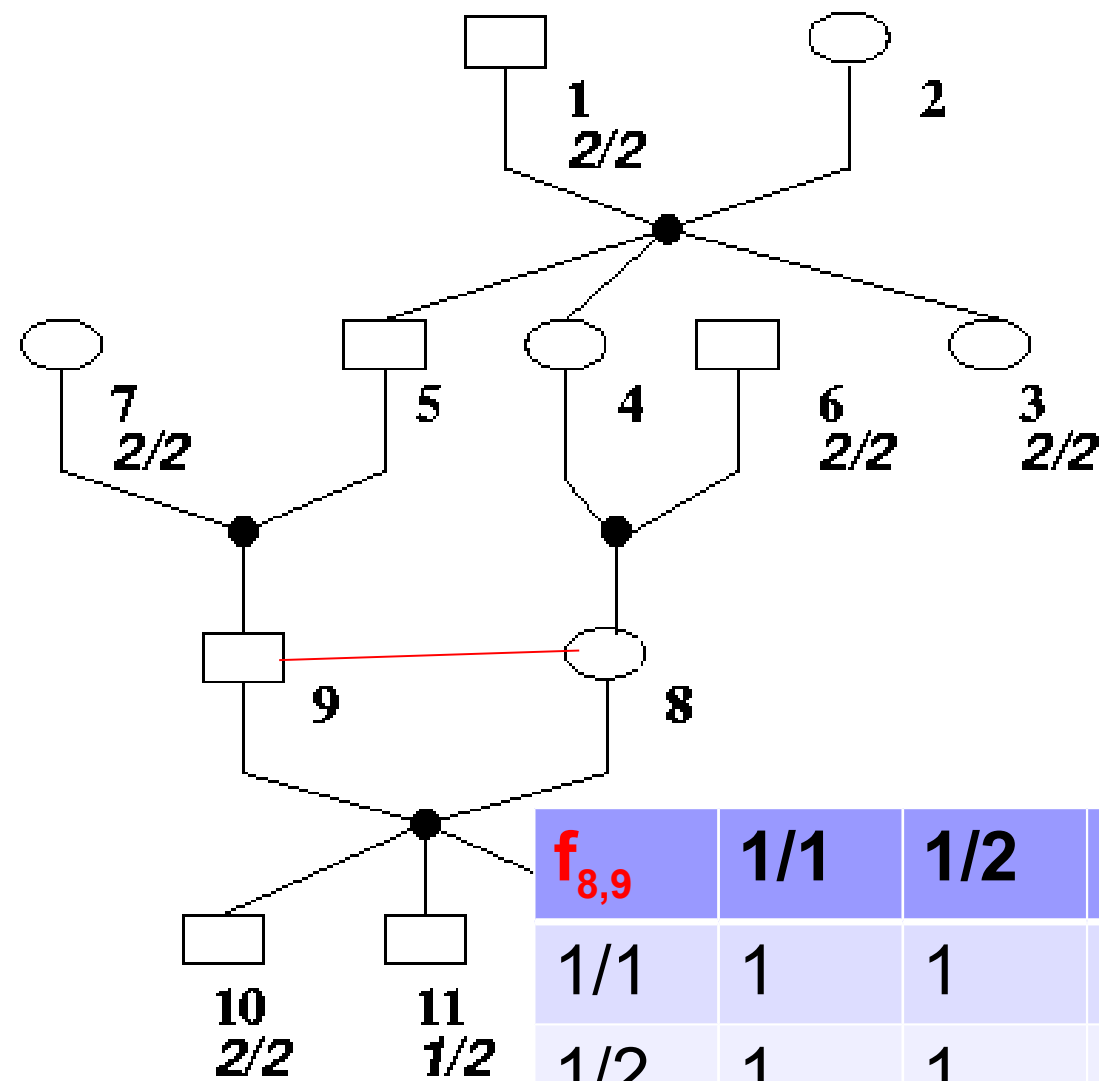
Variable elimination



	1/1	1/2	1/3	2/2	2/3	3/3
f_{12}	1	1	1	1	0	1

(Dechter, AIJ 1999)

Variable elimination



$f_{8,9}$	1/1	1/2	1/3	2/2	2/3	3/3
1/1	1	1	1	1	1	1
1/2	1	1	0	1	0	0
1/3	1	0	1	0	0	1
2/2	1	1	0	1	0	0
2/3	1	0	0	0	0	0
3/3	1	0	1	0	0	1

(Dechter, AIJ 1999)

(Givry et al, Constraints 2008)

Real data

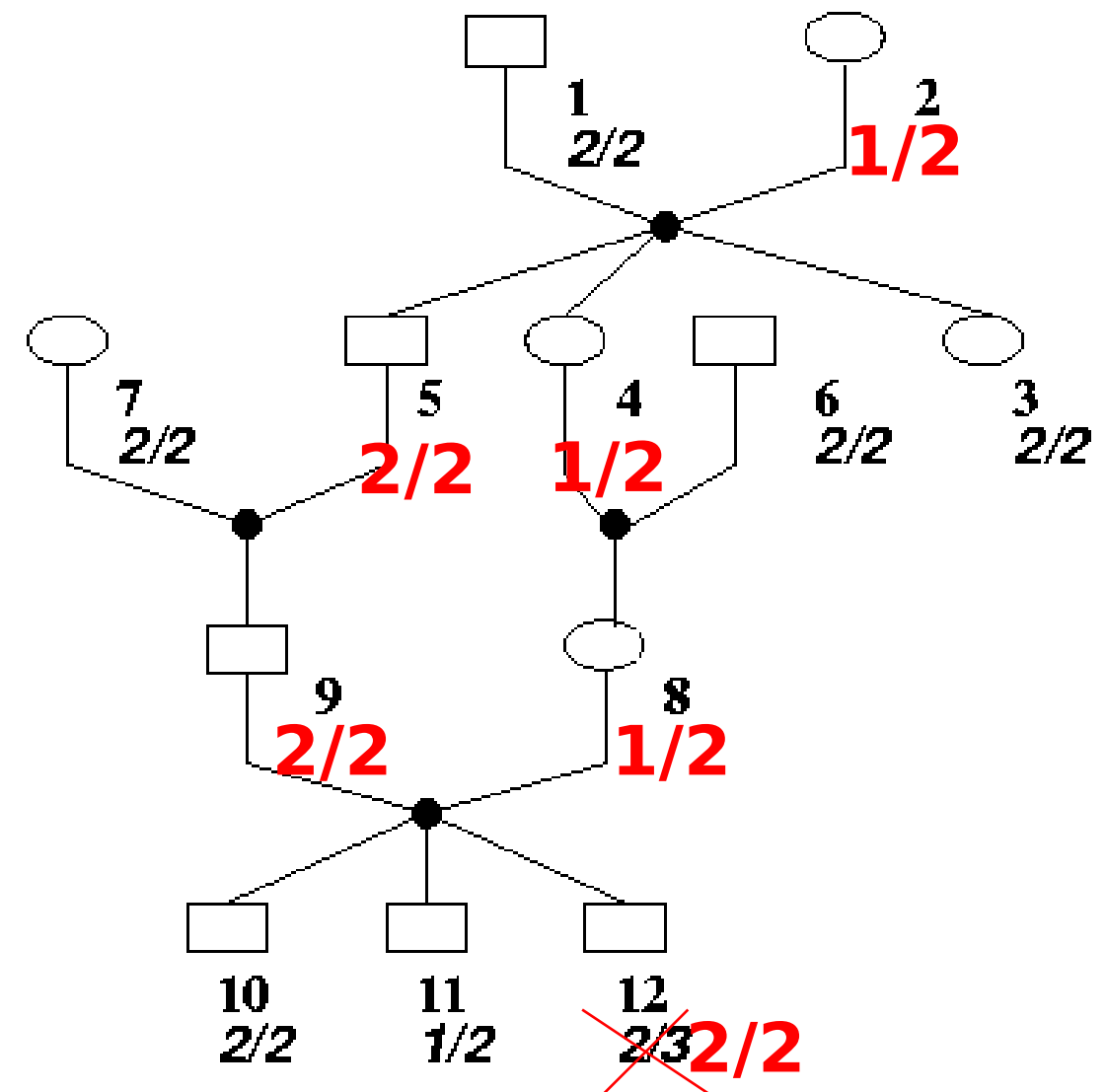
CPU time in seconds to find and prove optimality
on a linux PC 3 GHz with 16 GB using toulbar2 v0.5

	ind	vars	genotyped	alleles	nf	ngen	treewidth ub	B&B-VE(2)		
								errors	time	nodes
<i>eye</i>	36	36	28	6	11	4	2	1	0.02	0
<i>cancer</i>	49	48	37	8	18	5	2	1	0.21	0
<i>parkinson</i>	37	34	13	4	7	7	5	0	0	6
<i>berrichon</i> _{1nc}	129516	9947	2448	4	8821	17	262	2	4.73	8805
<i>berrichon</i> ₁	129516	10017	2483	4	8786	17	330	23	5.81	8384
<i>berrichon</i> _{2nc}	27255	19337	10215	4	4719	19	-	41	5.89	6170
<i>berrichon</i> ₂	27255	19562	10215	4	2381	19	-	106	17.23	15445
<i>langlade</i> ₁	1355	1209	711	9	298	13	84	38	12.28	391
<i>langlade</i> ₂	1355	1223	715	7	298	13	82	89	60.56	17857
<i>langlade</i> ₃	1355	1258	787	5	298	13	85	39	14.19	6731
<i>langlade</i> ₄	1355	1186	672	8	298	13	83	43	59.7	3520
<i>moissac</i> ₁	283	260	183	2	81	5	6	0	0	5
<i>moissac</i> ₂	283	244	167	7	81	5	6	0	0.51	6
<i>moissac</i> ₃	283	225	151	3	81	5	6	0	0	4
<i>moissac</i> ₄	283	256	179	2	81	5	6	0	0	5
<i>moissac</i> ₅	283	237	161	8	81	5	6	0	1.02	5
<i>moissac</i> ₆	283	201	131	11	81	5	5	0	5.64	6

Task 3: Error Correction using Probabilistic Model

- Finds a complete assignment with maximum posterior probability

✉ Bayesian network

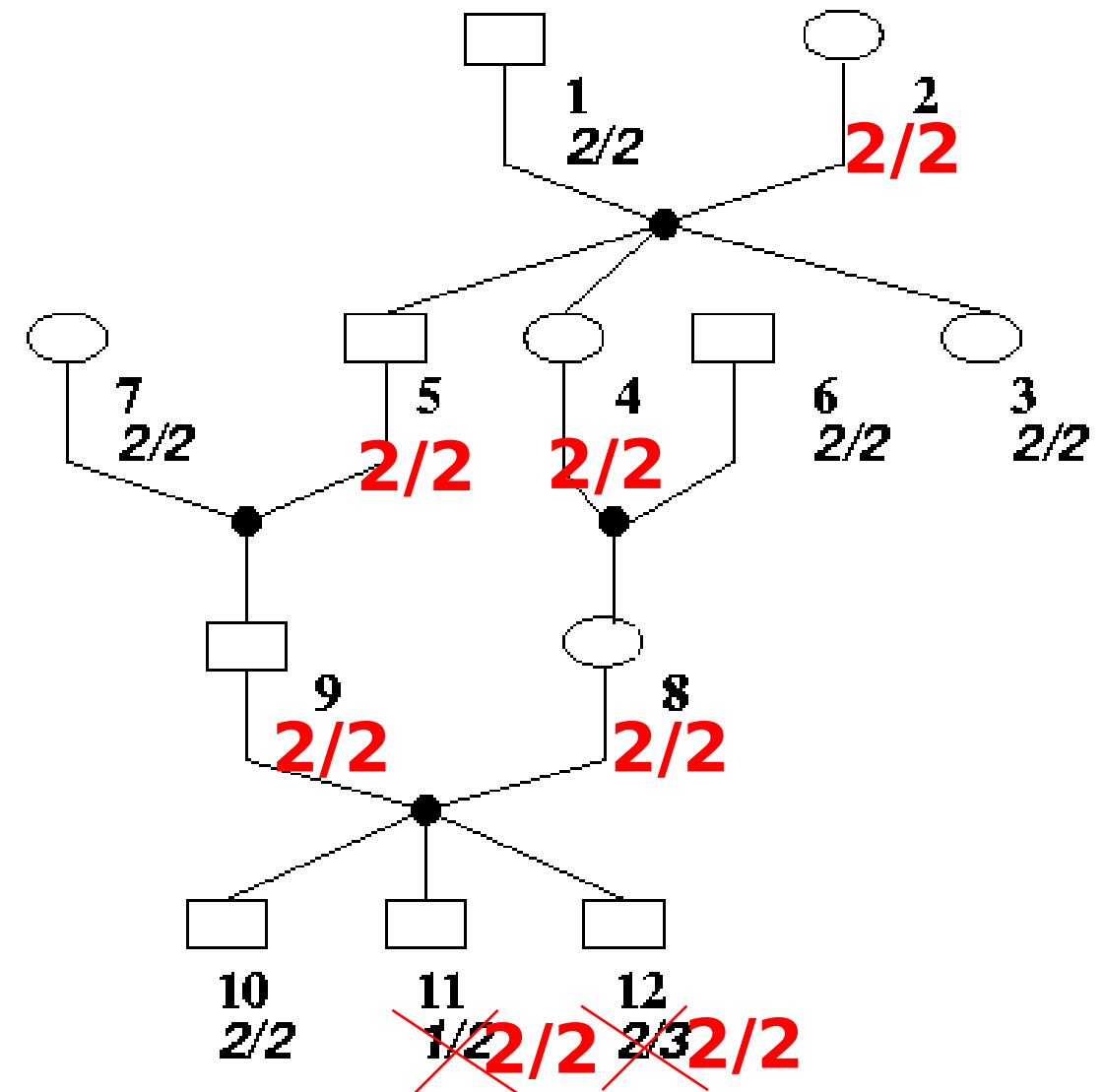


Prior on genotyping error: 1%
(and equifrequent alleles)

Task 3: Error Correction using Probabilistic Model

- Finds a complete assignment with maximum posterior probability

✉ Bayesian network

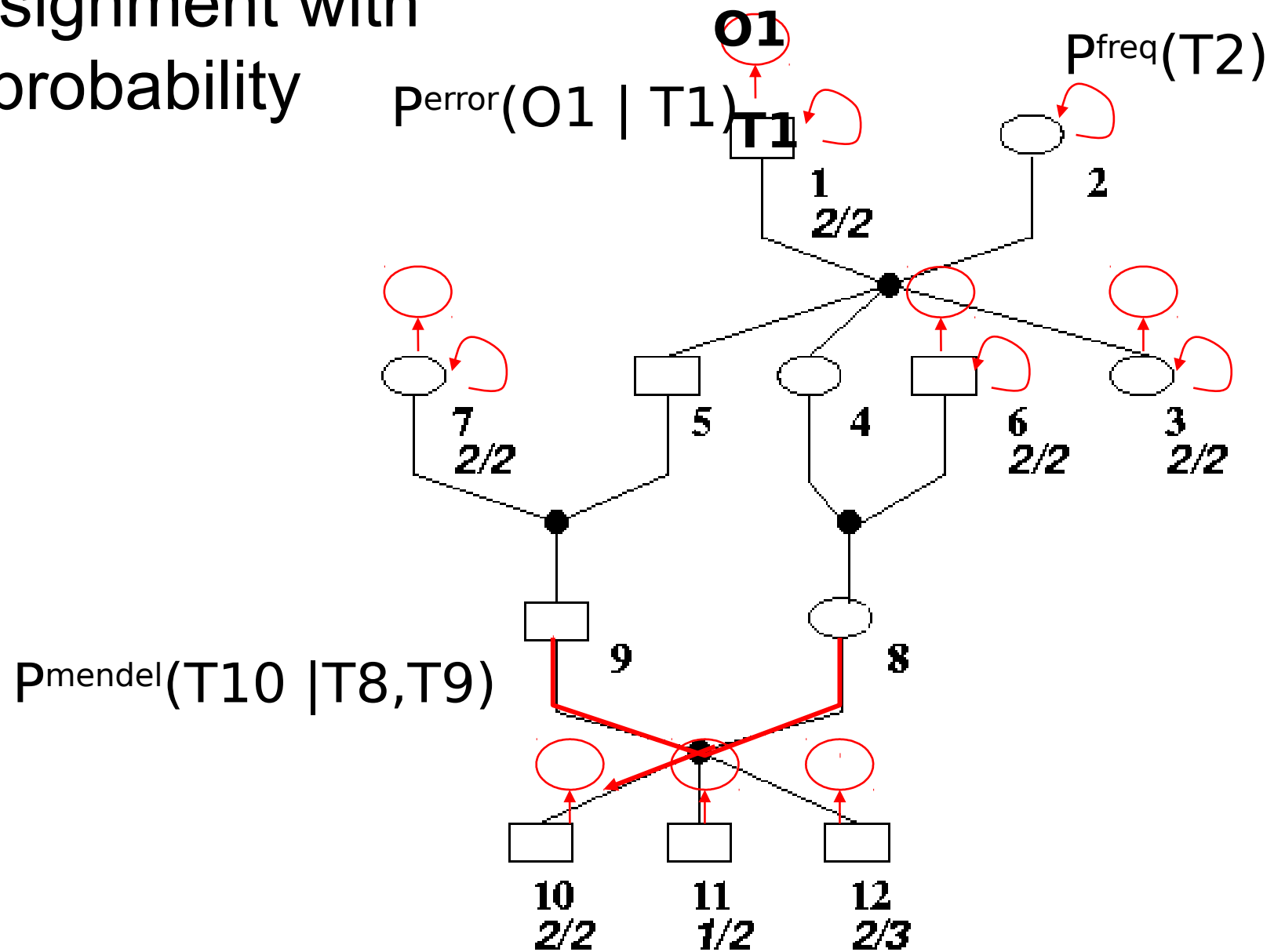


Prior on genotyping error: 10%
(and equifrequent alleles)

Task 3: Error Correction using Probabilistic Model

- Finds a complete assignment with maximum posterior probability p

Bayesian network



$$P(\mathbf{O}, \mathbf{T}) = \prod P^{\text{error}}(\mathbf{O}_i \mid \mathbf{T}_i) \times \prod P^{\text{mendel}}(\mathbf{T}_i \mid \text{parents}(i)) \times \prod P^{\text{freq}}(\mathbf{T}_i)$$

Graphical model

- X , a set of n variables
- D , finite domains of maximum size d
- $F = \{f_{S_1}, \dots, f_{S_e}\}$, a set of e functions with $S_i \subseteq X$

► Probabilistic models
(Markov Net, Bayes Net,...)

► Deterministic models
(Max-SAT,
Weighted CSP,...)

$$P(X) \propto \prod_{i=1}^e f_i(S_i)$$

$$T = 0$$

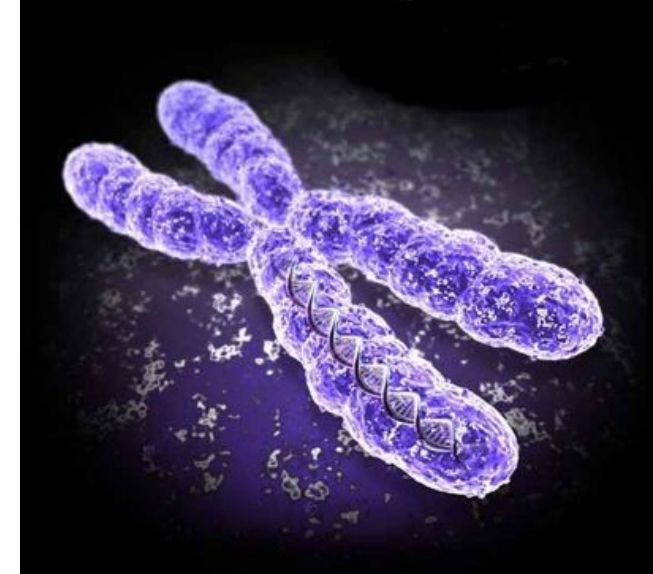
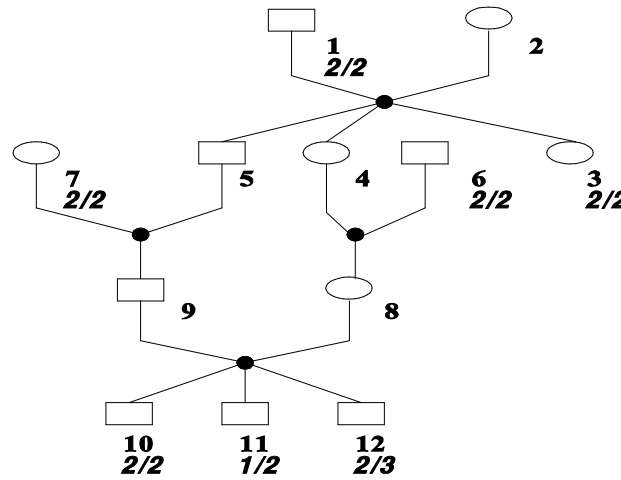

-log transform

$$score(X) = \sum_{i=1}^e f_i(S_i)$$

$$T = k$$

(finite or infinite positive integer)

Genetics

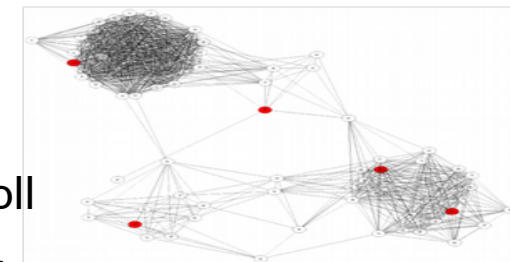


- Restoring consistency on large animal pedigrees

- M. Sánchez, S. de Givry, and T. Schiex. Mendelian error detection in complex pedigrees using weighted constraint satisfaction techniques. Constraints, 13(1):130-154, 2008
- MendelSoft <http://www7.inra.fr/mia/T/MendelSoft/>

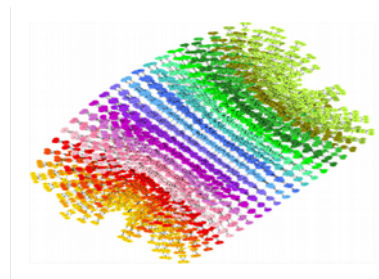
- TagSNP selection

- D Allouche, S de Givry, M Sanchez, and T Schiex. TagSNP selection using Weighted CSP and Russian Doll Search with Tree Decomposition. In CP-09 workshop on Constraint Based Methods for Bioinformatics, Lisbon, Portugal, 2009



- Optimal haplotype reconstruction in half-sib families

- A. Favier, J-M. Elsen, S. de Givry, and A. Legarra. Optimal haplotype reconstruction in half-sib families. In ICLP-10 workshop on Constraint Based Methods for Bioinformatics, Edinburgh, UK, 2010



- Reference population in a genomic selection design

- J-M. Elsen, S. de Givry, G. Katsirelos, F. Shumbusho. Optimizing the reference population in a genomic selection design. In WCB'13. Uppsala. Sweden. 2013





Laboratoire d'Ingénierie
des Systèmes Biologiques
et des Procédés

Catalysis & Enzyme Molecular Engineering Group

COMPUTATIONAL PROTEIN DESIGN

David Allouche, Jessica Davies, Simon de Givry, George Katsirelos, Thomas Schiex

UBIA, UR-875, INRA, F-31320 Castanet Tolosan, France

Isabelle André, Sophie Barbe, Seydou Traoré

LISBP, INSA, UMR INRA 792/CNRS 5504, F-31400 Toulouse, France

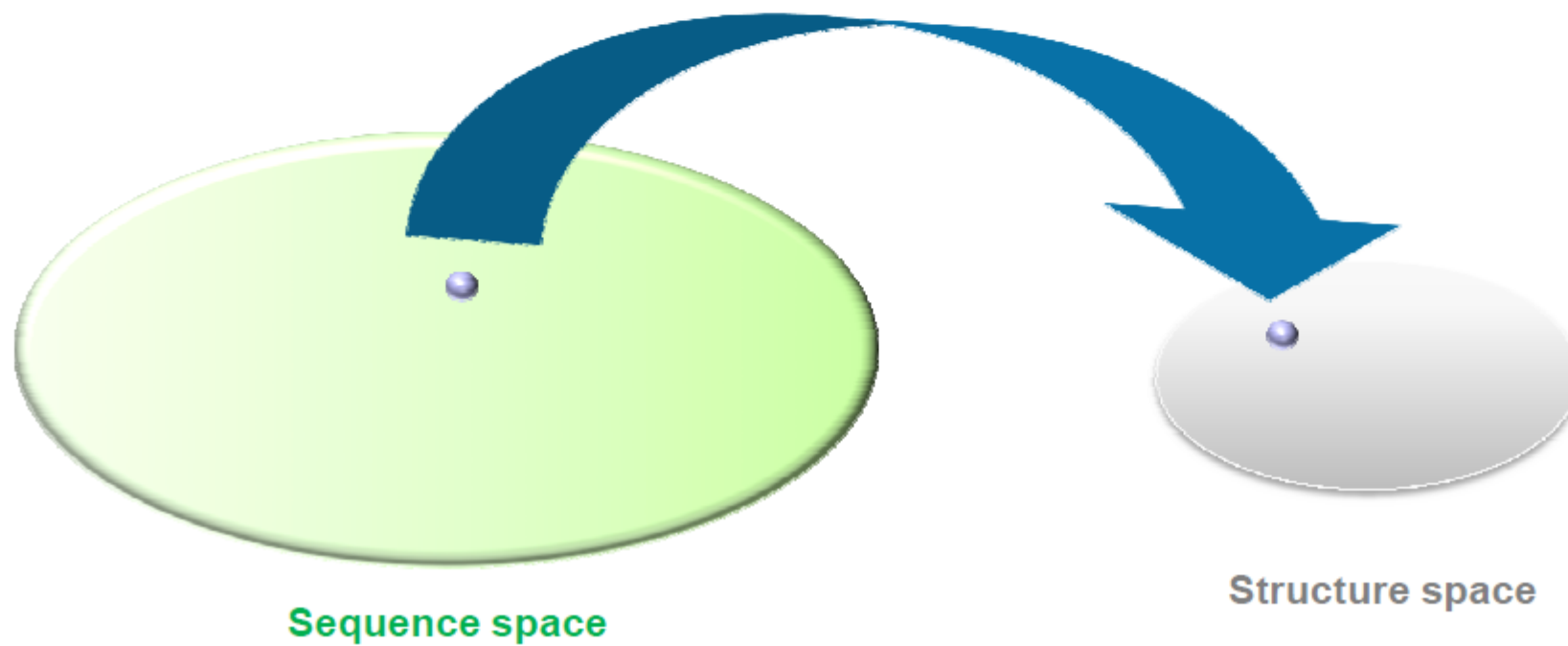
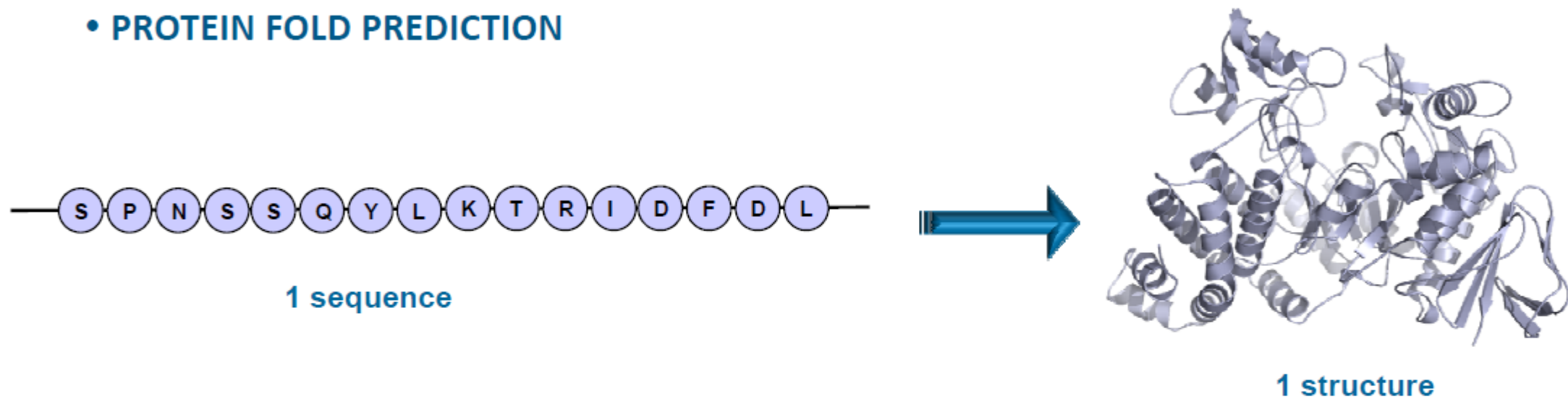
Steve Prestwich, Barry O'Sullivan

Cork Constraint Computation Centre, University College Cork, Ireland

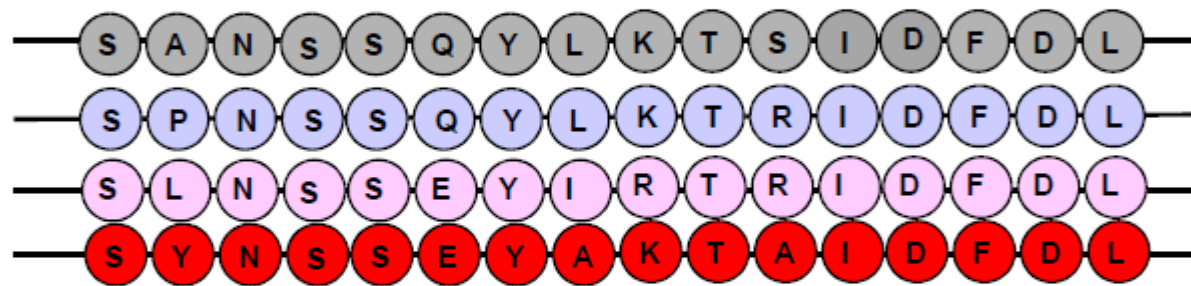
Meeting 30 Nov 2011



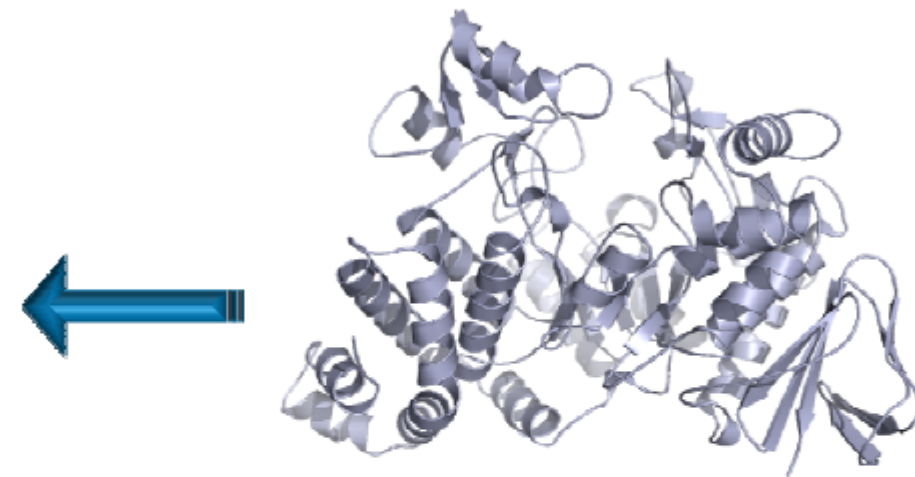
- PROTEIN FOLD PREDICTION



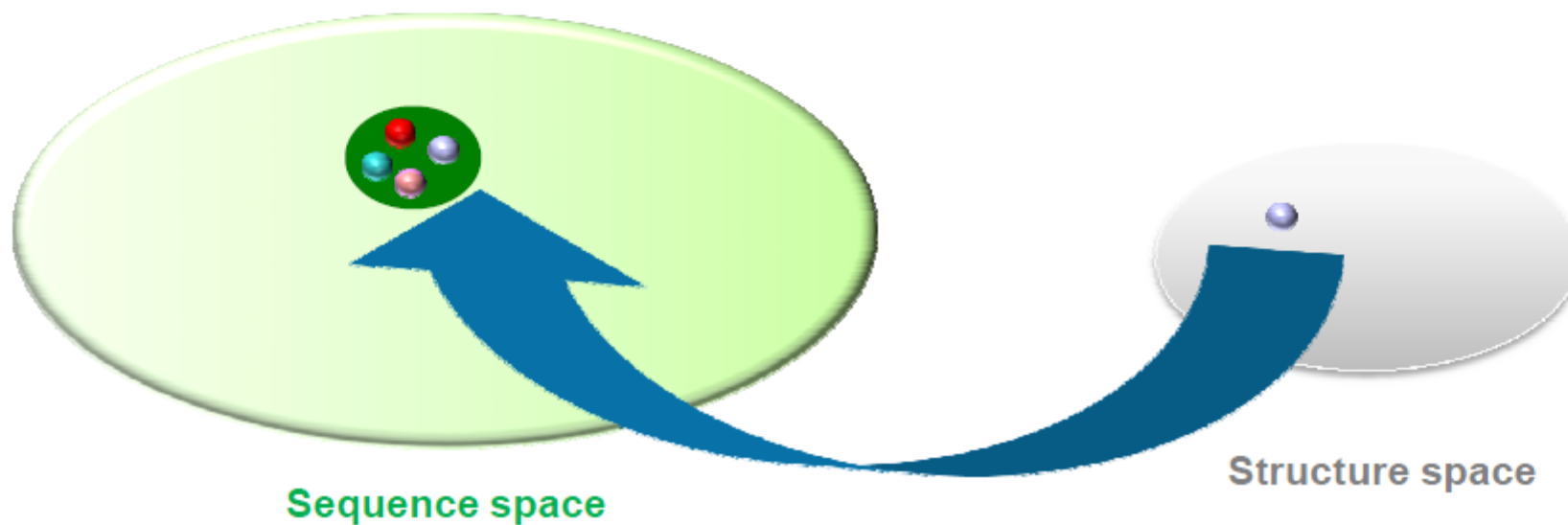
• PROTEIN (RE)DESIGN



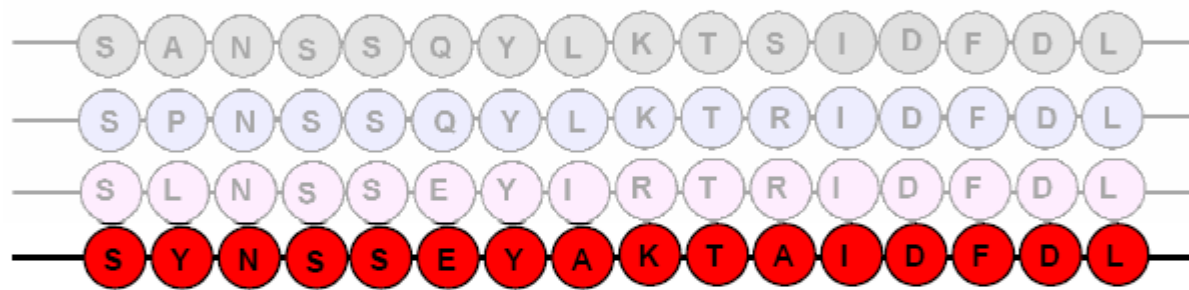
multiple sequences



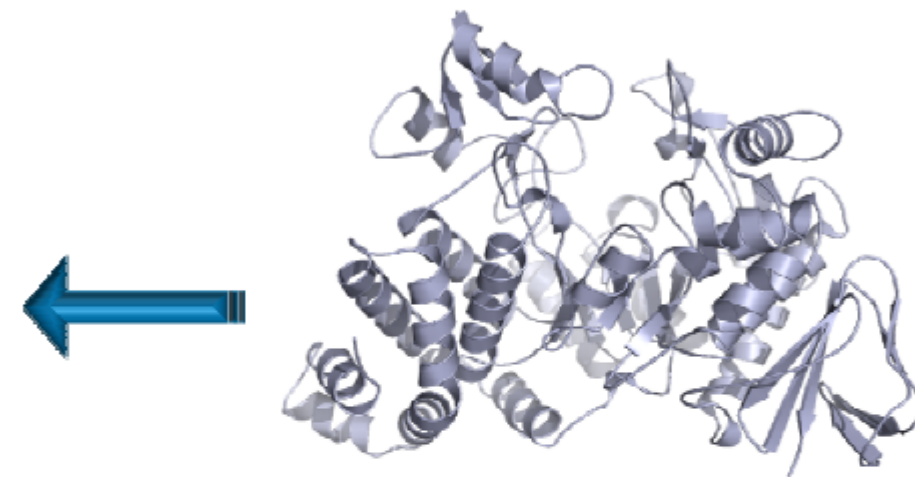
1 structure



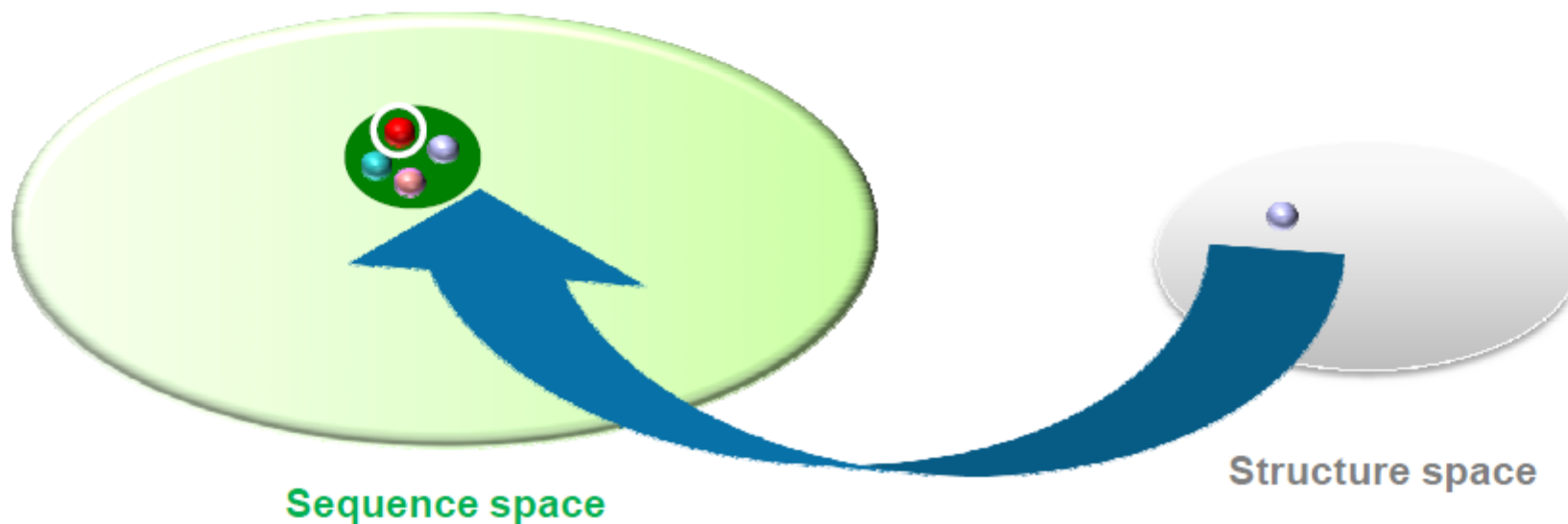
• PROTEIN (RE)DESIGN



« optimal » sequence

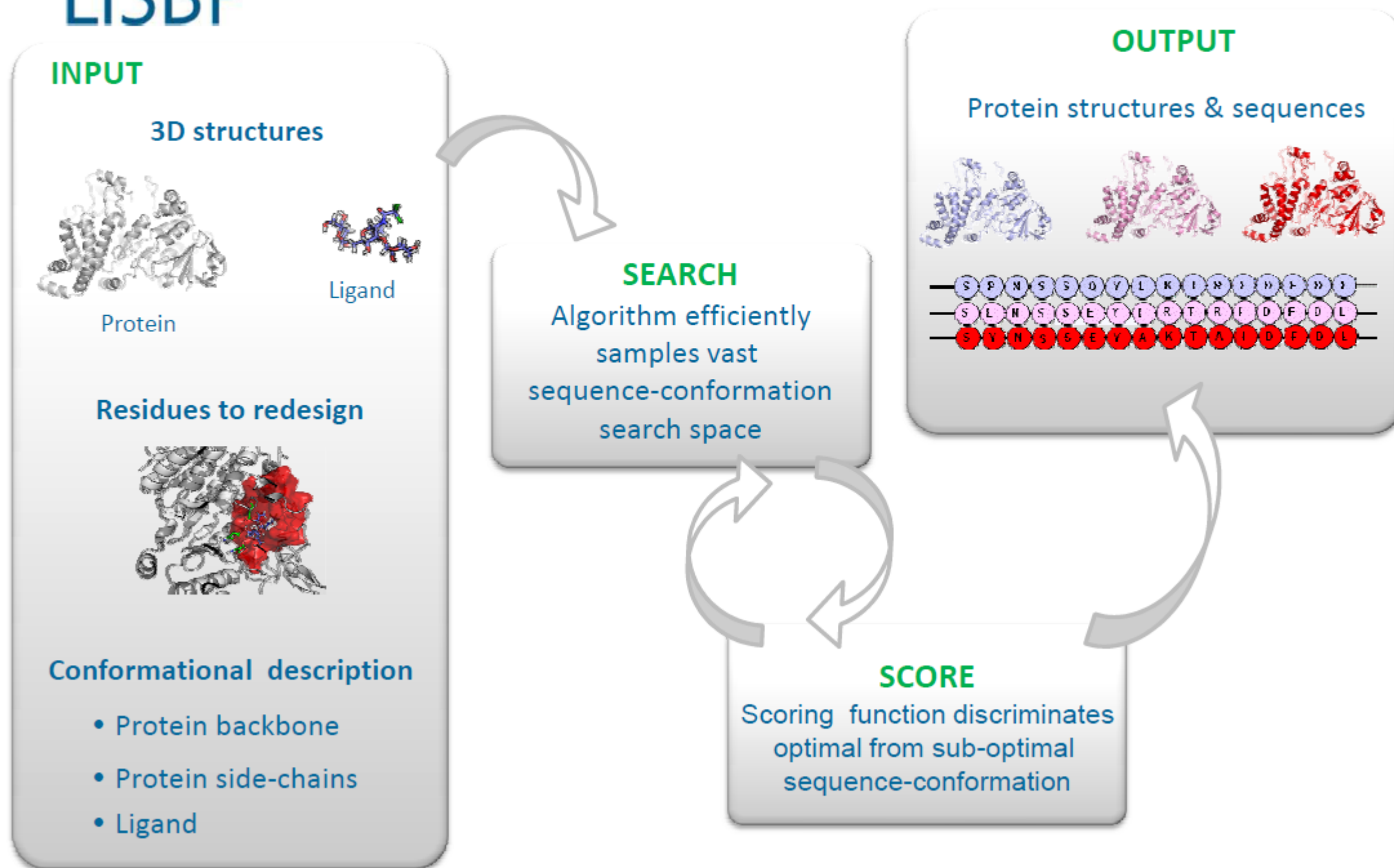


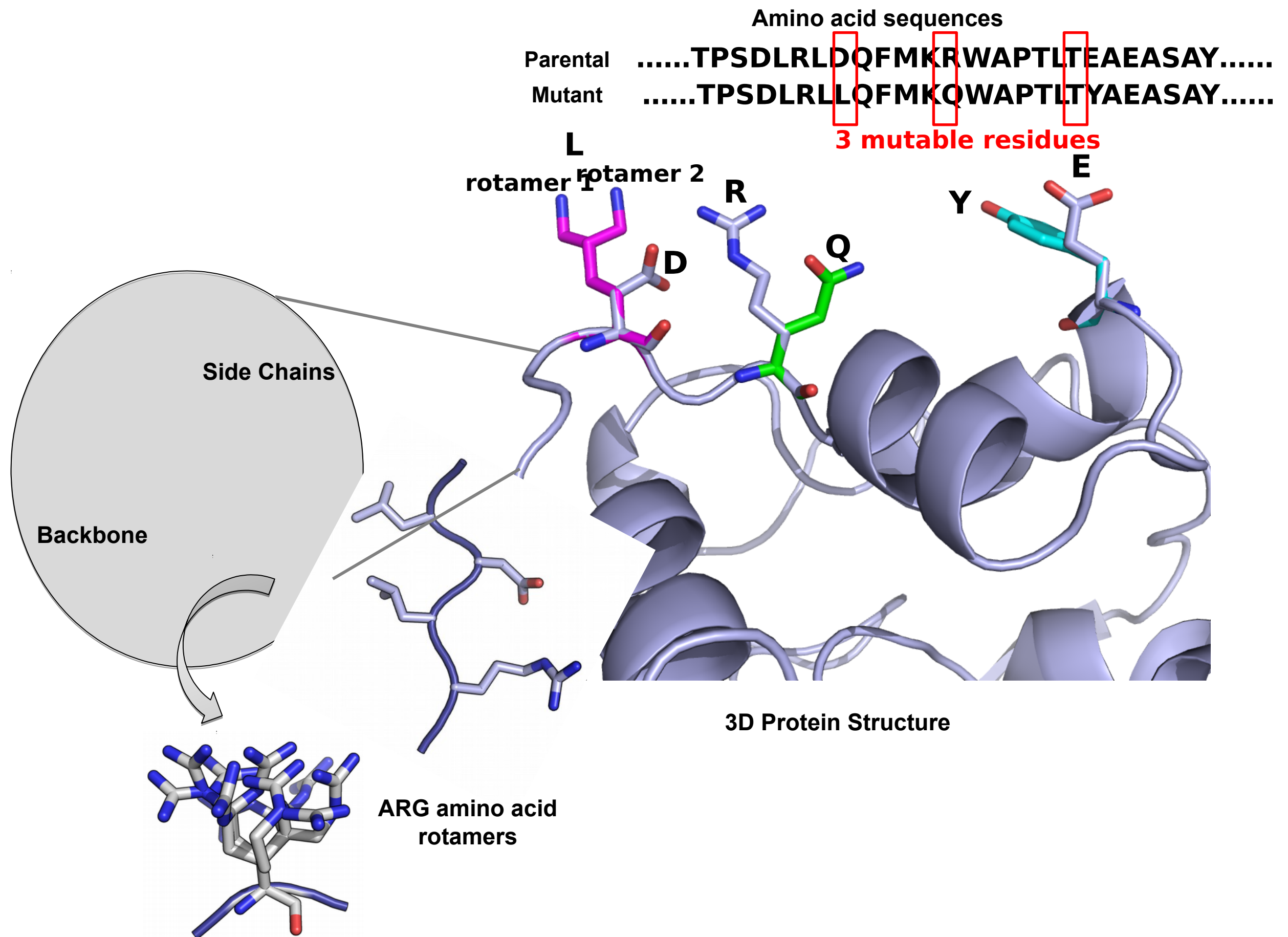
1 structure



GENERAL COMPONENTS OF A COMPUTATIONAL PROTEIN DESIGN (CPD) TOOL

LISBP





Representation of a sequence-conformation model. A) Partial view of protein 3D structure showing as example three mutable positions. B) Zoom on a polypeptide segment. C) Illustration of accessible rotamers for an amino acid type.

Weighted Constraint Satisfaction Problem (X,D,F)

- ▶ **X**: one variable per mutable residue
- ▶ **D**: domain of every variable is defined as the set of all combinations of amino acids (20) and spatial conformations
- ▶ **F**: unary and binary soft constraints to encode the energy function for every pair of residues

$$E = E_{\emptyset} + \sum_i E(i_r) + \sum_i \sum_{j>i} E(i_r, j_s)$$

E_{t'} : Self-energy of the rigid region

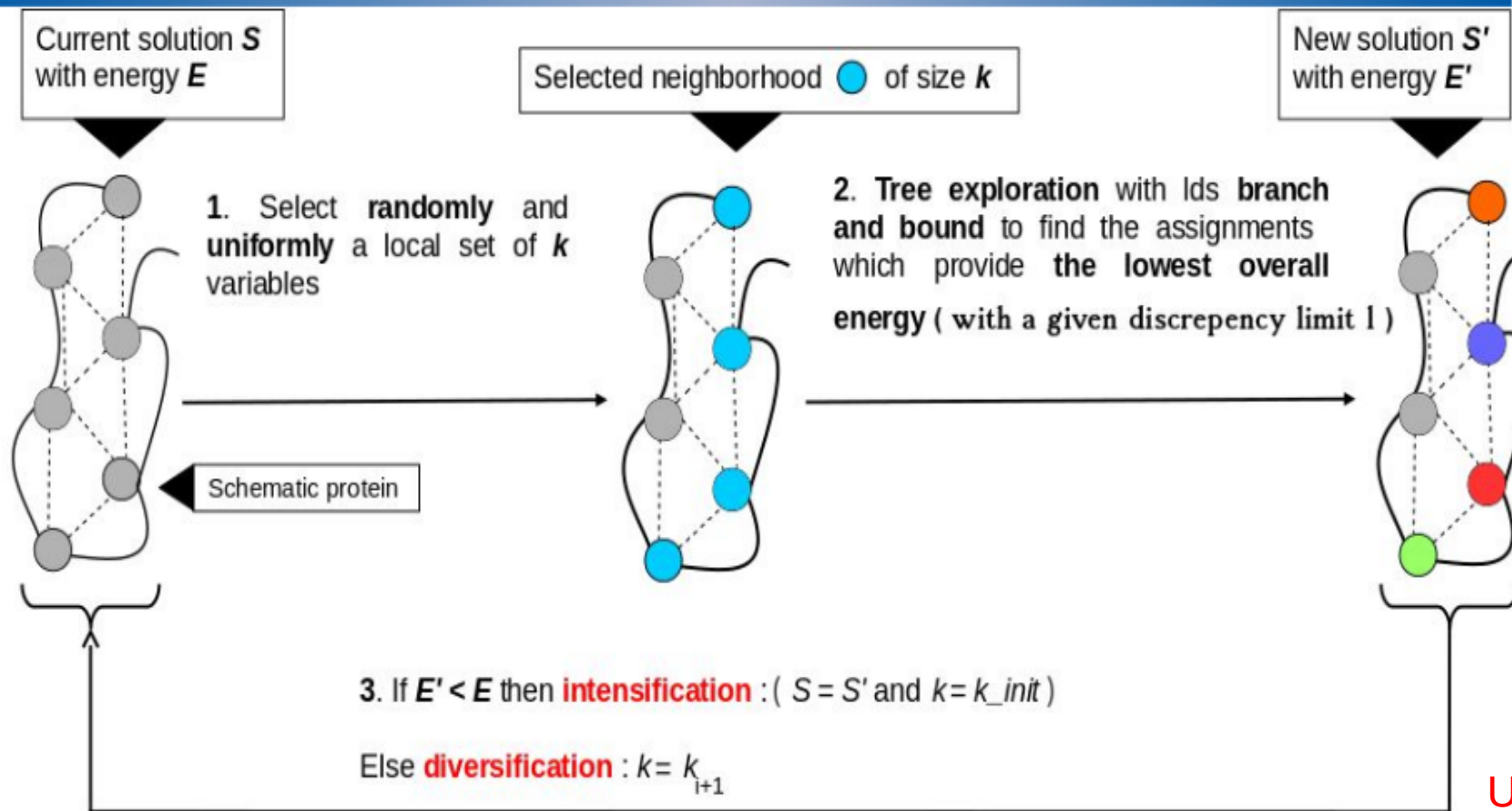
*E(ir) Interaction energy between rotamers
and rigid region*

E(ir, js) Interaction energy between rotamers

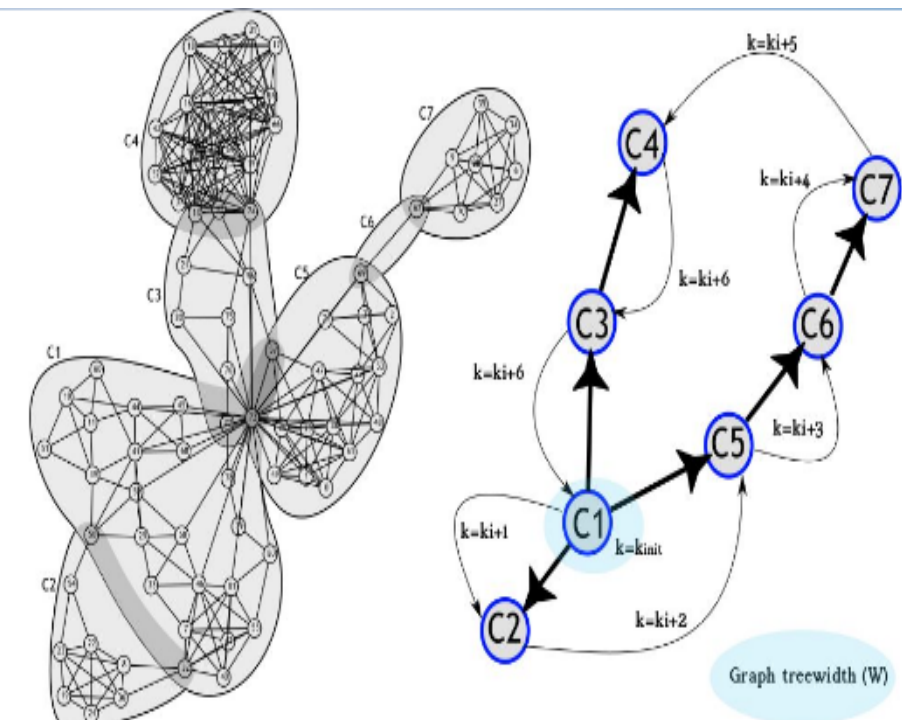
Table 1: For each instance: protein (PDB id.), number of mutable residues, maximum domain and CPU-time for solving using maxhs, daoapt, osprey, cplex, mplp, and toulbar2. A '-' indicates the corresponding solver did not prove optimality within the 9,000-second time-out. A '!' indicates solver stops with a SEGV signal.

PDB id.	n	d	maxhs	daoapt	osprey	cplex	mplp	toulbar2
2TRX	11	44	4,086	268.6	31.5	2.6	2.8	0.1
1PGB	11	45	5,209	300.4	135.3	3.6	0.5	0.1
1HZ5	12	45	5,695	350.2	75.0	7.6	16.7	0.1
1UBI	13	45	-	826.9	2,812.6	139.2	37.3	0.2
1PGB	11	148	-	-	8,695.2	-	1,291	4.3
1HZ5	12	148	-	-	2,398.3	1,555	1,217	2.4
1UBI	13	148	-	-	-	-	-	1,557
2PCY	18	44	-	-	1,281.1	26.9	14.5	0.2
2DHC	14	148	-	-	-	-	5,388	14.1
1CM1	17	148	-	-	138.4	473.1	87.5	3.3
1MJC	28	182	3,698	631.7	4.6	4.1	0.8	0.1
1CSP	30	182	-	-	200.0	1,380	1,264	0.8
1BK2	24	182	-	-	93.2	125.0	114.9	0.6
1SHG	28	182	-	-	138.0	39.4	!	0.2
1CSK	30	49	-	-	41.7	12.5	9.6	0.1
1SHF	30	56	-	-	44.3	8.6	3.1	0.1
1FYN	23	186	-	-	622.0	2,548	3,136	2.8
1PIN	28	194	-	-	-	-	-	3.7
1NXB	34	56	-	-	11.1	17.0	4.5	0.2
1TEN	39	66	-	-	113.0	45.4	17.1	0.2
1POH	46	182	-	-	77.9	29.0	13.1	0.3
2DRI	37	186	-	-	-	-	4,458	42.8
1FNA	38	48	-	-	3,310	124.9	121.2	0.5
1UBI	40	182	-	-	-	2,572	979.4	2.4
1C9O	43	182	-	-	2,310	1,635	155.7	1.8
1CTF	39	56	-	-	-	263.2	549.2	0.7
2PCY	46	56	-	-	2,080	54.0	20.3	0.4
1DKT	46	190	-	-	5,420	1,254	3,103	2.5
2TRX	61	186	-	-	487.0	765.0	344.1	0.9
1CM1	42	186	-	-	-	-	-	17.4
1BRS	44	194	-	-	-	-	-	346.5
1CDL	40	186	-	-	-	-	-	341.8
1LZ1	59	57	-	-	-	601.6	1,084	1.5
1GVP	52	182	-	-	-	-	-	361.8
1RIS	56	182	-	-	-	-	8,483	288.4
2RN2	69	66	-	-	-	480.8	565.2	1.2
1CSE	97	183	-	-	367.0	172.9	60.9	0.7
1HNG	85	182	-	-	5,590	2,360	5,934	2.8
3CHY	74	66	-	-	-	-	8,691	59.6
1L63	83	182	-	-	-	1,480	1,779	2.9
			3	5	25	29	33	40

Variable Neighborhood Search

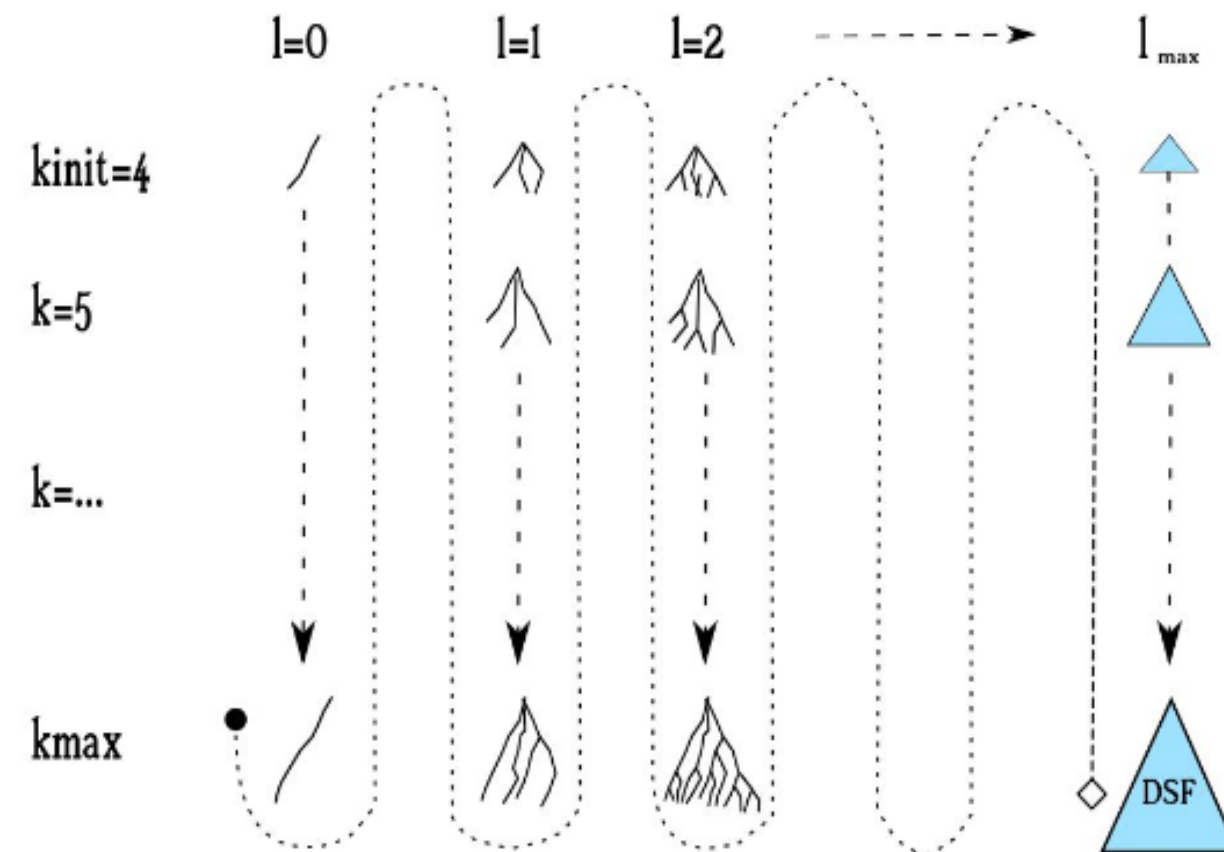
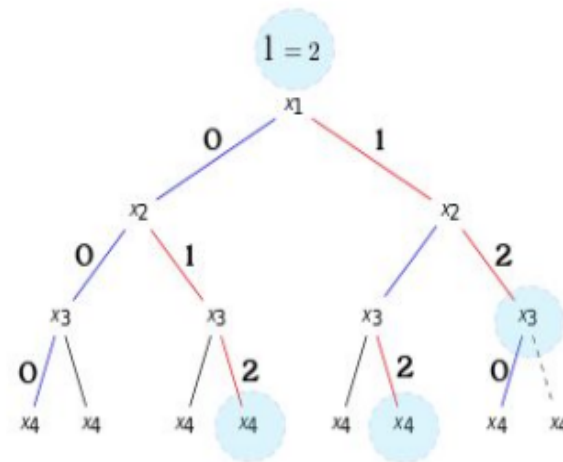


Neighborhoods based on tree decomposition



UDGVNS = iterative (VNS + LDS + TREEDEC)

Limited discrepancy search



Comparative evaluation on CPD instances. In bold instances solved by toulbar2. Number of successful runs and average CPU times over 10 runs to reach the best known solution within a 1 hour time limit . ncl is the number of clusters. Parallel UPDGVNS with npr=96 processes. ΔE is in Rosetta Energy unit. It has been obtained by the difference of solutions costs divided by the cost shift used during modeling, in this case 10^8 . For toulbar2, reported CPU times to find the best known solution within 24 hours.

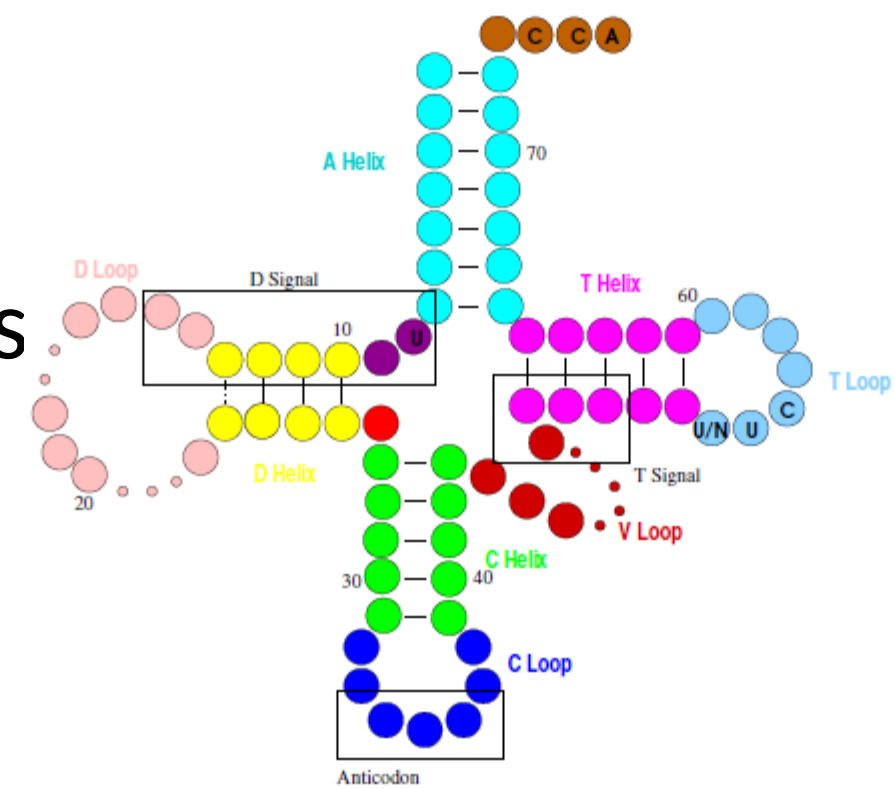
Instance	ncl	(1)		(2)		(3)		(4)	(5)		Speed-up		
		Succ.	Time (s)	Succ.	Time (s)	Succ.	Time (s)	ΔE	Time (s)	ΔE	(1/2)	(1/3)	(2/3)
5dbl	87	10/10	2761± 67	10/10	963± 53	10/10	149± 18	0.27	783	0	2.86	18.53	6.46
5jdd	168	0/10	TO	10/10	3248 ± 162	10/10	646± 72	4.08	20662	0.04	-	-	5.02
3r8q	157	0/10	TO	10/10	3478± 82	10/10	397± 47	4.19	12762	0	-	-	8.76
4bxp	108	10/10	1,213± 32	10/10	1352± 40	10/10	216± 24	0.26	2966	0	0.89	5.61	6.26
1f00	177	0/10	TO	10/10	2575± 29	10/10	542± 30	4.38	9749	0	-	-	4.75
2x8x	131	10/10	3312± 111	4/10	2801± 1,071	10/10	521± 58	4.16	69213	3.61	1.18	6.35	5.37
1xaw	66	0/10	TO	10/10	581± 38	10/10	260± 57	2.73	2804	0	-	-	2.23
5e10	74	10/10	1667± 86	10/10	1412± 21	10/10	132± 24	0.26	1171	0	1.18	12.62	10.7
1dvo	82	10/10	940± 22	10/10	940± 22	10/10	197± 25	2.90	34142	0.18	1	4.77	4.77
1ytq	67	10/10	2235± 211	10/10	1304± 40	10/10	280± 26	1.67	17063	0.31	1.71	7.98	4.65
2af5	140	0/10	TO	10/10	2659± 75	10/10	894± 247	4.37	86029	0.60	-	-	2.97
1ng2	86	10/10	1065± 69	10/10	547± 18	10/10	260± 26	1.14	38730	5.93	1.94	4.09	2.10
3sz7	79	0/10	TO	10/10	2952± 276	10/10	320± 50	3.11	82625	0.54	-	-	9.22
2gee	110	10/10	1,647± 8	10/10	1276± 19	10/10	286± 38	1.68	5021	0	1.29	5.75	4.46
5e0z	73	10/10	621± 15	10/10	995± 18	10/10	105± 8	0.16	999	0	0.62	5.91	9.47
1yz7	87	10/10	2148± 9	10/10	945± 24	10/10	457± 85	2.91	83816	3.20	2.27	4.70	2.06
3lf9	72	10/10	1636± 31	10/10	894± 41.17	10/10	216± 21	2.41	2667	0	1.82	7.57	4.13
3e3v	91	0/10	TO	10/10	2327± 621	10/10	263± 38	2.57	81574	0.15	-	-	8.84
lisl	107	10/10	3178± 51	10/10	2274± 421	10/10	329± 48	3.45	63832	0.42	1.39	9.65	6.91
5eqz	89	10/10	1849± 49	10/10	697± 6	10/10	225± 16	2.20	12768	0	2.65	8.21	3.097
4uos	118	10/10	2304± 988	10/10	2109± 62	10/10	469± 79	5.20	58589	17.86	1.09	4.91	4.5
TO: TimeOut (1): VNS/LDS+CP ($k++$, $\ell = 3$) (2): UDGVNS ($k++$, $\ell = 3$) (3): UPDGVNS ($npr, 1, k++$, $\ell = 3$) (4): FIXBB (5): TOULBAR2													

Abdelkader Ouali, David Allouche, Simon de Givry, Samir Loudni, Yahia Lebbah, Francisco Eckhardt, and Lakhdar Loukil
Iterative Decomposition Guided Variable Neighborhood Search for Graphical Model Energy Minimization
In Proc. of UAI-17, pages 550-559, Sydney, Australia, 2017

Bioinformatics

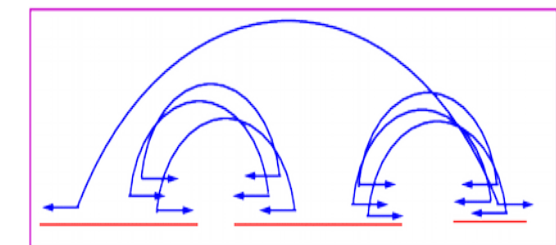
- RNA motif search in genomic sequences

- P. Thébault, S. de Givry, T. Schiex, C. Gaspin. Combining constraint network processing and pattern matching to describe and locate structured motifs in genomic sequences. *Bioinformatics*, 22(17), 2006
- M. Zytnicki, C. Gaspin, T. Schiex. DARN! A weighted constraint solver for RNA motif localization. *Constraints*, 13(1), 2008



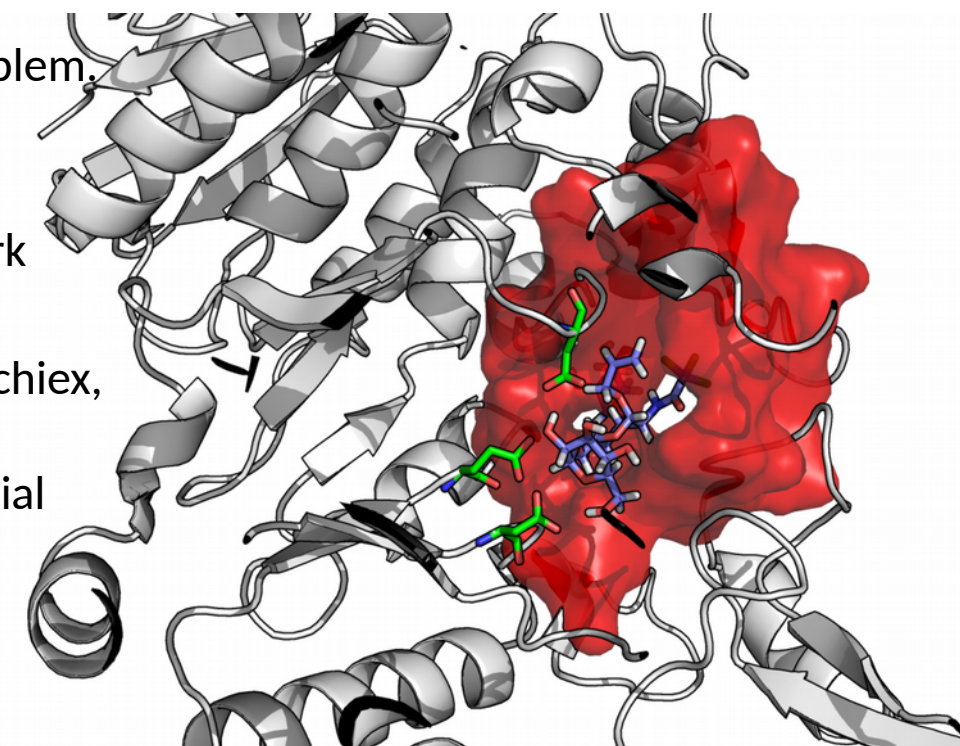
- Genome scaffolding

- N Briot, A Chateau, R Coletta, S de Givry, P Leleux, and T Schiex. An integer linear programming approach for genome scaffolding. In *CP-14 workshop on Constraint-Based Methods for Bioinformatics*, Lyon, France, 2014.

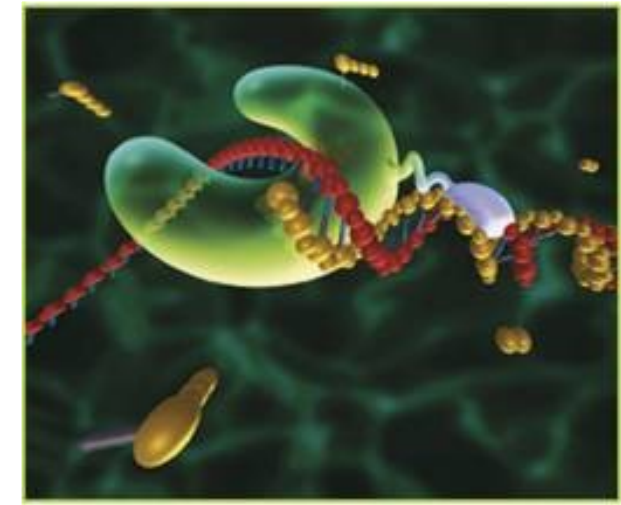


- Computational Protein Design

- D. Allouche, S. Traoré, I. André, S. de Givry, G. Katsirelos, S. Barbe, T. Schiex. Computational Protein Design as a Cost Function Network Optimization Problem. In *CP'2012*, Québec, Canada.
- S. Traoré, D. Allouche, S. de Givry, G. Katsirelos, T. Schiex, S. Barbe. A New Framework for Computational Protein Design through Cost Function Network Optimization. *Bioinformatics*, 2013
- David Allouche, Jessica Davies, Simon de Givry, George Katsirelos, Thomas Schiex, Seydou Traoré, Isabelle André, Sophie Barbe, Steve Prestwich, and Barry O'Sullivan. Computational protein design as an optimization problem. *Artificial Intelligence*, 212:59-79, 2014
- D. Simoncini, D. Allouche, S. de Givry, C. Delmas, S. Barbe, and T. Schiex. Guaranteed discrete energy optimization on large protein design problems. *Journal of Chemical Theory and Computation*, 11(12):5980-5989, 2015



Gene regulatory network

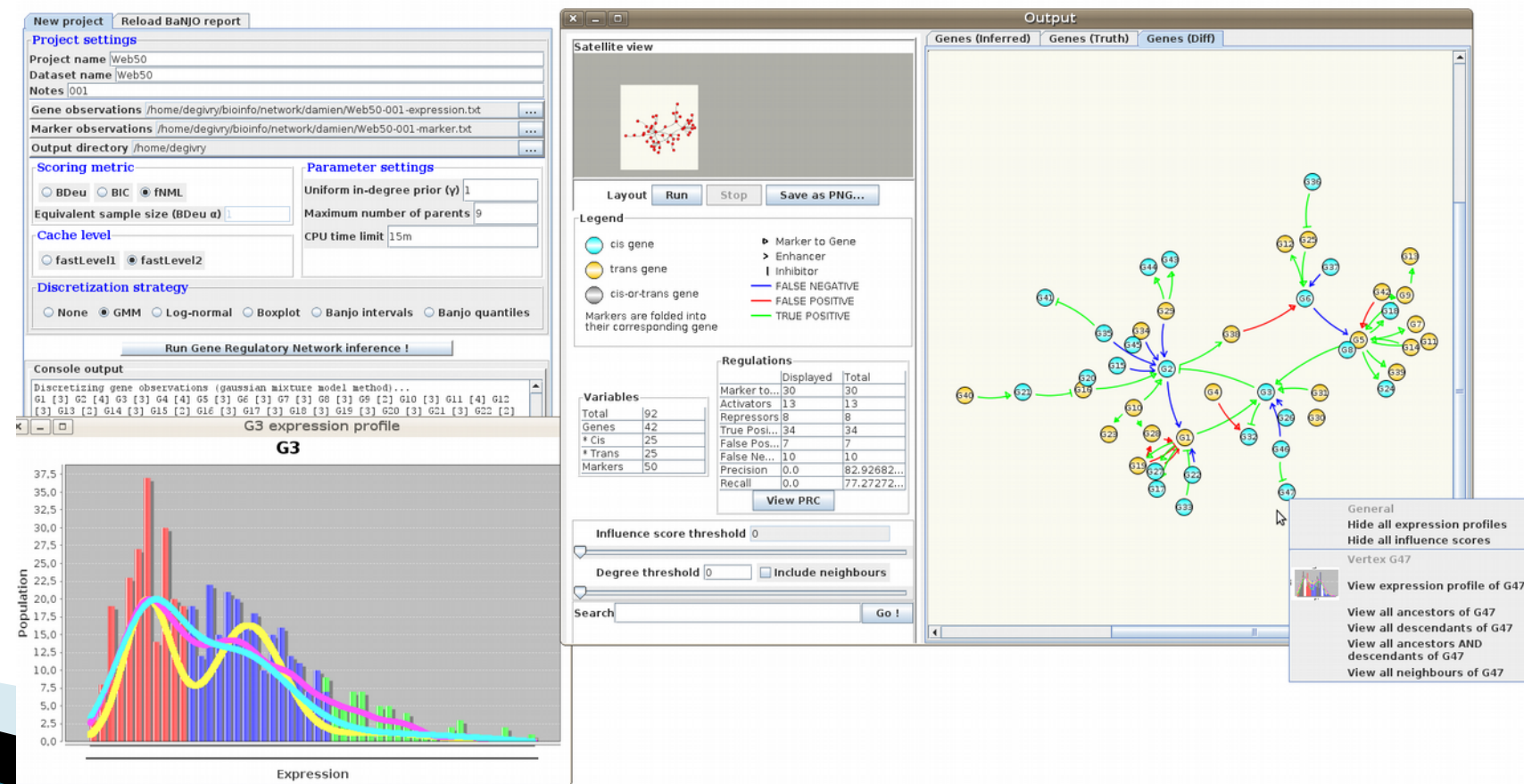


► Structure learning (Bayesian net, GGM, RF)

- M. Vignes, J. Vandael, D. Allouche, N. RamadanAlban, C. CiercoAyrolles, T. Schiex, B. Mangin, S. de Givry. Gene regulatory network reconstruction using bayesian networks, the Dantzig selector, the lasso and their meta-analysis. PLoS ONE, 6(12), 2011.
- J Vandael, B Mangin, and S de Givry. New Local Move Operators for Bayesian Network Structure Learning. In Proc. of PGM-12, Granada, Spain, 2012.
- David Allouche, Christine Cierco-Ayrolles, Simon de Givry, G Guillermin, Brigitte Mangin, Thomas Schiex, Jimmy Vandael, and Matthieu Vignes. Gene Network Inference, chapter A Panel of Learning Methods for the Reconstruction of Gene Regulatory Networks in a Systems Genetics Context. Springer, 2014

○ GeneBayesNet

<http://carlit.toulouse.inra.fr/genebayesnet>



Solvers and benchmarks

- ▶ **Toulbar2** <http://www.inra.fr/mia/T/toulbar2>
(Open source WCSP, MaxSAT, MPE solver in C++) (Debian package toulbar2)
 - Contribution by UPC (J. Larrosa's team) & CSIC (P. Meseguer's team)
 - Contribution on soft global cost functions by CUHK (J. Lee's team)
 - Contribution on parallel large neighborhood local search by Caen University (P. Boizumault's team)
- ▶ **Numberjack** <http://numberjack.ucc.ie>
 - Multi-solver (Mistral, CPLEX, Gurobi, SCIP, Minisat, toulbar2) platform in python
 - minizinc reader
- ▶ **Large set of graphical model benchmarks**
(wcsp, uai, wcnf, lp, mzn formats) (Givry et al, Constraints 2016)

<http://genoweb.toulouse.inra.fr/~degivry/evalgm>