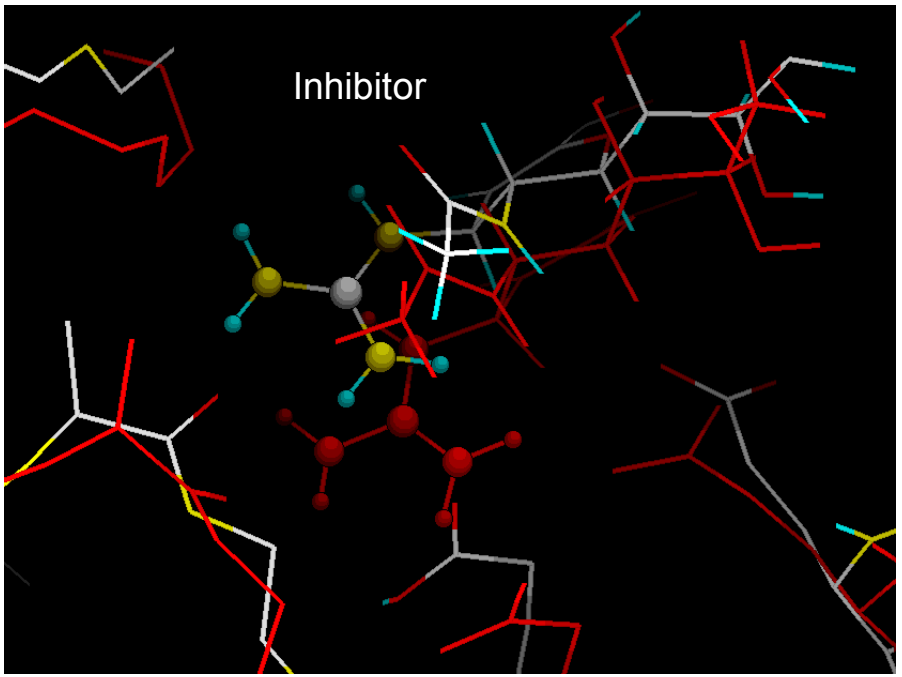
Protein Modeling:
Structure Preparation and Contact Analysis

- A Problem
- The Crystal Structure
- The Geometry
- The Contact Analysis
- No Problem!

November 2001

A Problem: From a MD Simulation

Is the large deviation in the position of the inhibitor relevant?



or rather:
- error in the original structure (crystal structure)?
- error in the computational procedure?

The Crystal Structure: Flow Chart

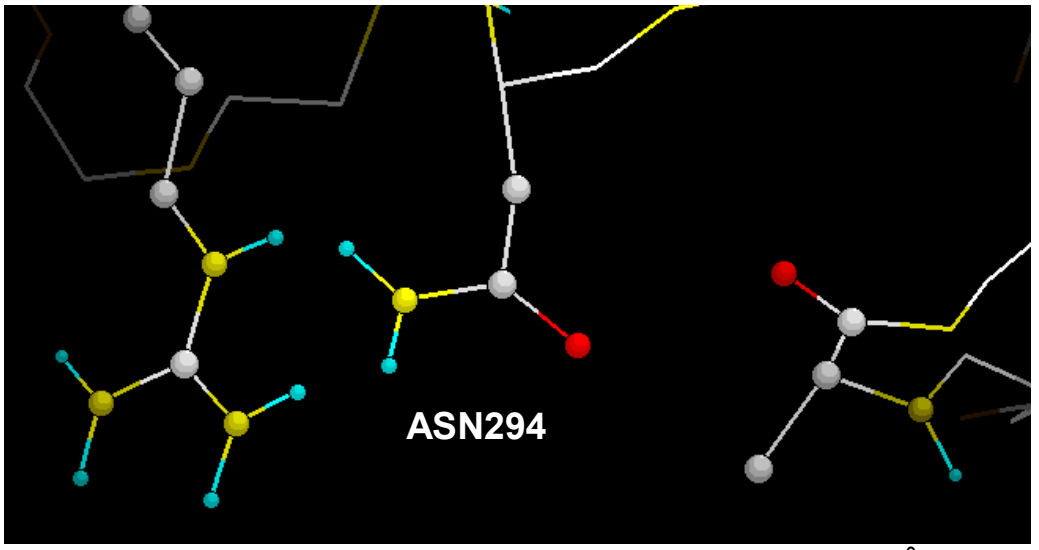
- **Measurement:** X-ray diffraction pattern
- **Derivation:** electron density (calc. vs. exp.)
- **Interpretation:** least squares fit for spatial coordinates and temperature factors, /w the known problems of optimizations

- **Result:** Coordinates /w assignments of atomic numbers for heavy atoms (C, N, O, etc.)

⇒ **Crystal structure is a mixture of objective measurement and subjective interpretation**

The Crystal Structure: Interpretation I

Amide Sidechain: wrong result (N/O-assignment)

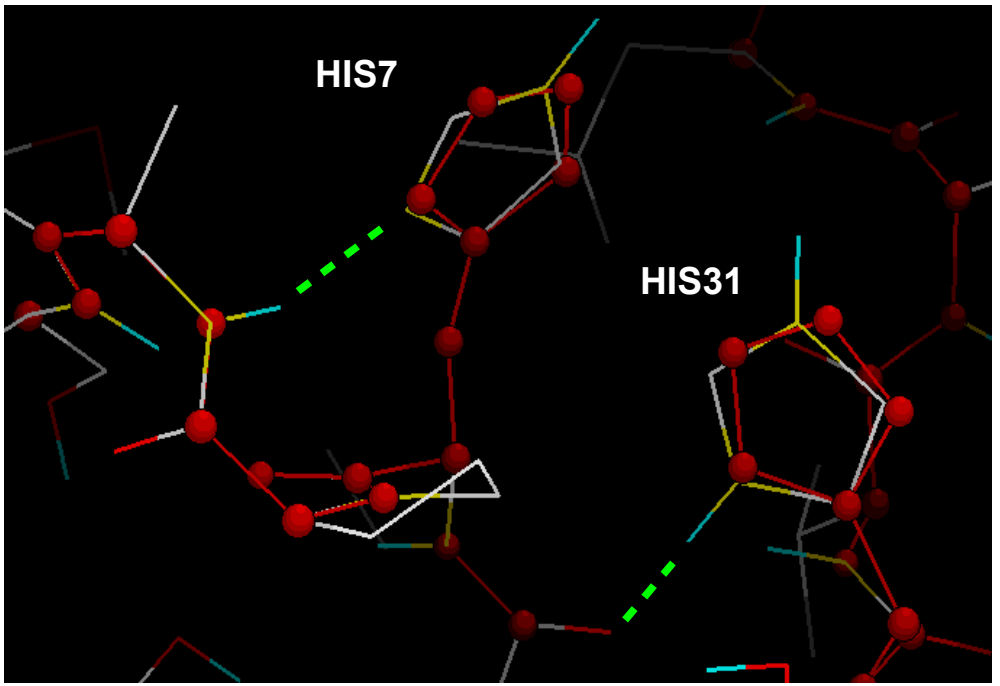


(1NNC, 2.2Å, 1995)

Hydrogen bonding network possible only after N/O exchange

The Crystal Structure: Interpretation II

2 Histidines: wrong result (C/N-assignment in aromatic rings)

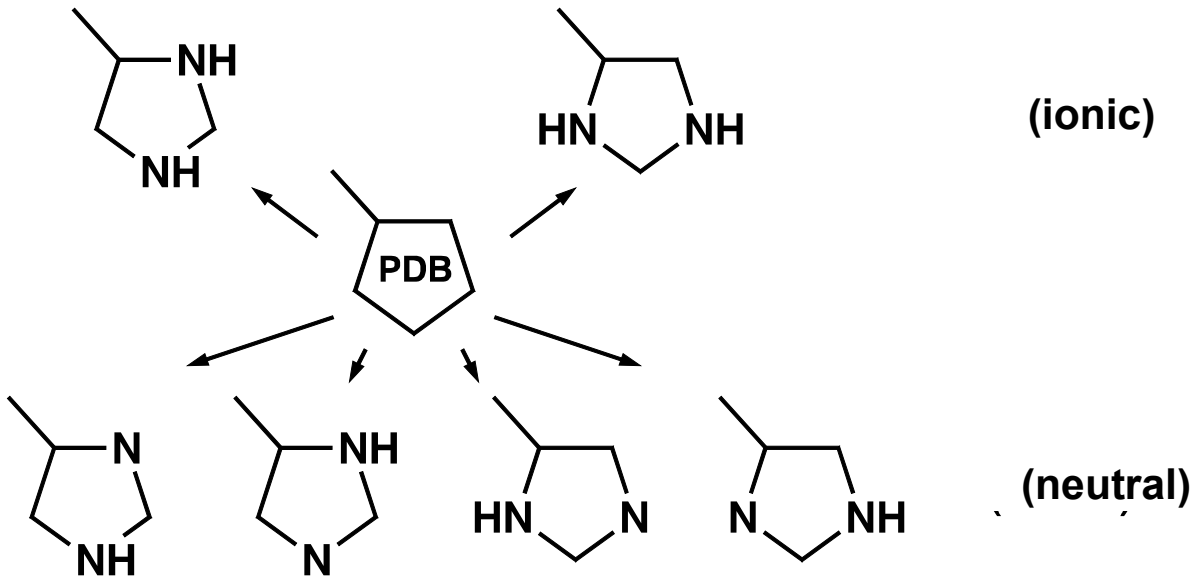


(1BX9, 2.6Å, 1998)

Hydrogen bonds possible only after C/N exchange (rotation of rings!)

The Crystal Structure: Interpretation III

Histidine Sidechain: C/N-assignment and protonation

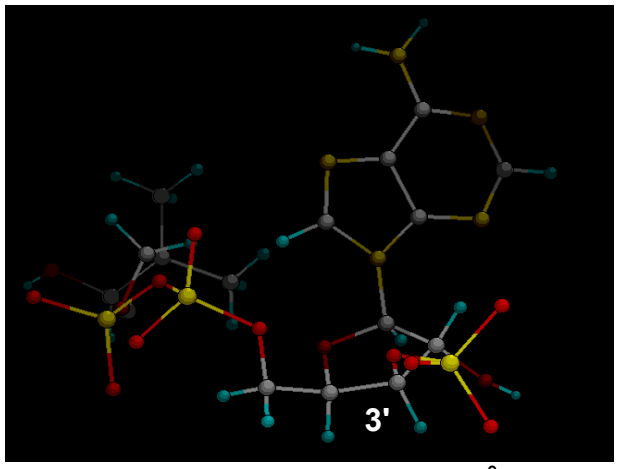


interpretation of physical measurement (orientation of the ring plane) in one of six ways (/w respect to surrounding)!

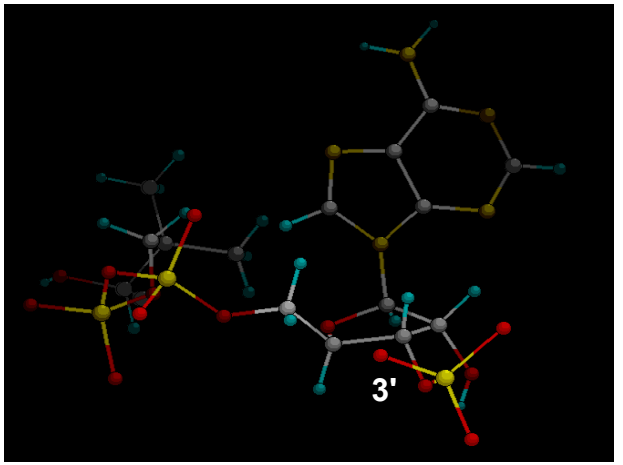
The Crystal Structure: Interpretation IV

Hetero-Groups: geometry parameters have to be specified

Coenzyme A: configuration of C-3'



wrong (1DQ8, 2.1Å, 1999)



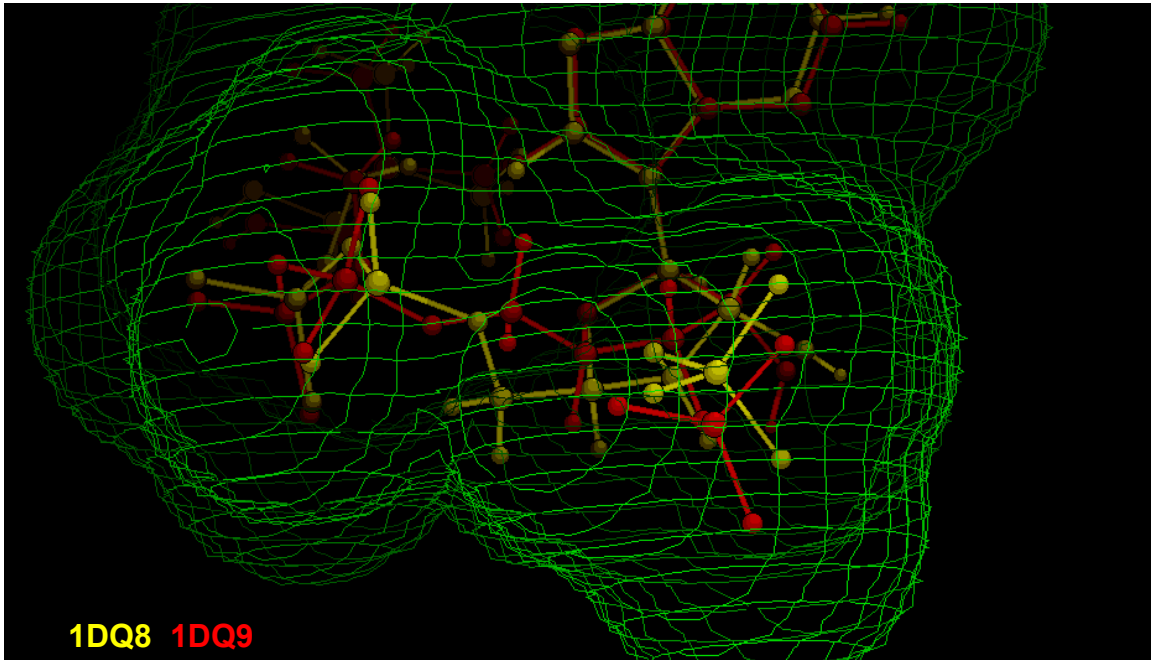
right (1DQ9, 2.8Å, 1999)

both structures from the same article: the same parameters lead to different results? Wrong result at better resolution!

The Crystal Structure: Interpretation V

Hetero-Groups: either structure fits electron density

Coenzyme A: superposition of structures



no decision possible in experiment:
external information must be regarded!

The Crystal Structure: Quality Criteria I

- **Resolution:** phys. limits due to crystal imperfections
  NOT: precision of atomic positions

- **B-Factors:** degree of freedom in interpretation
  NOT: vibrational motion of atomic positions

- **R-Factor:** measure of global regularity
  NOT: quality of local coordinates

Refinement is a fit to experimental and theoretical constraints
rather than
a determination of geometry parameters

The Crystal Structure: Quality Criteria II

Statistical Criteria for Geometry Parameters /w Reference to other Crystal Data (PROCHECK, WHAT_IF, etc.):

- bond lengths and angles: Engh and Huber
- dihedral angles:
  - Ramachandran, ω
  - preferred conformations for χ_1, χ_2
- close van der Waals contacts
- close lying water molecules

$\Rightarrow$ criteria do not ensure that computations on the structure are possible (legitimate)!

The Crystal Structure: Quality Criteria III

Comparison: ENGH/HUBER-Param's to AMBER-Ref. values

(Stat. Standard Deviation) and [Deviation of Ref. Values from Parameters]

	C-N	CA-C	CA-CB	N-CA	C-N-CA	CA-C-N	CB-CA-C	N-CA-C	N-CA-CB
Pro	1.341	-	-	1.466	122.60	116.90	-	111.80	103.00
	(0.016)			(0.015)	(5.00)	(1.50)		(2.50)	(1.10)
Amber	1.335	-	-	1.449	121.90	116.60	-	110.10	109.70
	[-0.006]			[-0.017]	[-0.70]	[-0.30]		[-1.70]	[+6.70]
Gly	-	1.516	-	1.451	120.60	116.40	-	112.50	-
		(0.018)		(0.016)	(1.70)	(2.10)		(2.90)	
Amber	-	1.522	-	1.449	121.90	116.60	-	110.30	-
		[0.006]		[-0.002]	[1.30]	[0.20]		[-2.20]	
Any	-	-	1.530	1.458	121.70	116.20	110.10	111.20	110.50
			(0.020)	(0.019)	(1.80)	(2.00)	(1.90)	(2.80)	(1.70)
Amber	-	-	1.526	1.449	121.90	116.60	111.10	110.10	109.50
			[-0.004]	[-0.009]	[0.20]	[0.40]	[1.00]	[-1.10]	[-1.00]

excellent agreement of experimental and reference values!
but:
crystal structures often deviate from these parameters!

The Geometry: Basic Requirements

- appropriate Parameters for Hetero-Groups

- favourable Energy Contributions
  - for valence terms
  - for van der Waals terms (close contacts)
  - other non-valence terms

  $\Rightarrow$ no energetical 'hot spots'

- optimum Hydrogen Bonding Network

$\Rightarrow$ Requirements *have to be met* Prior to Theoretical Studies

CHEOPS Structure Preparation

- **Protonation of Ionizable Residues:**
  depending on the actual surrounding

- **Positions of Protons and Hetero-Atoms:**
  optimum hydrogen bonding network

- **Water Surrounding:**
  essential water molecules and solvation

- **Valence Optimization:**
  optimization of bond lengths and bond angles

⇒ **MD Simulation s*tarts* with the best Resultant Geometry and needs *no* Special Protocol**

Contact Analysis

Functionality Vectors:
The spatially restricted ability of chemical functionalities like hydrogen bond donor/acceptor or electronic π-systems to interact are represented by vectors together with more spherical van der Waals groups

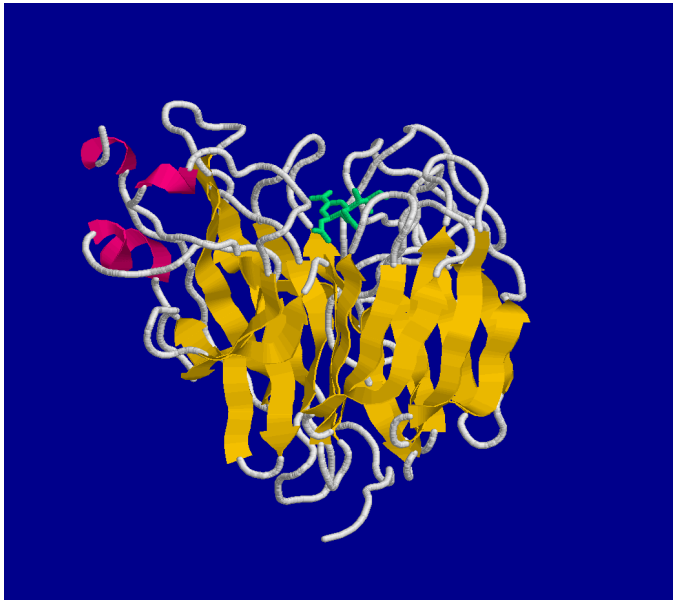
Contacts:
Favourable interactions determined by requirements on the relative orientation of the vectors are called contacts.

Groups of Contacts:
The structrual fragments interacting like backbone or sidechain centers or centers in hetero-groups or especially water are used to form groups of contacts

$\Rightarrow$ Coordinate-free Representation of a Geometry

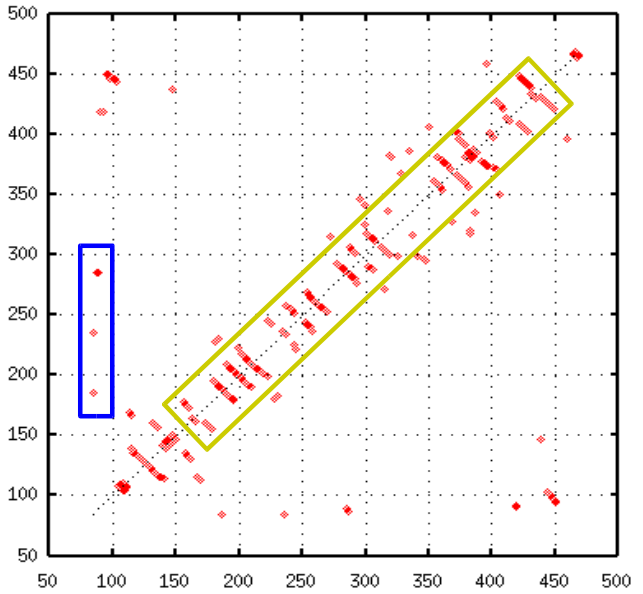
Contact Analysis II

Representation of Secondary Structure



(1NNC, 2.2Å, 1995)

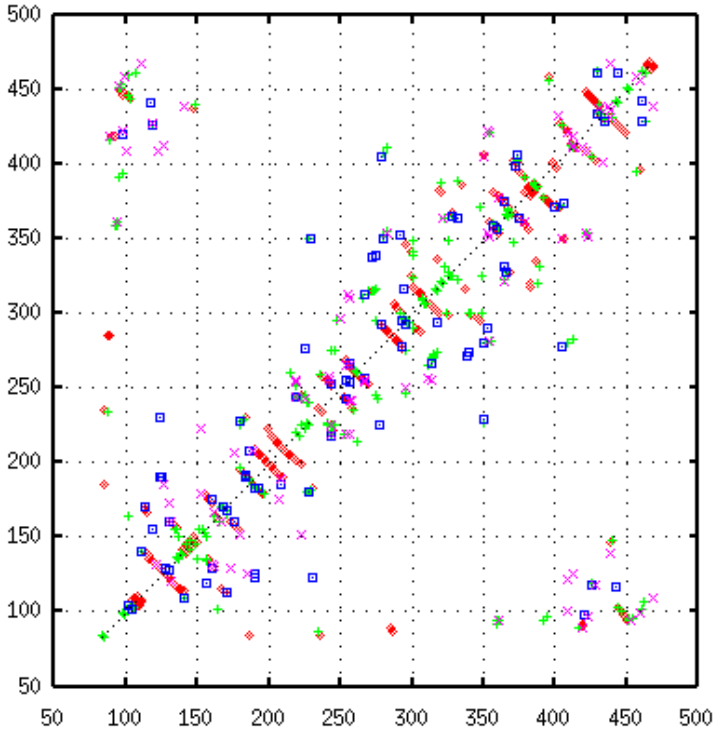
Contact Diagram (Backbone only)



The backbone contact diagram shows secondary structure forming (e.g. antiparallel β-sheet) together with other (e.g. long-range) contacts

Contact Analysis III

Complete Contact Diagram



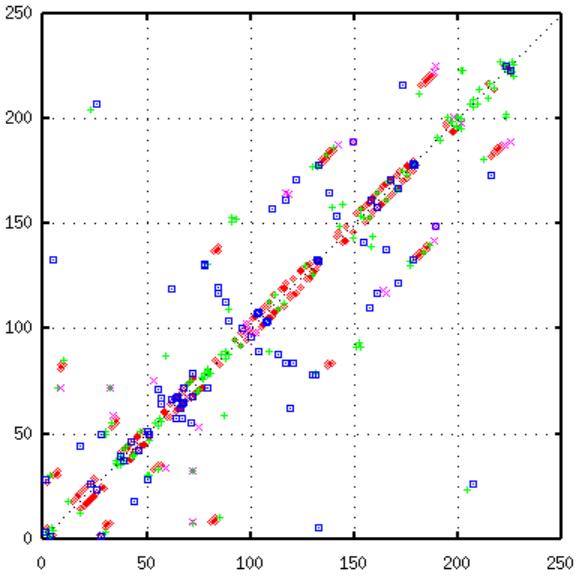
HBO:
Backbone-Backbone
Backbone-Sidechain
Sidechain-Sidechain

VDW:
specific van der Waals

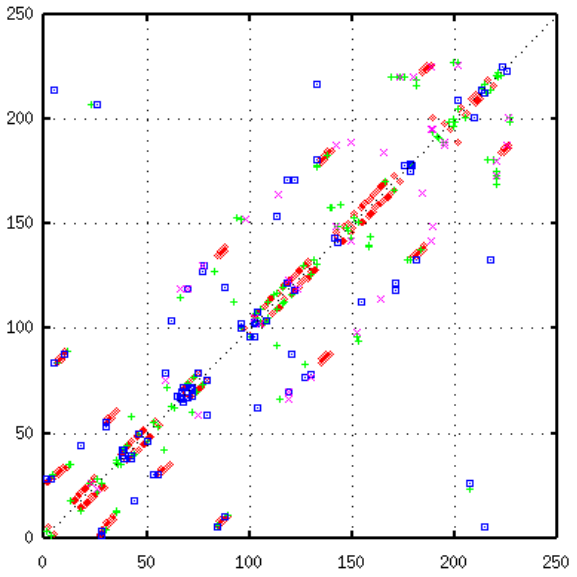
Previous contact diagram with all specific sidechain contacts

Contact Analysis IV

Application: Homology Models



Model 1



Model 2

Contact diagrams serve as fingerprints for easy comparison: two solutions to a homology problem show different diagrams

Contact Analysis V

Contact Vector: List of All Contacts in a Geometry

Here: Inhibitor Binding (PDB: 1NNC)

HBO contacts to backbone

```
HET3/L3A   GNA200B  ASP151A  SDBA X
HET3/x9A   GNA200B  TRP178A  SDBA X
```

HBO contacts to sidechain

```
HET3/B1EA  GNA200B  ARG118A  SASD X
HET3/L3A   GNA200B  ASP151A  SDSA X
HET3/L3A   GNA200B  ARG152A  SASD X
HET3/L14A  GNA200B  ARG371A  SASD X
```

VDW contacts to sidechain

```
HET3/x9A   GNA200B  TRP178A  SSV  X
HET3/L7A   GNA200B  ILE222A  SSV  X
HET3/B1GA  GNA200B  ARG224A  SSV  X
HET3/L9A   GNA200B  ALA246A  SSV  X
HET3/B2HA  GNA200B  ARG292A  SASR X
HET3/x21A  GNA200B  TYR406A  SSV  X
```

HBO contact to water

```
WAT/HET3   HOH123B  GNA200B  SDSA X
WAT/HET3   HOH285B  GNA200B  SDSA X
```

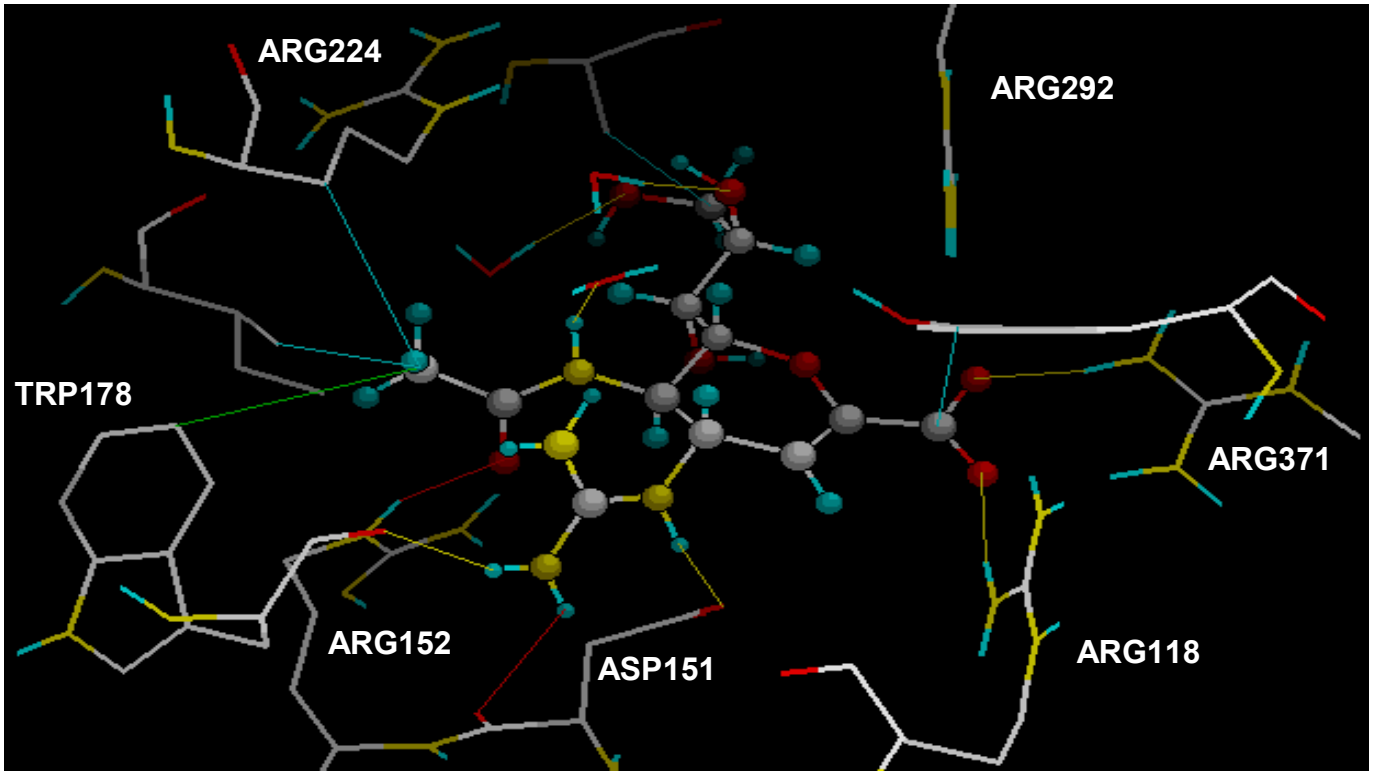
Bridging water

```
WAT/L8A    HOH121B  GLU227A  SDSA X
WAT/x14A   HOH121B  GLU277A  SDSA X
WAT/HET3   HOH121B  GNA200B  SASD X
```

Group of inhibitor contacts yields intuitive binding picture

Contact Analysis VI

Inhibitor Binding:
Geometric Representation of the Contact Vector



Contact Analysis VII

Dynamics of Secondary Structures: Contact Matrix

A stable β-sheet:

```
-----------
B1JA SER353A DSSP  AAAAAAAAAAA/AAAAAAAAAAA/AAAA/AAAA/AAAAAAAAAAAAAAAAA   92
B1JA TYR354A DSSP  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  100
-----------
B2JA TRP361A DSSP  MMMMMAMMMMMMAMMMMMMAMMMMMMMMMAAMMMMMMMMMMMMMMMMAMMMM  100
B2JA LEU362A DSSP  MMMMMMMMMMMAMMMMMMMMMMMAMMMMAAMMMAMMMMMMMMMMMMMMMMM  100
B2JA GLY363A DSSP  AAAAAAAAAAAAAAMAAAAAAAAAAAAAAAAAAAAAAAAAAAMMMAMMAAAAAAAA  100
B2JA ARG364A DSSP  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  100
-----------
B3JA GLU375A DSSP  MMMMMMMMMMMMMMMMMMMMMMMMMMMAMMMMMMMMMMMMMMMMMMMMM  100
B3JA MET376A DSSP  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM  100
B3JA LEU377A DSSP  MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMAMMMMMMMMMMMMMMMMM  100
B3JA LYS378A DSSP  MMMMAAMMAAAA/MAAMMAAAMAAMAAAMA/MMMMMMMMMMMMMMMMAMMMM   97
-----------
B4JA GLN392A DSSP  AAAA/AAA/////A//AA/A/A//A///AA/AAAAAAAAAAAAAAAAAAA   72
B4JA GLY394A DSSP  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   99
B4JA GLN395A DSSP  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA   99
B4JA THR396A DSSP  AAAAAAAAAAAAAAAAAAAAAAAAAAA/AAAAAAAAAAAAAAAAAAAAAAA   99
-----------
```

Gathering the contact vectors for various geometries from a MD simulation (columns) into a matrix shows the time evolution of the system

Contact Analysis VIII

Dynamics of Inhibitor Binding: Short Form of Matrix

HET4/x21A	GNA200B	TYR406A	SSV	--|X	82
HET3/x21A	GNA220B	TYR406A	SSV	XX|-	32
HET10/B2HA	GNA221B	ARG292A	SASD	-X|X	98
HET10/L14A	GNA221B	ARG371A	SASD	XX|X	00
HET10/B1EA	GNA222B	ARG118A	SASD	XX|-	35
HET10/L14A	GNA222B	ARG371A	SASD	XX|X	99
HET8/x5A	GNA241B	GLU119A	SDSA	--|X	48
HET8/L3A	GNA241B	ASP151A	SDSA	XX|-	39
HET9/x9A	GNA242B	TRP178A	SDBA	XX|-	41
HET9/x14A	GNA242B	GLU277A	SDSA	--|X	56
HET9/x5A	GNA243B	GLU119A	SDSA	--|X	60
HET9/L3A	GNA243B	ASP151A	SDBA	X-|-	6
HET9/x9A	GNA243B	TRP178A	SDBA	XX|X	50
HET7/L3A	GNA250B	ARG152A	SASD	XX|-	25
HET7/L3A	GNA250B	ARG152A	SSV	-X|X	46
HET7/x9A	GNA250B	TRP178A	SSV	XX|X	81
HET7/L7A	GNA250B	ILE222A	SSV	XX|X	45
HET7/B1GA	GNA250B	ARG224A	SSV	X-|X	47
HET5/B1HA	GNA262B	GLU276A	SDSA	XX|X	94
HET5/B2HA	GNA262B	ARG292A	SASR	X-|X	44
HET6/L9A	GNA263B	ALA246A	SSV	XX|X	93
HET6/B1HA	GNA263B	GLU276A	SDSA	XX|-	32

Contact:
stays
appears
vanishes
with respect to initial geometry

The development in the course of a MD simulation can be represented in a single column (*Representative Contact Vector*)

No Problem: Two MD Simulations

MD simulations starting from crystal structure (1NNC) with

- **Preparation**
  - **conventional (named PDB) or**
  - **according to our procedure (*CHEOPS*).**
- **identical Standard Protocol for 100 ps**

Both simulations stable according to global criteria **(cf. p. 27).**

Some aspects of a detailed inspection using our contact analysis follow...

No Problem: Inhibitor Binding

PDB

GNA200B	TYR406A	SSV	--|X	82
GNA220B	TYR406A	SSV	XX|-	32
GNA221B	ARG292A	SASD	-X|X	98
GNA221B	ARG371A	SASD	XX|X	00
GNA222B	ARG118A	SASD	XX|-	35
GNA222B	ARG371A	SASD	XX|X	99
GNA241B	GLU119A	SDSA	--|X	48
GNA241B	ASP151A	SDSA	XX|-	39
GNA242B	TRP178A	SDBA	XX|-	41
GNA242B	GLU277A	SDSA	--|X	56
GNA243B	GLU119A	SDSA	--|X	60
GNA243B	ASP151A	SDBA	X-|-	6
GNA243B	TRP178A	SDBA	XX|X	50
GNA250B	ARG152A	SASD	XX|-	25
GNA250B	ARG152A	SSV	-X|X	46
GNA250B	TRP178A	SSV	XX|X	81
GNA250B	ILE222A	SSV	XX|X	45
GNA250B	ARG224A	SSV	X-|X	47
GNA262B	GLU276A	SDSA	XX|X	94
GNA262B	ARG292A	SASR	X-|X	44
GNA263B	ALA246A	SSV	XX|X	93
GNA263B	GLU276A	SDSA	XX|-	32

CHEOPS

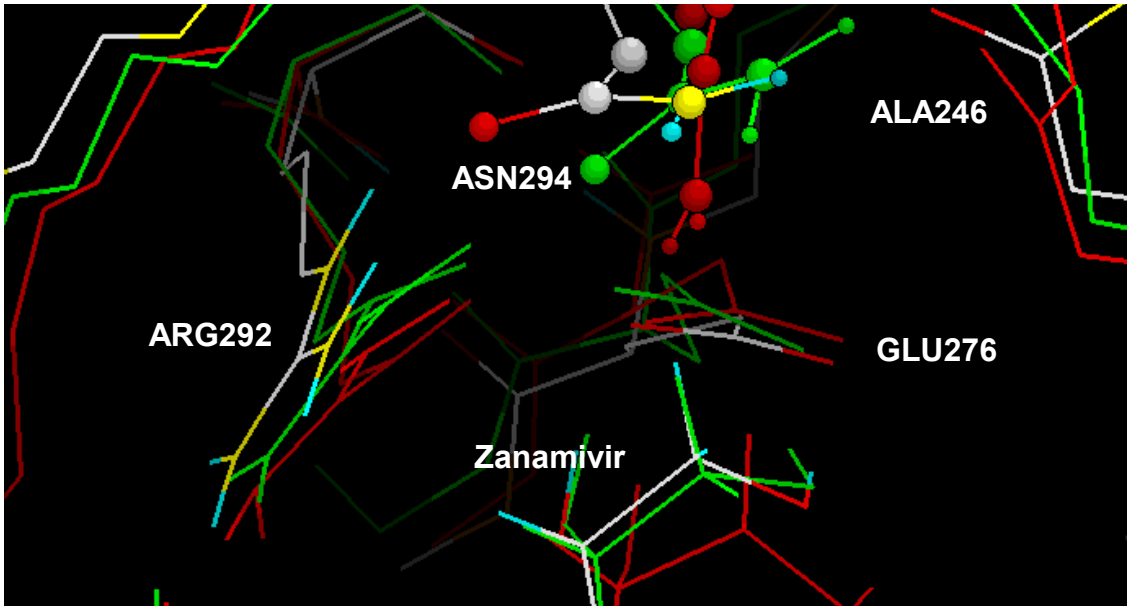
GNA200B	TYR406A	SSV	--|X	66
GNA220B	TYR406A	SSV	X-|X	53
GNA221B	ARG292A	SASD	-X|X	84
GNA221B	ARG371A	SASD	XX|X	97
GNA222B	ARG118A	SASD	XX|X	100
GNA222B	ARG371A	SASD	XX|X	60
GNA241B	ASP151A	SDSA	XX|X	67
GNA242B	TRP178A	SDBA	XX|X	76
GNA243B	GLU119A	SDSA	--|X	97
GNA243B	ASP151A	SDBA	X-|-	0
GNA243B	TRP178A	SDBA	X-|X	99
GNA250B	ARG152A	SASD	X-|-	9
GNA250B	ARG152A	SSV	-X|X	48
GNA250B	TRP178A	SSV	XX|X	90
GNA250B	ILE222A	SSV	XX|X	87
GNA250B	ARG224A	SSV	XX|-	40
GNA262B	GLU276A	SDSA	XX|X	95
GNA262B	ARG292A	SASR	X-|X	42
GNA263B	ALA246A	SSV	XX|X	85
GNA263B	GLU276A	SDSA	XX|X	90

***CHEOPS*-Preparation results in much more stable binding!**

No Problem: ASN294, Sidechain

PDB	ASN294A	ALA246A	SDBA	----------	0
	ASN294A	GLU276A	SDSA	--XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	98
	ASN294A	ARG292A	SASD	----------	0
CHEOPS	ASN294A	ALA246A	SDBA	XXXXXXXXXXXXXXXXXXXXX---XXXXXXXXXX-XX-XX-XXXXXXXXXX	89
	ASN294A	GLU276A	SDSA	----------	0
	ASN294A	ARG292A	SASD	XXXXXXXXXXXXXXXXXXXX--XXXXXXXXXX-XX-XXXX--XXXXX-X-X	86

Starting Geometry *CHEOPS*, 100 ps PDB, 100 ps *CHEOPS*



PDB: incorrect orientation of ASN294 (cf. p. 4) is not improved during simulation and bumps against inhibitor (cf. p. 2)

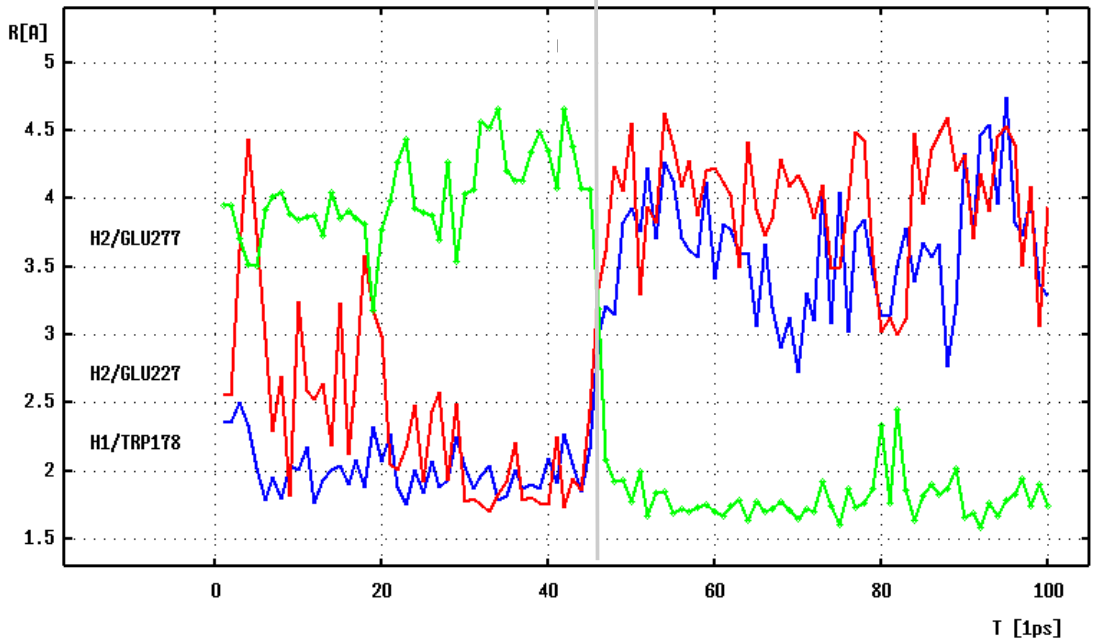
***CHEOPS*: No Problem!**

No Problem: Guanidine Group (Inhibitor)

```
PDB      GNA242B  TRP178A  SDBA XXXXXXXXXX-XXXXXXXXXXXXX--------------------------  42
         GNA242B  GLU227A  SDSA -----X-------X--XXXXXXX-X--------X--X----XX------XX-  30
         GNA242B  GLU277A  SDSA ------------------------XXXXXXXXXXXXXXXXXXXXXXXXXXX  54

CHEOPS   GNA242B  TRP178A  SDBA XXXXX-XXXXXX-XXXXXXXXXXXX----XXXXXXXX-XXXXXXXXX-X--X  76
         GNA242B  GLU227A  SDSA -----X-------------------XX-------------------------   8
         GNA242B  GLU277A  SDSA ------------------------------------------------------   0
```

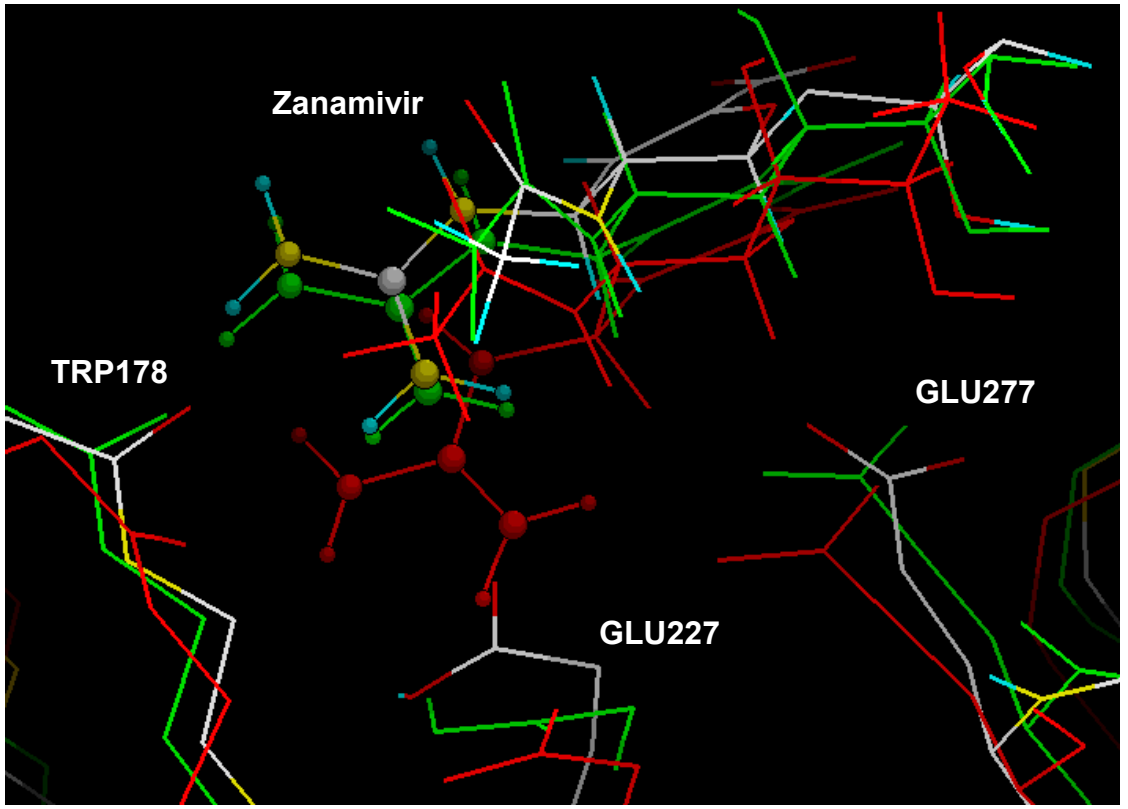
Significant Hydrogen Bond Distances for PDB



The contact changes of the guanidine group (GNA242B) would hardly be detected looking at atom-atom distances!

No Problem: Guanidine Group (Inhibitor)

Starting Geometry *CHEOPS*, 100 ps PDB, 100 ps *CHEOPS*



CHEOPS **shows no deviation for Inhibitor!** (cf. p. 2)

No Problem: Evaluation of Crystal Structure

		Preparation PDB	Preparation *CHEOPS*
Δ_{cart}	[Å]	1.2	1.0
R_{gyr}	[Å]	19.5	19.4
E_{stab}	[kcal/mol]	-159.0	-157.8
ASN294		rot. by 60°	in place
Guanidine		dislocated	in place
Δ (Inh.)	[Å]	1.8	0.8
Starting Structure		not stable	stable

⇒ **Crystal structure - only when properly prepared - supports experiment and can be studied further**

Summary

X-ray structure suffers from subjective Interpretation:
- **statistical criteria for quality are not sufficient**
- **established geometry parameters are disregarded**

Stable starting geometry is prerequisite for Protein Modeling:
- **determine position/orientation of hydrogen atoms**
- **errors cannot be corrected by optimization runs**

The *CHEOPS* way:
- **automized structure preparation (incl. ionization *and* hydrogen bonding network)**
- **remaining hot spots (if any) treated manually**
- **MD simulation with standard protocol w/o restraints**
- **contact analysis: innovative condensed representation of 3D structures and their dynamics**

⇒ **Assessing Stability of the Crystal Structures**