REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE:	3. REPORT TYPE AND DATES COVERED
		Final Report 15-Sep-2004 - 14-Oct-2006

4. TITLE AND SUBTITLE

Multiscale Modeling of Complex Systems Conformational Transitions in Proteins

5. FUNDING NUMBERS

W911NF-04-1-0381

6. AUTHORS

Thomas Beck, Nimal Wijesekera (postdoc), David Rogers (grad), Roman Petrenko (grad)

8. PERFORMING ORGANIZATION REPORT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES

University of Cincinnati
Office of Sponsored Programs
Suite 7148, Edwards One
Cincinnati, OH 45221 -0627

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

47076-MS.1

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The abstract is below since many authors do not follow the 200 word limit

14. SUBJECT TERMS

protein loop modeling, multiscale method, Monte Carlo, regulatory domains, devices

15. NUMBER OF PAGES

Unknown due to possible attachments

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Report Title

Multiscale Modeling of Complex Systems Conformational Transitions in Proteins, final report

ABSTRACT

This project concerns development of new simulation methods for modeling protein conformational changes. The purpose is to develop accurate methods for generating a coarse Hamiltonian for use in Monte Carlo simulation. Our method is unique in that we derive the coarse scale energy function based on fine scale (all atom) simulation. We are developing a general scheme which employs both replica exchange (multiple temperatures) with resolution exchange (multiple scales). The method will lead to order-of-magnitude speedup in accurate simulations of loop conformations and protein folding more generally.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Several papers resulting from this project are in progress.

Number of Papers published in peer-reviewed journals: 0.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

1) Seminar at the "State-of-the-art, developments and perspectives of electronic structure techniques in condensed matter and molecular physics" CECAM workshop in Lyon, France, 20th-24th June 2005. The PI (Beck) was a co-organizer of this international conference.
2) Seminar at the "Biological ion channels: structure and function" workshop at Telluride, CO. July 31-Aug 5, 2005. This meeting was attended by 25 top ion channel modelers and experimentalists. Our work on modeling the pH sensor loop in ClC-2 attracted a lot of attention at the workshop.
3) Seminar at the "Epithelial ion Channels" workshop at Telluride, CO. July, 2006. This meeting was attended by roughly 25 top ion channel modelers and experimentalists. This talk presented a new model for the bacterial ClC transporters, and how the transporter mechanism works. Following this presentation, we became aware of a new experimental paper by Chris Miller (who discovered the transporter behavior), which closely agreed with our modeling. In addition, work from our group was presented related to coarse-graining models and simulations of folding heat capacities for the villin headpiece peptide. Our CG models agree closely with all-atom simulations for the heat capacity curve.

Number of Presentations: 3.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

(d) Manuscripts

This project has been large methods-development effort, and the results are just now leading to manuscripts for submission. We have several manuscripts in preparation: 1) a new configurational bias Monte Carlo technique for re-generating fine scale chains following coarse-scale simulations. This 'completes the cycle' in the multiscale modeling, which allows for iterative communication between the fine and coarse scales during a folding simulation. 2) We have derived a new formula for molecular potentials of mean force (PMFs) utilizing the Potential Distribution Theorem developed in our book (co-authored with Mike Paulaitis and Lawrence Pratt, Cambridge University Press). This formula partitions the PMF into a direct gas phase part and a solution-phase correction. We have used this formula to derive PMFs for simple polyethylene chains and for peptides, and have employed those PMFs in coarse-level simulations. These PMFs allow one to reproduce global features (like heat capacity curves) but do not lead to accurate folds. Local short-ranged features are now being built in to lead to accurate folds. This work is now being written up for publication. 3) We have derived a new formula for free energy bounds for molecular solvation. These bounds will lead to much more accurate and efficient means of obtaining solvation free energies, and resulting PMFs for polymers and proteins. Those PMFs will figure prominently in our coarse-scale simulations. 4) New results pertaining to loop modeling of the ClC-2 pH sensor domain with our new multiscale methods will lead to a manuscript also.

Number of Manuscripts: 4.00

Number of Inventions:

Graduate Students

NAME	PERCENT_SUPPORTED	
Roman Petrenko	0.25	No
David Rogers	0.10	No
FTE Equivalent:	**0.35**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED	
Nimal Wijesekera	1.00	No
FTE Equivalent:	**1.00**	
Total Number:	**1**	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

TOC:

Our work has progressed on several fronts.

1) We have written a full Monte Carlo simulation code for peptides using the CHARMM forcefield. This code was written from scratch by my graduate student Roman Petrenko. He successfully tested his code by comparison with CHARMM MD simulations. He also has employed the forcefield developed by Hansmann, one of the pioneers in applying the replica exchange Monte Carlo method to proteins.

2) We have implemented the replica exchange Monte Carlo method in simulations of the 35 residue pH sensor loop domain of the human ClC-2 channel. This channel is involved in the acid secretion mechanism of the stomach, and has possible device applications (bio-sensors and energy transduction). We have developed a homology model of the ClC-2 channel based on the bacterial ClC structure, and we found that the pH sensor sequence discovered by Cuppoletti's group (PI of our MURI award) is located on the extracellular side of the channel. The loop is too large to model accurately with homology modeling, however. We plan to perform extensive Monte Carlo simulations of this loop when the key glutamate residue is charged and neutral to investigate the conformations of the loop related to its regulatory role. Petrenko has performed preliminary replica exchange MC simulations of the loop and has located several trial conformations for further analysis.

3) Both Petrenko and my postdoc Nimal Wijesekera have developed coarse-graining Monte Carlo methods based on the original idea of Bai and Brandt (see attached nimal_report). They successfully compared their codes to results obtained in the original paper by Bai and Brandt. Therefore, we have the coarse-graining step developed and in hand. A key feature of the general multiscale approach is to be able to make corrections on the fine scale once much cheaper (at least an order of magnitude) simulations have been performed on the coarsened chains. We have now developed a stochastic method for generating a new fine scale chain based on a coarse configuration. Modeling on the coarse level allows for large amplitude moves in the configuration space. This is the general spirit of the multiscale method in which the fine and coarse levels interact repeatedly to rapidly obtain equilibrum statistics (removal of critical slowing down). We have now successfully tested a configurational biased Monte Carlo (CBMC) technique for the fine-scale chain generation. These methods are new and will allow for the proper coarse-fine interaction for the first time. All previous coarse-graining procedures generate the coarse hamiltonian in an ad hoc manner.

4) The next step is to unite the coarse-graining and replica exchange methods, combining the best ideas of both. This combination will allow for highly efficient sampling of the conformational space, yet based on the molecular level hamiltonian. That work is in progress. After initial testing, we will employ this method in modeling of the pH sensor domain of ClC-2.

5) We have developed a new multiscale loop-modeling strategy (attached 'roman_report') for sampling the ClC-2 pH sensor domain. This work is in progress and should lead to efficient modeling of the loop in the presence of the rest of the protein.

6) A new method for accurate generation of coarse-level potentials is under development (see attached rogers_CG_report). Instead of postulating a functional form for the coarse potential related to the fine Lennard-Jones potential, we generate the potential accurately with our new PMF methods for complex molecular systems.

Detailed reports:

Wijesekera: Outline of the multiscale Monte Carlo method to study flexible domains of proteins

The main objective of this work was to find a scheme to communicate between two different scales of representation of a complex system such as a polymer or a protein. Such a scheme will allow us to develop a highly efficient multiscale simulation technique to study complex systems. Methods are already available to move from a fine-level simulation to a coarse-level one, and one such approach is found in Ref. [1]. First, we opted the method developed in Ref. [1] as the scheme to coarsen a fine chain and implemented it successfully. The original coarsening method was applied for a polymethylene-like bonded linear chain. We applied the coarsening scheme to more general case of branched chain with the intention of using it for proteins in a later stage. However, it is done with the following two modifications. First, on the fine scale, only the dihedral angle is allowed to vary while the bond length and bond angle are kept fixed for all atoms. However, on the coarse scale all three internal coordinates are varied. Second, only the side chains of the fine chain are coarsened while the fine-scale backbone with the points of attachment is not coarsened (i.e., the coarse backbone is the same as the fine backbone including the points of attachment).

We employ configurational-bias Monte Carlo (CBMC) technique, introduced by Rosenbluth and Rosenbluth [2], on the fine-scale simulation of the branched chain. Since our simulations are not grid-based, we are utilizing the off-lattice CBMC technique in this study, and as a result, we have encountered many technical difficulties coding the algorithm we have proposed. In CBMC the atoms are grown one by one to generate a new conformation of a chain. One of the advantages of this method is that the growing procedure can be started from an already partially grown chain, and this allows us to grow only the side chains on a backbone obtained by some other means. In our algorithm, backbone confirmations are generated on the coarse level using the Metropolis Monte Carlo technique, and from time to time the coarse backbone is used as a feedback

to the fine simulation. Thus, we have developed a device to transfer coarse-scale information to fine scale. The advantage of using a coarse level simulation for the backbone is that the long range movements of the chain are simulated in a computationally less expensive and efficient way compared to the fine-level simulation of the whole chain.

The combining of the fine level and coarse level algorithms into a single algorithm is achieved in the following way. We begin with the CBMC on the fine level, and after every several thousands of iterations, the coarse level simulation, which involves only the backbone, is initiated. Next, the CBMC is performed on an accepted backbone confirmation from the coarse simulation, and the fine-level CBMC is continued. The shuttle between the two scales can be done as many times as preferred. The coarse-level Hamiltonian is constructed [1] based on the information obtained from the restricted fine chains. In addition, the coarse Hamiltonian is continually updated while CBMC is running on the fine level so that it keeps evolving continuously. Thus, we have found a method to cycle back and forth between coarse scale and fine scale representations.

We have just completed the main code of this multiscale approach to study complex systems such as polymers. The code is written using C++ language. The code development has been done in number of stages. First, a Metropolis Monte Carlo technique to simulate branched polymers with or without an external potential (Lennard-Jones potential) is developed, and then a code is developed to implement the coarsening scheme described in Ref. [1]. Next, a complete CBMC code is written to simulate a branched polymer, and finally a unified code is obtained to perform multiscale simulations. The preliminary results are promising but not without any issues; one of them is that we have observed a very low level of acceptance (less than 5 percent) of the coarse backbone on the fine level, which is contrary to what we have expected, and currently we are investigating this issue. The code is now in the stage of further testing and fine-tuning, and this is required before producing reliable results.

[1] Dov Bai and Achi Brandt, Multiscale Computational Methods in Chemistry and Physics, A.Brandt et al. (Eds.) IOS Press, 2001, pp. 250-266.

[2] M.N. Rosenbluth and A.W. Rosenbluth, J. Chem. Phys. 23, 356 (1955).

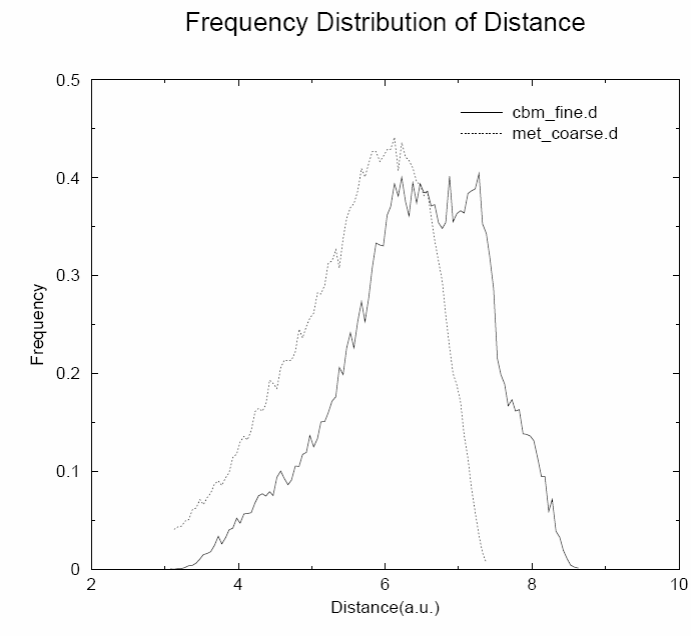


Figure 1. This figure displays the end-to-end distance distribution for fine and coarse branched polyethylene chains.

Report on multiscale monte carlo method in protein loop structure prediction

Roman Petrenko

December, 2006

The project according to the proposal was split on two parts:

- development of the effective potential to be applied on reduced (coarse-grained) structures, which are free in space;
- multiscale simulation of protein loops.

The ECEPP/2 force field developed by Scheraga (Scheraga) was chosen, in which bond lengths and bond planar angles are fixed and the only degrees of freedom of protein are dihedral angles. To reduce computational cost water environment was modeled with SASA (solvent accessible surface area) approach. I implemented parallel tempering Monte Carlo method in my own program and fine scale simulations were carried out on test proteins: villin headpiece and polyalanine-10 to reproduce the results of SMMP-software package (Hansmann). Fine-scale simulations were good for poly-alanine (the final structure is just a helix), but not very good for villin headpiece.

After that the coarse-grained energy potentials were constructed as in (Bai). Namely, the coarse-grained model of protein consisted of all backbone atoms and one atom

representing the side-chain. The potentials correctly identified temperature dependent global parameters of protein, as the energy and specific heat of a protein (pic.1), but it fails to reproduce the local properties (pic.3, 4). The stumbling point happened to be the reconstruction of the chain on fine scale.

Given efficient algorithm for sampling dihedral angles while keeping the ends fixed, the simulation of protein loops must be easier since the conformational space of a chain is greatly reduced. But the sampling itself becomes a big problem – a simple pivot or crankshaft update of dihedral would brake the internal structure of loop. And naïve way of imposing constrained on one of the ends in the loop with the anchor residue yields very low acceptance rate. Additionally acceptance rate is greatly reduced when taking the rest of the protein into account.

It is very important to realize that loops are located in the cradle of protein environment, which cannot be neglected. Despite some success of loop modeling with flexible stem geometries (Floudas, Monningam), other researchers claim that protein environment greatly improves loop prediction. With currently available methods it is possible to build a loop up to 15 residues long (Friesner group) by inserting loop fragments from extracted from PDB database of loop segments.

The next stage of the development of modeling loops should be based on experience with fine scale simulations of loops. That is before building coarse grained model, we should clearly devise a sampling procedure for loops with fixed ends. In the spirit of multiscale I propose new method to handle that. Take a loop of N residues, cut it in the middle. Simulate two pieces with biasing of the free ends towards each other. Then measure the middle point of the two free ends. The next step is to cut loop in three pieces and place the middle piece to one of the positions of the middle point we measured above. Now simulate loop (1: N/3 and 2N/3:N) with one fixed end and one free end and the middle loop with two ends free. Again bias the ends to meet with each other. Then iteratively proceed to smaller pieces. This method is very similar to the one we used for free proteins, the more pieces we have, the higher frequencies are being sampled.

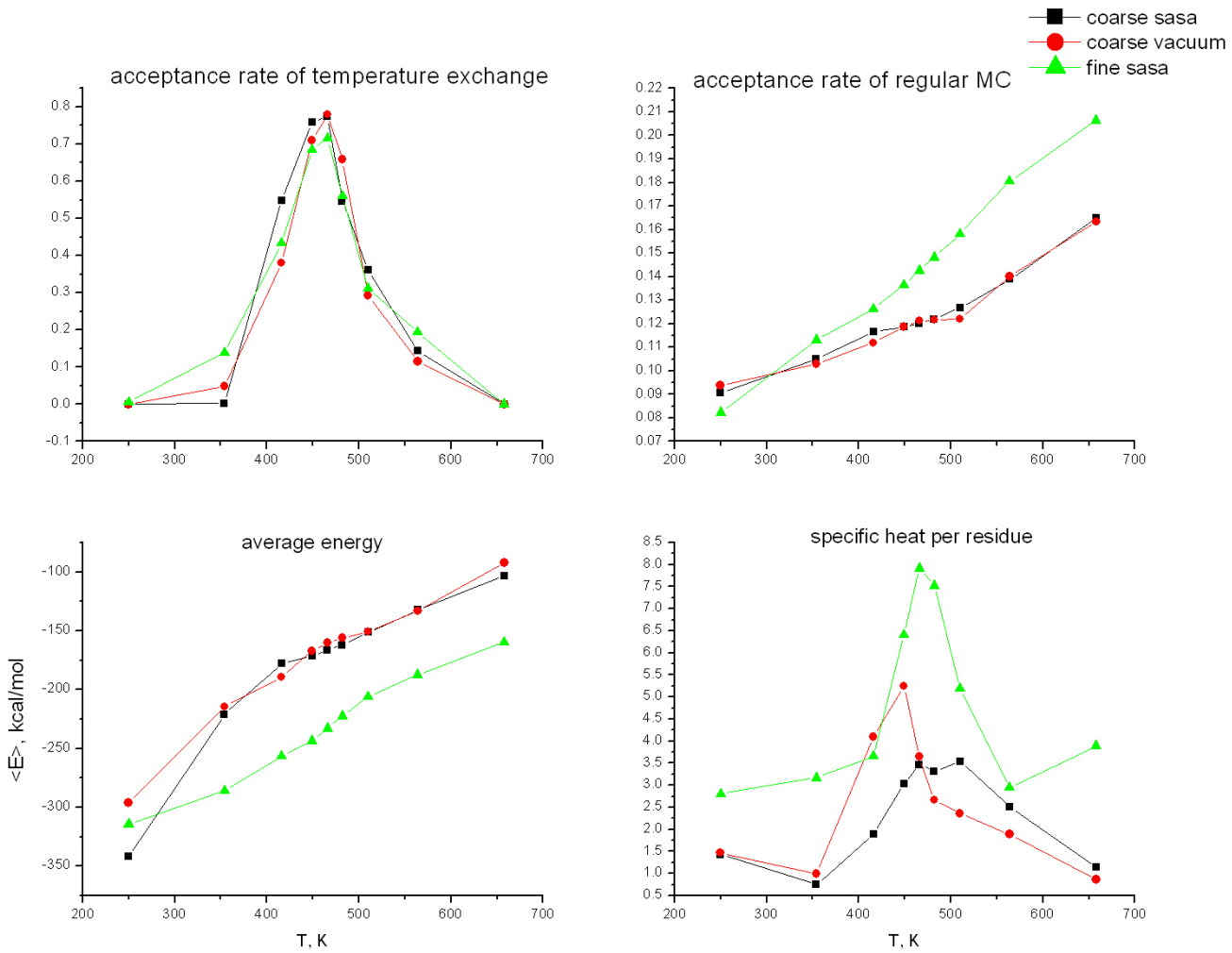
Recent studies (Rose) have shown that protein folding problem is mainly governed by three factors:

- compactness of the structure (gyration radius greater that the one for the globular polymer of the same length should be avoided),
- backbone hydrogen bonds should are favored to intermolecular contacts,
- secondary structures bias to backbone torsions should be used.

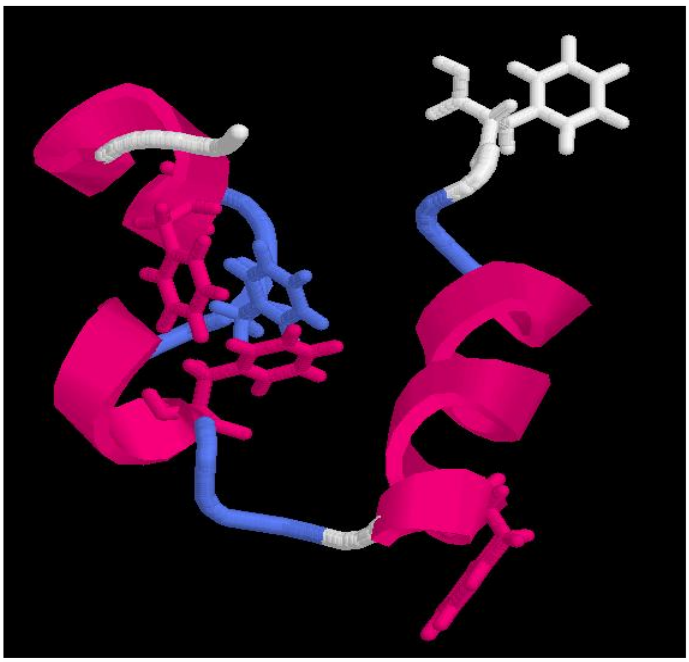
To avoid frequent local clashes of the atoms in the neighboring residues a database called DASSD (**D**ihedral **A**ngle and **S**econdary **S**tructure **D**atabase of Short Amino acid Fragments) must be used in the sampling procedure of backbone dihedral angles (DASSD). The database contains dihedral angle values and secondary structure details of short amino acid fragments of lengths 1, 3 and 5.

The analysis of protein structures in PDB database shows (pic. 5) that if two residues are separated by less than 6 peptide bonds the specificity of side chains almost don’t play any role. That is, on this length scale the backbone conformation of one of the limited set of all allowed conformations without steric clashes and side chains give the backbone a preference to one state or another.

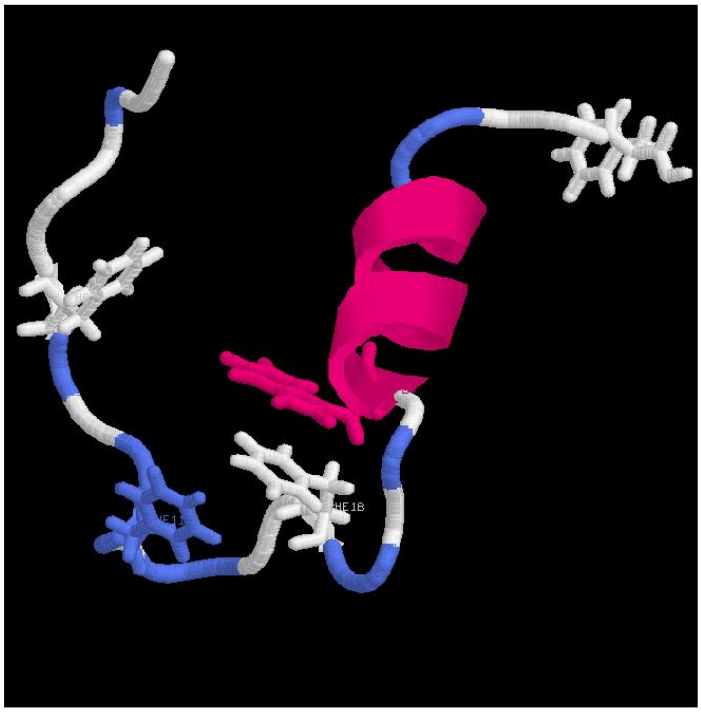
Pic. 1


coarse sasa
coarse vacuum
fine sasa
acceptance rate of temperature exchange
acceptance rate of regular MC
average energy
<E>, kcal/mol
T, K
specific heat per residue
T, K

Pic. 2

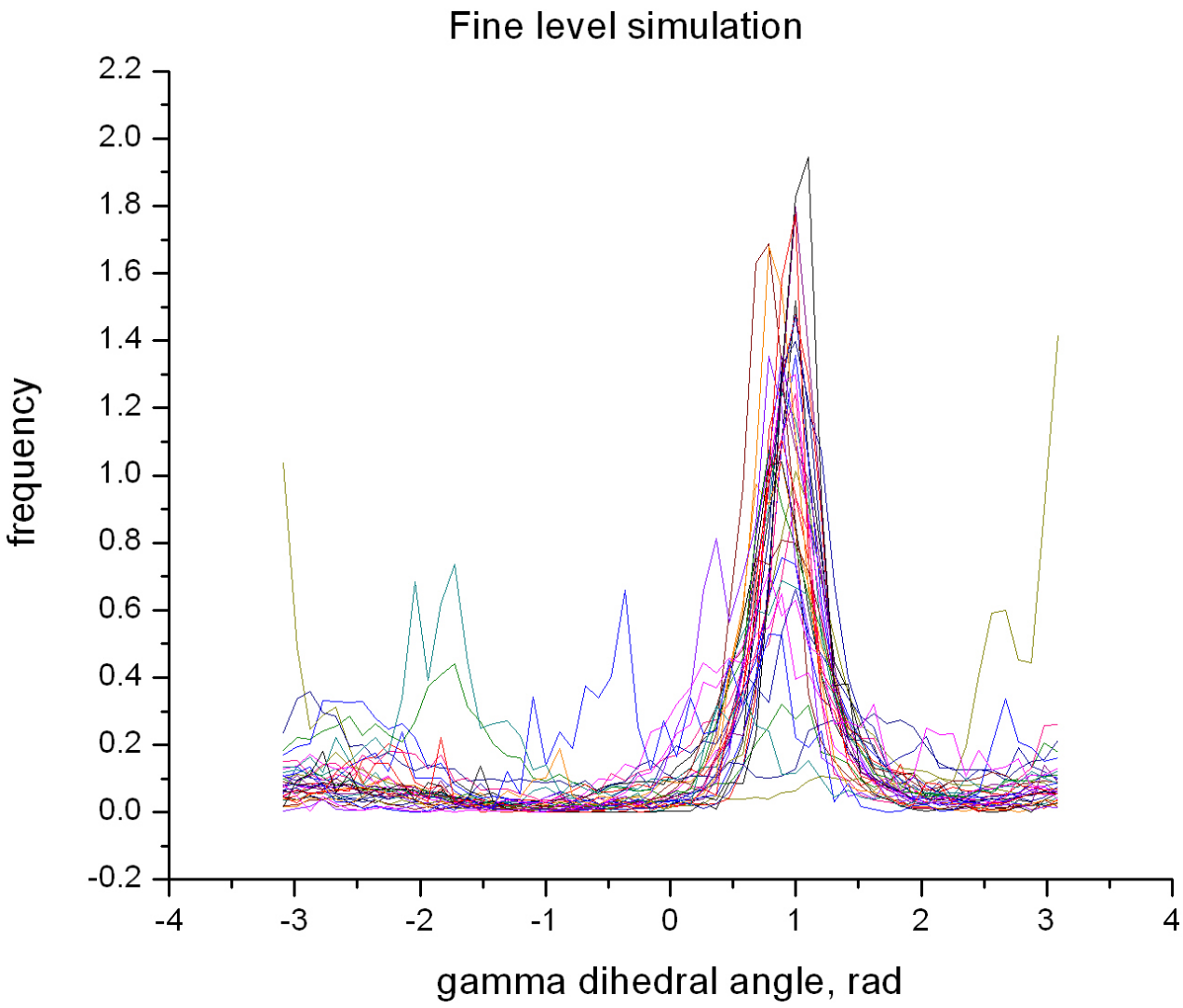


Pdb structure of villin headpiece (1vii.pdb)

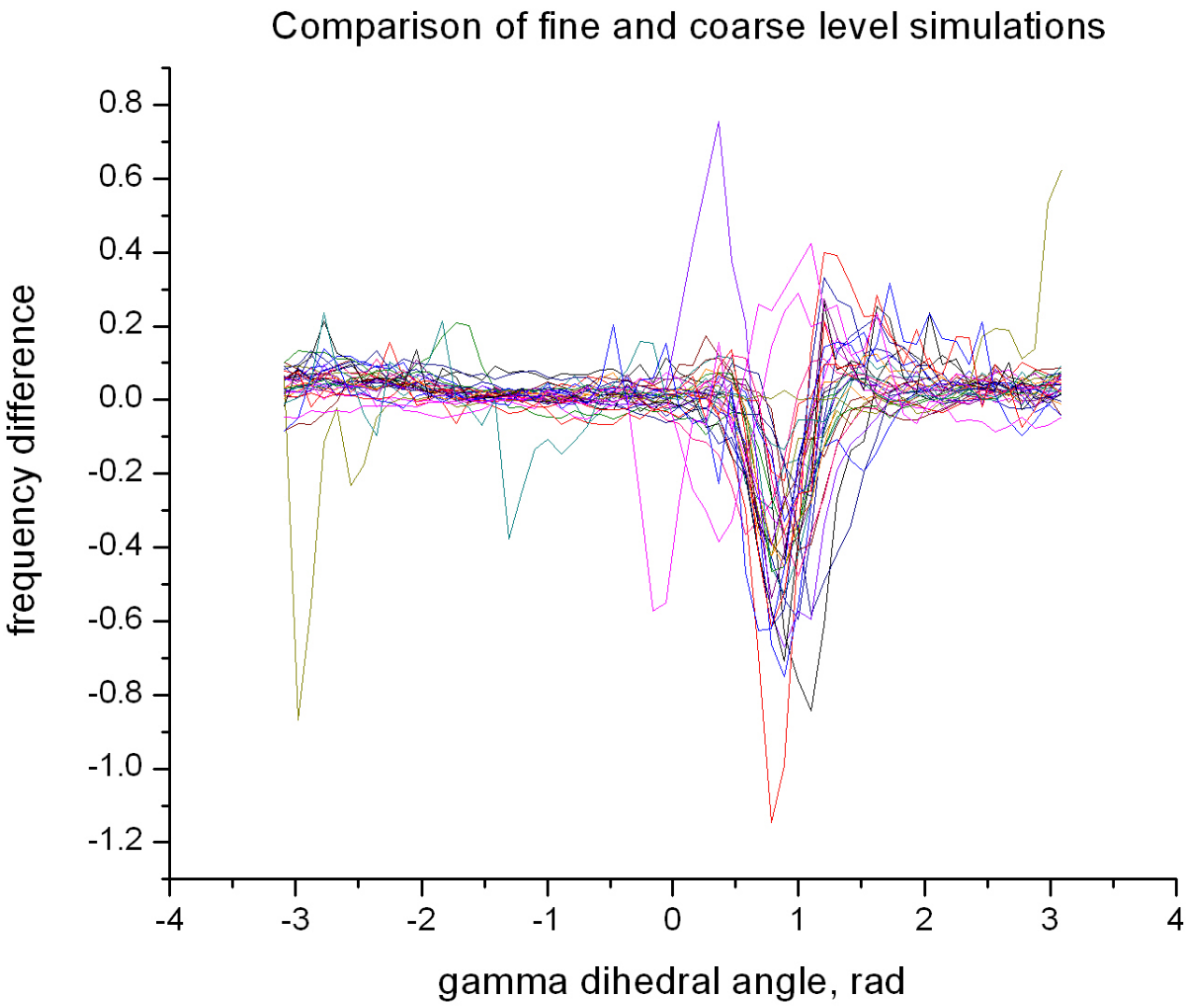


Lowest energy structure after 4000 mc sweeps at 9 temperatures

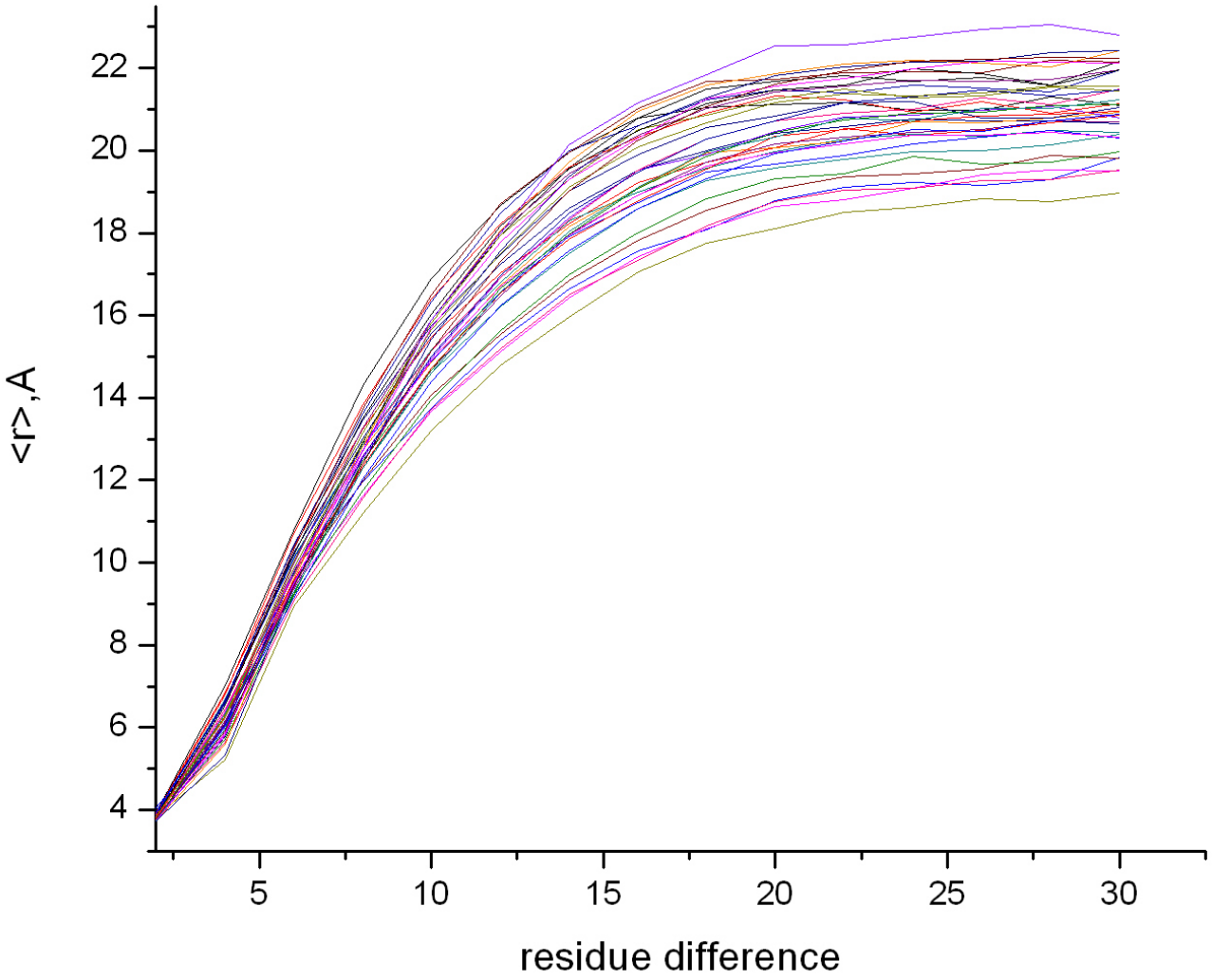
Pic. 3


Fine level simulation
frequency
gamma dihedral angle, rad
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-4
-3
-2
-1
0
1
2
3
4

Pic.4


Comparison of fine and coarse level simulations
frequency difference
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-4
-3
-2
-1
0
1
2
3
4
gamma dihedral angle, rad

Pic. 5



References:

(Scheraga)
Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 6472.

(Hansmann)
F. Eisenmenger, U.H.E. Hansmann, S. Hayryan, and C.-K. Hu, *[SMMP] A Modern Package for Protein Simulations*, Comp. Phys. Comm. **138** (2001) 192-212.

(Bai)
Bai, D.; Brandt, A. NATO Science Series III, 177, IOS Press, 2001, pp. 250-266.

(Floudas)
Mönnigmann M and Floudas CA. Protein Loop Structure Prediction with Flexible Stem Geometries. Prot Struct Funct Bioinf. 2005;61:748-762.

(Friesner)
K.Zhu, D.Pincus, S.Zhao, R.A.Friesner, Prot Struct Funct Bioinf, 65, 438-452 (2006)

(Rose)
N.Fitzkee, P.Fleming, G.D. Rose, Prot Struct Funct Bioinf, 58, 852-854 (2005)

(DASSD)
S. Dayalan, N.D. Gooneratne, S. Bevinakoppa, H. Schroder, Dihedral angle and secondary structure database of short amino acid fragments,
Bioinformation 1(3): 78-80, (2006)

David Rogers
December 13, 2006
Coarse Graining Work Report

I have attempted to methodically test the assumptions implicit in course graining studies. The basic idea of this field is that unimportant degrees of freedom of the system can be accounted for in an approximate way. One of the major assumption that has been made in light of the practical difficulty of constructing a course-grained Hamiltonian is the assumption that intra-molecular interactions (e.g. bonds, angles and dihedrals) that are not directly connected behave independently and thus their contribution to the potential energy is strictly additive. Notably, this assumption has figured prominently in the development of traditional atomic-scale forcefields, where bond and angle potentials handle the local interactions along the polymer chain and Lennard-Jones and electrostatic interactions (which do not apply for pairs of atoms involved in the same bond or angle) handle the long-ranged part of the potential.

Unfortunately for coarse-grained studies, the fine-scale Lennard-Jones force acts between atoms which would be considered bonded neighbors at the coarse level. This force effectively generates a correlation between all coarse-scale interactions. Thus in generating a coarse-scale Hamiltonian, we are facing a variant of the problem that atomic-scale forcefield designers struggle with. The common solution is to adopt an empirical dispersion or Lennard-Jones interaction and to fit the bonded interactions to reproduce the remaining potential energy difference. Although this strategy is useful in principal, the fitting usually requires an empirical form for the bonded interactions as well. These forms are well studied for atomic interactions (e.g. harmonic, cubic stretch, Ryckaert-Bellmans, etc.), but mesoscale forms have received less attention. An interesting consideration is that if a proper coarse-scale dispersion interaction could account for nonlocal effects at the coarse scale, then it would make the coarse-scale bonded interactions uncorrelated at longer ranges – making them truly local in nature.

In order to test this assumption, the degree of correlation between bonds, angles, and dihedrals induced by the dispersion potential in a united-atom model was investigated. The model used is the polymethylene system studied by Bai and Brandt. Since the dispersion term for their energy function is to be neglected for atoms separated by less than four bonds (where most models use three), it serves to find a lower bound on the amount of correlation between successive bonded interactions that can be accounted for by a dispersion potential. The results shown in Figures 1, 2, and 3 represent the difference in occupancy probabilities from a system where bonds, angles, and torsions were considered independent and the actual joint probabilities from the original data. Note that if no Lennard-Jones interactions had been included in the polymethylene simulation, the difference would be zero and gives an estimate of the statistical noise – about 0.0004 occupancies per sample per bin.

Interestingly, the bond-angle correlation is only slightly above the statistical uncertainty, whereas the angle-dihedral correlation is clear. This shows that the bulk of the polymer's propensity to collapse into a globule manifests itself in the correlations between local angles and dihedrals of the polymer. Figure 3 shows an expanded view of this probability change induced by the nonlocal dispersion interactions. It also shows the subtlety involved in the choice of dispersion interactions.

Because of the Lennard-Jones potential, it is assumed that correlations between bonded interactions further apart are present and decreasing in magnitude with distance. However without this potential present at the coarse scale, the only observable difference in joint probability densities is very small and expected to decrease with distance along the chain. This suggests a measure of goodness for the choice of long-ranged interactions that cannot be calculated, as it involves a large number of quantities which are very close to the limit of detection.

An alternative criteria for choosing long-ranged interaction potentials can be arrived at by considering how the form of the long-ranged potential should behave. If two units of the polymer were coarsened into one "bead," then the pairwise potential at large separations should behave as the sum of the two atom's contributions – both acting from the coarse bead position. At close distances, it should represent the average interaction between the atoms and its target, excluding any sort of bond the atoms and the target may share. This, however, is just the form of a potential of mean force (PMF) for the Lennard-Jones potential. Figure 4 shows the PMF between two segments of polymer of length 2 (PMF and Cubic Spline) and the pairwise Lennard-Jones energy multiplied by 4. Since there are 4 pairwise interactions between the two segments, these two should be approximately equivalent at large separations. Note that the uncertainty in the PMF grows as the distance becomes small due to the large repulsion between beads. In practical calculations involving the PMF, the interaction energy for short ranges can be calculated via a fitted "soft" potential (e.g. 9-6 form). For larger separations where the original data is accurate, a cubic spline interpolation can be used to generate a continuous potential.

Using this PMF, it is possible to construct the complete coarse-scale potential energy function by re-weighting techniques. The additivity of dispersion potentials and local potentials implies that the probability of the observed configuration is the product of the probability of the given local variables multiplied by the Boltzmann factor for the coarse-scale potential. Intuitively, local probabilities are given higher weight where the PMF is larger (less favorable non-local configuration) to find the corrected coarse-scale probabilities for local variables. Since the correlation introduced to the local bonds by the LJ force is small, this re-weighting (intuitively doing away with the LJ-induced correlation) usually has a small effect on the correlation of the coarse local variables. This means that whatever correlation is present must still be accounted for by cross-terms in the local energy function – even though the choice of LJ-exclusion separations is still arbitrary at this point.

Figure 5 displays an example fine-scale configuration (stick model) and its coarsened counterpart (spheres). The data mentioned here makes use of a polymer with chain length 12 and a 2:1 coarsening ratio.

Unfortunately, this process still does not guarantee that the local variables will be uncorrelated – only that there will not be some strange correlation arising between atoms very far apart in the bonded chain that must be accounted for with bond/angle potentials. In fact, the very idea of eliminating degrees of freedom from a system implies that the eliminated variables must make themselves felt in some way through the remaining particles – suggesting large correlations are possible at the coarse scale. Indeed this is what has been found. Figures 5 and 6 show the unweighted and re-weighted coarse variable joint probabilities for all central bonds, angles and torsions. The distributions involving end-bonds, angles, or torsions are not significantly different, however. The re-weighted probabilities are nearly identical to the unweighted. On the coarse-scale, the bond-distances

and angles are clearly distributed differently conditional on the value of the other. This is due for the most part not because of the fine-scale Lennard-Jones force, but because of the geometry of the fine-scale bonds and angles themselves.

Although the aforementioned probability-based technique works when the bonded interactions are assumed to be independent, when this assumption is not made the joint probabilities of neighboring interactions along the chain are correlated with one another in a complicated way – making an iterative approach necessary to reproduce their distributions on the coarse-scale. In these cases, force-matching is the best method as it can be used to generate coarse-grain potentials in a non-iterative fashion and mathematically minimizes the squared difference between coarse and average fine energies for any given form of the coarse potential. As noted by Izvekov et. al. (2004), one of the most powerful choices for coarse-potential form is the cubic spline. A singular-value-decomposition approach was used in the implementation of a computer program to carry out force-matching. In order to make the most use of sample data, the cubic spline mesh points are chosen so that an equal number of fine-scale data points falls in each mesh interval. This program is still in the testing and validation stages. When it is operational, it will be able to produce coarse potential functions which take local bond, angle, and torsion correlations into account. Finally, the complete cubic-spline mesh data will show the form of the bond-angle and bond-torsion cross-terms – allowing for the design of much-needed empirical expressions which can be fit much more rapidly from fine-scale data for new systems.

Figure 4 – PMF between 2 polymethylene segments of length 2:

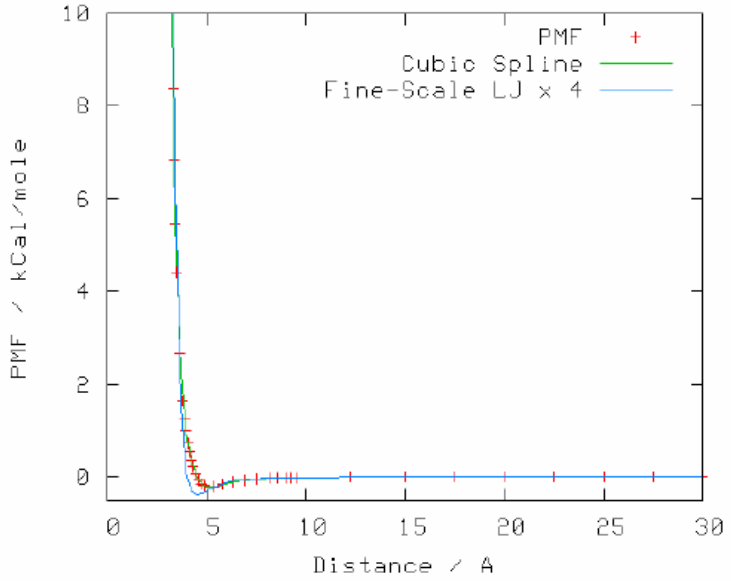


Figure 5 – Example Fine and Coarse-Scale Configuration: